

REFRAMING MENTAL CAUSATION

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
George J. Aulisio
May 2024

Examining Committee Members:

David Wolfsdorf, Advisory Chair, Philosophy
Miriam Solomon, Philosophy
Eli Alshanetsky, Philosophy
Joseph Corabi, External Member, St. Joseph's University

©
Copyright
2024

by

George J. Aulisio
All Rights Reserved

ABSTRACT

This dissertation explores the relationship between mental properties and physicalism to confront the apparent inconsistency between mental realism and the tenets of physicalism. As I see it, the major obstacle to fully integrating mental properties into physicalism is the feasibility of downward mental causation. Specifically, stringent physicalists find it contradictory to maintain that the mind can affect the body without contradicting the tenets of physicalism. This inconsistency claim is most notably addressed in the Causal Exclusion Argument.

Though I am not personally committed to physicalism as an absolute worldview, I respect its prevalence and the reasons for its dominance. Rather than reject physicalism, I approach the puzzle with epistemological humility and attempt to work within the scope of physicalism. This exploration involves critically examining physicalism's leading mental-physical relationships, focusing on emergence as a plausible means to reconcile mental and physical properties without undermining either. Ultimately, I propose a modified form of physicalism that maintains its metaphysical and epistemological theses but in a milder form that is more conducive to emergent mental phenomena and the aspects of reality that are nonlinear and indeterminate.

Guided by the work of Jaegwon Kim and Gerald Vision, this dissertation moves beyond their ideas, challenging reductionist perspectives within physicalism. The key contribution is the introduction of Dynamically Stable Causal Holism (or DSC Holism in brief), which represents a significant departure from traditional reductionist approaches, promoting a more holistic understanding of physicalism. Through nonlinear emergence

and DSC Holism, I confront the Causal Exclusion Argument. A secondary original contribution is my approach to these puzzles. I integrate and synthesize concepts from the philosophy of science and special sciences to offer a fresh perspective on physically compatible mental realism and downward causation.

DEDICATION

For my son Leonardo and my wife Lindsay.

You're in my heart always.

ACKNOWLEDGEMENTS

I am grateful to many people who helped me on my journey, culminating in this dissertation's writing and defense. I thank Gerald Vision for his time and expertise, as he was my primary advisor throughout my writing. I wish he were able to continue until the end. I owe a debt of gratitude to the other members of my committee: Joseph Corabi from St. Joseph's University for serving as my External Examiner; Eli Alshanetsky for joining my committee toward the end and being willing to participate in my defense while on sabbatical; Miriam Solomon, who served as a member of the committee from the beginning and helped significantly in my formation and understanding of concepts from the philosophy of science, and David Wolfsdorf who was a member from the start, helped me understand the rigor necessary to be a successful philosopher, and for taking on the elevated role of Advisory Chair toward the end. He was essential in helping shepherd this project through to completion. I thank you all for your time and expertise.

I would also like to take a moment to mention others who have helped me along the way. I am grateful to my first professors of philosophy, Steven Hales, Gary Hardcastle, Wendy Lee, and Kurt Smith, all of Bloomsburg University at the time, who sparked my academic interest. I am also indebted to Susan Schneider, who was at the University of Pennsylvania when I was there. She got me tuned into contemporary problems in the philosophy of mind and metaphysics and was how I was introduced to Jerry Vision and his work on emergence. Thank you to Colin Chamberlain for working closely with me on my history of philosophy preliminary paper and showing me how to stay persistent and thoroughly address critical feedback. Thank you to my fellow grad

students, who were also integral to my formation and thinking. I especially want to thank Brooke Sharp for reading drafts and meeting virtually while I was geographically away from Temple's campus. Thank you also to Nina Marie Campellone for her assistance and reassurance while copyediting this work.

I need to thank my colleagues at The University of Scranton for the variety of support I received through my doctoral studies. Thank you to Dean Charles Kratz, Bonnie Strohl, and Betsey Moylan for supporting my decision to enter a Ph.D. program while working full-time. I didn't fully understand what I was getting into then, but it was worth it. I also want to thank my faculty colleagues in the Weinberg Memorial Library, especially Sheli Pratt-McHugh and Donna Witek. Thank you for the roles you played in helping me complete the program.

I owe a debt of gratitude to the Provost's Office, including Provosts Harold Baillie, Joseph Dreisbach, Jeffrey Gingerich, and Michelle Maldonado, as well as the Treasurer's offices of The University of Scranton for supporting me in different ways. One such way was granting funding for an interdisciplinary "Clavius Seminar" on Emergence that I co-hosted with an incredible interdisciplinary colleague, Jason Graham, Professor of Mathematics. The discussions that came from that interdisciplinary discussion were very enlightening, and beyond this dissertation, I hope to co-author an article on emergence with Jason someday soon.

I also need to thank my colleagues in the philosophy department at The University of Scranton for allowing me to teach and inviting me to department talks and activities. Also, I need to thank the philosophy professors and graduate students at the

National Taiwan University for their invitation and participation in the two-part philosophy symposium between the National Taiwan University and The University of Scranton.

Thank you to my family for their love and support throughout this long journey and for understanding why I was always so busy, including my mother, Julie Aulisio, father, George Aulisio, and sisters, Denise Gurz and Meredith Oliver. Your support and encouragement through my highs and lows cannot be understated, and I will forever appreciate your thoughtfulness.

Thank you, especially to my wife Lindsay, for understanding all the time I needed to dedicate to my classes, studying, and writing. Many days of waking up hours before sunrise and getting home from work late were necessary to earn the doctorate. I'll never forget her understanding and dedication to my success and her unwavering encouragement, especially toward the end. I'll also always remember her love and support even while we were both grieving the loss of our son, Leonardo Aulisio. The strength of our love for one another not only survived the worst thing we could imagine, but it even grew stronger.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
DEDICATION.....	v
ACKNOWLEDGEMENTS.....	vi
LIST OF FIGURES.....	x
CHAPTER	
1. THE PUZZLE OF MENTAL PROPERTIES.....	1
2. MENTAL PROPERTIES UNDER PHYSICALISM.....	32
3. EMERGENT MENTAL PROPERTIES.....	70
4. CAUSALLY EFFICACIOUS MENTAL PROPERTIES.....	144
BIBLIOGRAPHY.....	205

LIST OF FIGURES

Figure	Page
1. Mental to Mental Direct Causation.....	176
2. Direct Mental to Physical Causation.....	176
3. Mental and Physical Causal Overdetermination.....	177
4. Emergent Mind without Intervention.....	194
5. Emergent Mind with Intervention.....	195

CHAPTER 1

THE PUZZLE OF MENTAL PROPERTIES

In this dissertation, I examine the challenges and potential resolutions of integrating mental properties within the framework of physicalism. The impetus of my exploration is the apparent contradiction between maintaining mental realism and physicalism. More specifically, I am interested in determining if metaphysically significant mental properties and their associated phenomena (e.g., the mind, conscious awareness, etc.) could play a legitimate causal role while subject to physicalism. To put it simply, is downward causation possible under physicalism?

My objective is to critically engage with physicalism while acknowledging its limitations. My intuition is that physicalism, at least as it is currently understood, does not accurately capture reality. That said, despite its flaws, physicalism remains our best worldview, and therefore, the best practice would be to work within the confines of that worldview before claiming it is irrevocably flawed. Acknowledging the limitations of physicalism aligns with Kim's views on physicalism, which he dubs "physicalism manqué." Unlike Kim, I posit that physicalism's primary limitation is that it is subject to reductionism even while advancing notions like non-reductive mental properties.

My attempt at resolving the dilemma between mental realism and physicalism requires addressing physicalism's leading mental-physical relationships. I critique these relationships in Chapter 2 and find that emergence is the most fruitful option. We have a plausible explanation for the relationship between mental and physical properties that

does not require severely harming mental properties and their apparent abilities, nor does it require sacrificing the critical tenets of physicalism.

The works of Jaegwon Kim and Gerald Vision primarily guide my position. I utilize Jaegwon Kim as an interlocutor mainly because of his pioneering work in physicalism, property dualism, reductionism, and mental causation. Conversely, Gerald Vision's work on emergence is foundational to my view of emergence. I largely agree with Vision but diverge from and expand on his work in several ways. Vision appears to find physicalism suspect, whereas I recognize a need to continue working within the confines of physicalism. Instead, I move to push our understanding of physicalism. Because of that difference in perspectives on physicalism, I refine and modify our understanding of physicalism.

I further distinguish my version of Emergence from Visions by introducing nonlinear dynamics and feedback loops while building up my position through ideas borrowed from the philosophy of science, such as *ceteris paribus* laws. The primary way that I modify our understanding of physicalism is to undermine reductionist thinking. In this work, I propose that all physicalists, including so-called non-reductive physicalists, have an underlying thread of reductionism in their methodology. In its place, I advance a form of holism called Dynamically Stable Causal Holism (DSC Holism). DSC Holism is ultimately the most significant contribution of this work and one that could be further refined in future studies.

I propose a revised physicalism that embraces my version of strong emergence focused on indeterminacy and nonlinear dynamics and a broader causal landscape through DSC Holism. I make inroads by looking beyond the underlying (or subvenient)

substrate to explain emergent phenomena. My holistic view of emergence expands the physicalist canvas so that emergent mental properties and minds remain part of the physical. Through this approach, I recognize the complex interplay of phenomena beyond reductive physical explanations.

Ultimately, this framework accommodates the possibility of downward mental causation while remaining compatible with physicalism.

Contextualizing the Puzzle of Mental Realism and Physicalism

When Princess Elisabeth pressed Descartes to explain how immaterial souls interact with material bodies, she commenced a dialogue on the mechanics of mental causation.¹ Over the centuries, few have found Descartes's explanation convincing, mainly because he held that the mind and body were metaphysically distinct substances.² Descartes's stark metaphysical distinction between mind and body made the interaction between the two substances unintelligible. Today, most philosophers have turned away from substance dualism and its troubled mechanics to focus on today's dominant worldview—physicalism. Though physicalism is today's dominant worldview, mind-body problems have not disappeared.

It is common for physicalists to acknowledge the reality of mental experiences (e.g., what it is like to be in pain). I refer to philosophers that acknowledge mental

¹ Princess Elisabeth writes, “[p]hysical contact is required for the first two conditions, extension for the third. You entirely exclude the one [extension] from the notion you have of the soul, and the other [physical contact] appears to me incompatible with an immaterial thing. This is why I ask you for a more precise definition of the soul than the one you give in your *Metaphysics*, that is to say, of its substance separate from its action, that is, from thought” (Elisabeth of Bohemia and Descartes 2007, 61).

² For instance, Yablo writes, “Descartes launches with his reply a grand tradition of dualist apologetics about mind-body causation that has disappointed ever since. Apologetics are in order because, as Descartes appreciates, his conception of mental and physical as metaphysically separate invites the question, ‘how, in that case, does the one manage to affect the other?’; and because having invited the question, he seems unable to answer it.” Yablo 1992, 245.

experiences as “mental realists.” Kim defines mental realism as “[m]ental properties are real properties of objects and events; they are not merely useful aids in making predictions or fictitious manners of speech.”³ Kim’s definition clearly states that mental properties are real properties and not linguistic constructs or the ideas of folk psychology in place to help us make sense of mental experiences. Mental realists believe that mental experiences, and possibly other mental aspects, constitute real properties just as physical properties, such as mass and shape, are real.

The combination of physicalism and mental realism manifests as property dualism in the literature. Lowe describes property dualism as “the doctrine that mental properties are distinct from and irreducible to physical properties, even if properties of both kinds may be possessed by the same thing, such as the human brain.”⁴ One way to distinguish mental and physical properties is through how we come to know them. Physical properties, such as mass, acidity, and viscosity, can be observed objectively and quantitatively. Conversely, mental properties, such as being in pain or tasting espresso, are experienced qualitatively in the private domain of one’s mind. Both property types are real to the property dualist, but they are distinguishable in how we come to know and examine them.

Notably, most property dualists doubt or outright deny the possibility of reducing mental properties. The physicalist property dualist maintains “that the human brain possesses both mental and physical properties but that these properties are distinct and

³ Kim 1993, 198.

⁴ Lowe 2008, 1018.

mutually irreducible.”⁵ Vision notes that “[p]roponents acknowledge that there is at most one substance, but maintain that in addition to physical properties, that substance also has irreducible mental properties.”⁶ Given the above explanations, another way to distinguish mental properties is that they are irreducible, whereas physical properties are reducible. For example, oil’s viscosity can be explained by referring to the molecules, physical principles, and properties; however, there is no similar reduction for the quality of the pain experienced during a headache.

Since physicalists acknowledge that mental properties are distinct and irreducible, they must posit a relationship between them and the physical world. Without a clearly defined relationship, mental properties could be considered independent of the physical world, or at least not tightly constrained by physicalist philosophy, because they are distinct and irreducible. Physicalists formulate the mental-physical relationship to limit the autonomy and capabilities of mental properties. By constraining mental properties through a tightly bound physical relationship, mental properties’ existence and causal role would be contingent upon and follow the principles of physical reality.

An important physicalist position is that *all causes are physical causes*. What entails a physical cause is a matter in need of further exploration, but for explanatory purposes, one example of physical causation is energy transference between two physical objects (e.g., when one billiard ball strikes another, the kinetic energy from the first billiard ball transfers to the second ball and propels it forward). With a view of physical causation like energy transference, one can easily identify why the causal interaction

⁵ Lowe 2008, 1019.

⁶ Vision 2011, 25.

between Descartes's immaterial souls and material bodies is unintelligible. Namely, the mental substance has none of the attributes of physical matter; it is aspatial and immaterial and, therefore, cannot transfer physical energy to matter. The two substances are incompatible because they abide by different rules and exist within two systems—a material world of physical causes and an immaterial world of non-physical causes.

Physicalism inherits its principles from the natural sciences that serve as the epistemological basis of physicalism. Crane writes:

[p]hysicalism must also contain the idea that explanations of our world must come to an end with physical principles and the appeal to purely physical entities. Explanations of natural phenomena (of whatever form they take) must bottom out in terms of explanations in the physical sciences.⁷

Though physicalism follows the natural sciences, the worldview holds physics in the highest regard. Witmer, for example, defines physicalism as:

[e]very law of nature and every particular fact is either physical or to be explained by the physical in such a way as to imply that the nonphysical facts are nothing over and above the physical facts, where the physical facts include the actual distribution of physical properties and the laws of physics.⁸

Physicalism's reliance on physics and the natural sciences is both its greatest strength and the root of its potential weaknesses. The fruitfulness and explanatory power of the natural sciences give physicalism considerable authority; however, even physicalists acknowledge that our best conception of physics is incomplete. Among the unanswered questions, physics does not provide a rich understanding of mental properties, including their origin (i.e., whether they are generated or fundamental), governing principles (i.e., if they serve a function and how), and universality (i.e., whether they are unique tokens or

⁷ Crane 2010, 28-9.

⁸ Witmer 2001, 69. Please note that I originally located this quotation in White 2016, 5.

repeatable). These unknowns challenge physicalists to define the relationship between mental properties and the rest of the physical world.⁹ As noted, property dualism presents physicalism with new mind-body problems. I am primarily interested in two problems that mental properties pose for physicalism.

First, how should we understand mental properties as members of the physical category if they are irreducible and distinct? Specifically, this is a question about the origin of mental properties and their relationship to the rest of the physical world. I contend that the physicalist needs to (a) posit the origin of mental properties, whether generative or fundamental, and (b) define the relationship between mental properties and the rest of the physical world. If the physicalist cannot do either (a) or (b), then the completeness of physicalism must be called into question.¹⁰ If physicalism is incomplete, then its role as the standard-bearing philosophical doctrine is unfounded so long as mental realism is upheld, and we should not be compelled to diminish mental properties based on physicalist doctrine. To make my point, I focus my critique on physicalists' most common mental-physical relationships, including type identity theory (i.e., reductive identity), token identity theory (i.e., non-reductive identity), functional reductions, supervenience, and emergence.

The second problem I am interested in is whether mental causal efficacy is viable under a revised physicalism. This question originates in each conscious agent's direct experience of what seems to be the mind's causal ability. One hallmark of mental life is

⁹ Because of the shortcomings mentioned above, some physicalists appeal to a moderately future physics that they believe will explain unanswered questions about mental properties.

¹⁰ Stoljar 2021 asks the completeness question. He writes, "[w]hat relation or relations must obtain between everything and the physical if physicalism is to be true?"

the mind's apparent ability to causally affect one's body, which can impact the wider physical world. The issue with mental causal efficacy is that physicalists have renounced the possibility that mental properties have causal powers because of their interpretation of the principles of physical theory. The physicalist position is most clearly articulated by the Causal Exclusion Argument, which I address in greater detail shortly.

I find these two problems important to address for several reasons. First, physicalism posits the total composition of reality and is, therefore, a worldview. As a worldview, physicalism limits possibilities and explanations because that which is possible must be consistent with the worldview. Mental causal efficacy, or the view that one's mind initiates at least some bodily actions, seems apparent to conscious agents. Though mental causal efficacy appears to be an evident feature of our experience of reality, we question its very possibility because physicalists maintain that it is incompatible with physicalism. I find it essential to press the notion that we should give up something apparent, like mental causal efficacy, because of an inconsistency with the tenets of a worldview that fails to explain mental properties in the first place.

Second, we should attempt to hold onto mental causal efficacy more closely or be certain about eliminating it because of its role in so much of our understanding of reality. Physics is a powerful source of knowledge, but unless a grand unified theory is true, it follows that parts of reality cannot be explained by physics.¹¹ Eliminating mental causal efficacy because of physicalism could affect all significant areas of human affairs, including cultural, economic, ethical, and humanistic considerations. These human affairs provide meaning to our lives and guide our aspirations, yet our best physics does not and

¹¹ For example, physics cannot replace biology if there is no grand unified theory.

presumably cannot address these aspects, meanings, and aspirations. Mental causal efficacy plays a crucial role in understanding these areas of reality. Still, if physicalism eliminates mental causal efficacy, human actions result from physical processes and our human affairs, and the meanings attributed and aspirations sought ultimately fall under the domain of physical theory as it is currently understood.

Let us assume that physical theory eliminates the possibility of mental autonomy and mental causal efficacy. Does this consequence entail dismissing the pretenses of teaching moral responsibility and discernment in favor of cutting to the chase and dealing directly with the most effective means of bringing about particular physical effects?¹² In other words, does physicalism entail embracing more direct approaches to achieving desirable outcomes, such as indoctrination and coercion? The ethically cognizant physicalists will balk at the suggestion, but what is the physicalist rationale for recoiling? Without physicalist (i.e., physics-based) justifications for moral education over coercion, the physicalist would have to appeal to higher-level concepts, such as moral responsibility, agency, inherent dignity, and so on, to explain why coercion is reprehensible. Physics does not deal with morality, culture, and economics, so physicalists must embrace higher-level concepts rooted in mental life. I am hard-pressed to see how this is a valid approach from the physicalist who is adamant that mental life is *just* a straightforward physical process. Yet, physicalism entails this incongruous

¹² Along similar lines, the notion of recidivism or believing that someone could change their behavior if they choose to make better decisions may also be misguided. Rather, if an agent's actions are merely physiological responses, then there may be more direct mechanisms to positively manipulate the behavior of past offenders. Perhaps some moderately future physics will reveal that lobotomies are in fact the most effective way of reducing criminal or immoral behaviors.

approach, denying the significance of mental life while appealing to and deferring to higher-level concepts as a necessary course of action.

Ultimately, I am asking if there is a certain point where physicalism, as it is currently understood, breaks down. Physicalists believe their worldview is the standard bearer that everything must conform to. Yet, physicalists regularly appeal to irreducible concepts, ideas, and things so as not to upend those aspects of reality that are, in fact, most precious to humanity. A stringent physicalist approach would abandon mental causal efficacy based on its inconsistency with the worldview. Contra this hardline physicalism, physics cannot explain all of reality without appealing to physically inexplicable facts, such as mental properties and their perceived effects on our beliefs and actions. I believe that the incongruency between physicalism, mental realism, and mental causal efficacy is not due to physicalism in and of itself. Instead, the issues stem from the reductive approach of physicalism.

Rather than outright reject physicalism, I propose that physicalists could find mental causal efficacy viable when positing that (i) mental properties are conceptually and explanatorily distinct from reductive physical properties while remaining (ii) metaphysically grounded in the physical through an emergence relationship. After accepting (i) and (ii), the reformed physicalist needs to acknowledge (iii) a levels view of reality based on emergence referred to as the layered conception, and (iv) non-reductive causation in the form of what I call Dynamically Stable Causal Holism—an interlevel causal theory that seeks system stability. If physicalists accept (iii) levels and (iv) causal holism, then (v) mental properties can be causally efficacious without violating the causal closure of the physical. A reformed physicalist can retain the important piece that

physical reality is fundamental without undermining our humanity and the aspects of reality closely rooted in mental life.

For the remainder of this chapter, I set the stage for the larger argument in this work by overviewing some of the initial mental-physical dilemmas and terminological and argumentative context for my discussion of mental causal efficacy under physicalism.

Physicalism and Mental Properties

The term “materialism” is sometimes used interchangeably with “physicalism;” however, it is possible to distinguish the terms. Materialism, which appears earlier in philosophical and scientific discourse, was traditionally used to describe the metaphysical thesis that reality is ultimately composed of matter. The more recent term “physicalism” was introduced because of the need to refer to more than inert matter when describing reality, including, for example, laws of nature and physical forces.¹³ Notably, physicalism emphasizes its connection to physics more so than materialism. Stoljar writes:

[...] it is tempting to use ‘physicalism’ to distance oneself from what seems a historically important but no longer scientifically relevant thesis of materialism, and related to this, to emphasize a connection to physics and the physical sciences. However, while physicalism is certainly unusual among metaphysical doctrines in being associated with a commitment both to the sciences and to a particular branch of science, namely physics, it is not clear that this is a good reason for calling it ‘physicalism’ rather than ‘materialism.’¹⁴

Though Stoljar notes the etymology of “physical” does not give it priority over the alternative term “materialism,” I will primarily use “physicalism” to emphasize the

¹³ See Papineau 2001 and Stoljar 2021 for brief but more detailed historical analyses of the transition from “materialism” to “physicalism.”

¹⁴ Stoljar 2021, 3-4.

doctrine's relationship to physics, but I also reference philosophers that utilize "materialism."¹⁵ In the context of this work, both terms should be understood to mean a worldview committed to physical theory (i.e., physics).

Physicalism has been a leading conception of reality among contemporary philosophers and scientists since the early 20th century.¹⁶ Hellman and Thompson note that physicalism comprises two principles, they write:

[t]here is first a principle of Ontological Physicalism, or what we have called the Principle of Physical Exhaustion, which provides a non-question-begging construction of the informal claim that everything is physical. [...] The principle of Ontological Physicalism holds that the universe so delineated embraces everything there is.

[...] Physical Determination principles comprise the second part of Physicalist Materialism. Where the informal statement [of physicalism is] "Physical facts determine all facts."¹⁷

Hellman and Thompson describe physicalism as an all-encompassing worldview with a metaphysical and epistemological basis. The principle of physical exhaustion is the metaphysical thesis that reality comprises only the physical. The principle of physical determination speaks to the epistemological basis and explanatory power of physicalism. For simplicity, I refer to these as physicalism's metaphysical and epistemological theses rather than Hellman's and Thompson's preferred Ontological Physicalism and Physical Determination, respectively.

The metaphysical and epistemological theses serve as solid foundations, but these principles also charge physicalism with being a complete worldview. For instance, to

¹⁵ Some examples of contemporary philosophers who use "materialism" in lieu of "physicalism" include Smart 1963 and Strawson 2008. Where Smart uses materialism as a true synonym, Strawson distinguishes his materialism from physicalism and its reliance on physics (*passim*).

¹⁶ Stoljar 2021 provides both the historical and contemporary context of physicalism.

¹⁷ Hellman and Thompson 1977, 310-1.

posit that all things are physical suggests that if there are non-physical things, then physicalism is metaphysically incomplete. Furthermore, the epistemological stance of physicalism is mistaken if physical facts do not determine all facts.

Physicalism's two theses appear to make physicalism a worldview without qualification; however, physicalism's formulation is not universal. Stoljar writes, "many contemporary philosophers assume that they understand physicalism *somehow*, and concentrate instead on arguments for and against it."¹⁸ In a particular light, physicalism is a firm worldview based on clearcut metaphysical and epistemological theses, but there is a gray area. There is the hard-to-deny presence of mental aspects or properties that comprise our experience of reality. Given the uniqueness and ubiquity of mental properties to conscious agents, physicalists must resolve, or at least explain, how mental properties are consistent with physicalism. So, even though physicalism is a leading conception of reality, mental and physical relationships remain puzzling. I now examine property dualism and how it fares against physicalism's metaphysical and epistemological theses. I begin with the epistemological thesis.

Epistemological Thesis

The epistemological completeness of physicalism could be called into question when acknowledging mental properties. Take, for example, the "hard problem" or the mystery of *why* at least some physical states come with qualitative experiential awareness. Chalmers writes:

[w]hy is all this [neural] processing accompanied by an experienced inner life? Sometimes this question is ignored entirely; sometimes it is put off until another day; and sometimes it is simply declared answered. But in

¹⁸ Stoljar 2010, 530-1 (italics added for emphasis).

each case, one is left with the feeling that the central problem remains as puzzling as ever.¹⁹

The hard problem asks why qualitative experience should exist when functionally (or otherwise) physical state changes without the corresponding mental experiences would be enough to explain transpiring events. The hard problem is a serious challenge for physicalism because the physical world and its principles should be enough. Yet, this qualitative component persists regardless of the otherwise closed and complete physical world. Robinson writes:

[...] within our present framework of concepts, theories, and basic principles. [...]. The difficulty will remain, therefore, so long as our conceptual framework maintains its present requirements for explanation. In other words, we will not be able to solve the Hard Problem unless we can come to accept something other than our present modes of explanation as providing something like the kind of intellectual satisfaction, or relief from the sting of curiosity, that we now get from explanations.²⁰

Put simply, the hard problem suggests that qualitative inner experiences (i.e., qualia) come along for free, which calls into question the epistemological completeness of physicalism.²¹

Physics and the special sciences do not explain the *why* of mental experience. However, the physicalist may contend that answering *why* is an unfair expectation. I argue in Chapter 3 that physics and the special sciences do not clearly explain the *how* of mental properties either, which should be even more troubling for a worldview that aims to be complete.

¹⁹ Chalmers 1996, xxi. I have added the bracketed text for clarity.

²⁰ Robinson 1996, 15-9.

²¹ McGinn 2004 believes there is an answer to the hard problem; however, humans are cognitively closed to that answer.

The hard problem is, of course, not the only mental-physical dilemma to challenge physicalism. Davidson lists “perceivings, remembering, decisions, and actions” as typical mental experiences.²² Davidson notes those mental experiences “resist capture in the nomological net of physical theory.”²³ He writes:

[a]nomalous monism resembles materialism in its claim that all events are physical, but rejects the thesis, usually considered essential to materialism, that mental phenomena can be given purely physical explanations. Anomalous monism shows an ontological bias only in that it allows the possibility that not all events are mental, while insisting that all events are physical.²⁴

Anomalous monism, Davidson’s particular brand of physicalism, acknowledges a central tension between physicalism and the mental: physical facts do not adequately explain mental facts. For example, there is evidence that when c-fibers are stimulated, a person experiences pain. Given this evidence, a direct line can be drawn between certain physical and mental facts. However, the physical fact (i.e., c-fibers firing) and the physical principles (i.e., the principles of physics, biology, neurology, etc.) still leave much to be desired as an explanation of the mental fact (i.e., pain).²⁵

The anomalous nature of mental properties is an epistemological problem for physicalism that confronts the possibility of it being a worldview without qualification. Though proponents of anomalous monism accept that physical facts cannot explain mental properties and facts, they do not find this to be a refutation of physicalism. My proposed solution, which culminates in Chapter 4, is consistent with anomalous monism

²² Davidson [1980] 2001, 207.

²³ Ibid.

²⁴ Ibid., 214.

²⁵ To further complicate the matter for physicalists, proponents of multiple realizability contend that we cannot directly connect certain physical facts with mental facts. For example, it is possible to experience pain in the absence of c-fibers firing as well as the inverse of those facts. I address multiple realizability in greater detail as a problem for physicalism in Chapter 2.

and the stance that physicalism is not refuted due to the anomalous nature of mental properties.

Another challenge for physicalism is recognizing that qualitative phenomenal mental experiences populate the whole of reality for conscious agents, and those experiences appear to fall outside of the descriptions and principles available to physicalists constrained by reductive physicalism. There is something it is like to be you. Whether your attention is pointing internally, externally, or toward an abstraction, your beliefs about those spaces are formed by mental properties. I am willing to grant it whether something exists in those spaces beyond your mental properties, at least for the external world. Still, the granting of “stuff” beyond one’s mental experience does not minimize the fact that we only know of said “stuff” by way of our mental experiences of them.

Suppose the epistemological principle is not to be abandoned because of the positions of anomalous monism. In that case, the physicalist must explain what “determines” means in the context of “physical facts *determine* all facts?” In other words, given the hard problem, anomalous monism, and other mental-physical dilemmas, physicalists that accept mental realism must define the relationship between mental facts and physical facts or else risk diluting or undermining the epistemological principle, which in turn calls into question the completeness of physicalism. Chapters 2 and 3 address this issue in greater detail. In Chapter 3, I posit emergence is the best hope for explaining the mental-physical relationship while addressing the metaphysical and epistemological theses.

Today, many philosophers identify as non-reductive physicalists. These physicalists admit that mental properties cannot be reduced to physical properties and principles. Melnyk describes physicalism as “a view about the relationship between people’s mental properties and their physical properties. It claims that people’s mental properties are nothing over and above – nothing additional to – their physical properties.”²⁶ Given Melnyk’s description, physicalism need not represent a world that is comprised of only clear-cut physical things and explanations. Instead, physicalism can remain intact by ensuring that mental properties conform to physical theory.

One way to make mental properties conform to physical theory is to maintain that the physical necessitates the mental. Under this physicalist interpretation, mental properties and facts cannot exist without physical properties and facts (i.e., physical properties have metaphysical priority). However, physicalists typically want to go beyond physical necessity and claim that all mental properties and facts are contingent upon physical states. Under this stronger physicalist view, mental properties exist because of and are determined by physical theory. However, the physicalist must be able to posit the relationship between mental properties and physical theory, or else this strong claim is unfounded. The proposed ways of formulating the relationship between the mental and physical are hotly contested. In Chapter 2, I address the mental-physical relationships most often postulated by physicalists and argue that they fall short. I now turn to the metaphysical thesis and property dualism’s shortcomings.

²⁶ Melnyk 2019.

Metaphysical Thesis

By distinguishing between physical and mental properties, property dualism indicates that mental properties are *sui generis* because they are qualitative, private, and irreducible. Conversely, physical properties and substances are quantitative, public, and reducible. By recognizing the uniqueness of mental properties, property dualists accept two distinct types of properties. *Prima facie*, if mental properties are distinct from physical properties, then it seems to follow that reality is not all physical. If everything is not physical, then the metaphysical completeness of physicalism is called into question. Thus, to resolve any potential contradiction, physicalists must define distinct mental properties as legitimate components of physical reality.

Property dualists generally remain committed to the completeness of physicalism despite accepting two property types because they adhere to one substance type (i.e., physical monism). Though property dualists are generally satisfied with the physical completeness of property dualism, there are avenues to challenge its metaphysical completeness. Schneider argues that the two leading metaphysical positions on substances, bundle theory and substratum theory, provide good reason to question if property dualism entails substance dualism or undermines monism. Schneider writes, “if your property dualism is to be a *bona fide* physicalist position, [...] you must explain how your substances turn out to be genuinely physical. This demands some metaphysical soul searching rather than a crude appeal to the doctrine that all substances are physical.”²⁷ On metaphysical grounds, Schneider contends that property dualism is not well-founded as a form of substance physicalism and that a hybrid substance view may be the best path

²⁷ Schneider 2012, 62.

forward.²⁸ Schneider's argument should be well-taken as it is a powerful challenge to the metaphysical completeness of any physicalist theory of mind that hinges on property dualism.

Lycan takes a different approach to challenge property dualism and turns nine objections to substance dualism toward property dualism. He concludes that property dualism "is perhaps a little better off than [substance dualism]. But not so much better off that property dualists should go on boasting that they are not so crazy as to be Cartesians."²⁹ However, he does find the problems unique to property dualism to be a serious challenge. One such challenge regards the origin of mental properties. Lycan raises Chalmers's concern that mental properties must either result from strong emergence or, as Paul Churchland proposes, these properties must be fundamental.³⁰ Neither option is ideal for a strong formulation of physicalism that endeavors to give up as little ground as possible to the challenge of mental realism.

Even the most respected physicalists tacitly acknowledge the fact that physicalists make metaphysical assumptions. For instance, Jaegwon Kim recognizes problems with physicalism, but he finds his version to be the best theory available. Kim writes:

[t]he position is, as we might say, a slightly defective physicalism—physicalism manqué but not by much. [...] Physicalism is not the whole

²⁸ Schneider writes, "[a]lthough many contemporary property dualists aspire to uphold [substance physicalism], today I have urged that they have little justification for doing so. At this point, the property dualist who espoused [substance physicalism] might choose to repudiate [property dualism] because he regards substance dualism as being too implausible. Still, insofar as he is willing to accept [property dualism], I suggest that he look towards the hybrid view as it is the lesser of the two evils; [...]" Schneider, 2012, 73.

²⁹ Lycan 2013, 541.

³⁰ Lycan credits Chalmers for sharing in conversation the strong emergence objection. Lycan quotes Paul Churchland [1984] 2013, where Churchland refers to the fundamental mental property view as "elementary property dualism." Lycan 2013, 540. Please note that Churchland refers to "*elemental-property dualism*." Churchland [1984] 2013, 20.

truth, but it is the truth near enough, and near enough should be good enough.³¹

I rely on his positions throughout this work because of Kim's stature in the field and his arguments against mental causal efficacy.

Part of Chapter 2 argues that physicalists make metaphysical assumptions that unfairly favor their worldview. I call into question the legitimacy of those assumptions. In Chapter 3, I explore the possible origin of mental properties. Using the dichotomy developed by Churchland, a noted physicalist, I explore if mental properties are generated or fundamental and whether either explanation affects the metaphysical completeness of physicalism.

Mental Causal Efficacy

Physicalists who are mental realists (henceforth MR physicalists) acknowledge the presence of mental properties in causally connected events. For instance, the MR physicalist must accept that some physical causes precede and lead to mental events (e.g., a paper cut leads to pain). Throughout this work, I refer to mental causation and mental causal efficacy. Mental causation includes three *directions* of causation involving mental properties or states. The three directions are figurative and not literal directions in space-time.

The first figurative direction is vertical determination, which might be most easily understood as bottom-up causation. In cases of vertical determination, physical causes have physical effects that engender mental effects. For example, dropping a stone on my foot causes various physical state changes within my body, leading me to experience

³¹ Kim 2005, 174.

various mental properties, including pain. MR physicalists generally find no issue with vertical determination. A reoccurring issue for vertical determination is that it entails epiphenomenalism. To be epiphenomenal is to have no causal force and ultimately means that it is merely an after-effect of physical processes. I address the problem of epiphenomenalism throughout this work.

Second, horizontal mental causation, where either (a) a mental cause leads to a mental effect, or (b) a mental-physical cause leads to a mental-physical effect. Property dualists could subscribe to either (a) or (b); however, MR physicalists will opt for (b) because physicalism maintains that where there is a mental cause, there must also be a physical cause for the event to be consistent with physical theory. Option (a) does not clearly reference physical causes, making it unviable for physicalists. Because of these considerations, many physicalists strive for a mental-physical relationship, such as identity, to inextricably tie mental causes to physical causes. Lastly, option (a) is problematic for the physicalist because there is no physical framework by which the physicalist can explain how a freestanding mental cause could have a freestanding mental effect.

Third, downward or top-down mental causation, where a mental property, state, or event causes a physical effect. Consider, for example, a series of beliefs leading to a volition that causes the body to act in a specific way to realize that volition. As an illustration, imagine that after reflecting on my well-being, I begin exercising regularly. While downward causation may refer to preceding and underlying physical states when explaining how a mental cause can have a physical effect, the mental properties involved

must play a necessary role in causing the effect for it to be a genuine instance of downward causation.

By and large, physicalists deny the possibility of downward mental causation, instead attributing all effects to physical causes. Specifically, they conclude that mental properties are unnecessary for physical effects. Though physicalists generally do not find mental properties to be causally efficacious, it is common for MR physicalists to attempt to save a causal role for mental properties by conjoining them to physical states. By doing this, MR physicalists allow mental properties to be causally relevant but not efficacious.

For the MR physicalist, causally relevant mental properties make for an explanatorily valuable picture that MR physicalists can utilize in folk discussions. For example, if a physicalist spectator inquires why I am jogging, they may accept my explanation that “I made up my mind to begin exercising.” However, they will maintain that only vertical determination or horizontal mental causation can provide a physically sound explanation for my running around the track.

If the physicalist spectator opts for vertical determination in the above scenario, then the mental properties are epiphenomenal and can, at best, play an explanatory role in causal explanation. If the spectator opts for horizontal mental causation, they must decide how to explain the role of mental properties in my explanation about “making up my mind.” Leading options for MR physicalists include identity theories or functional reductions. I address both options in Chapter 2.

In the case of identity and functional reductions, the MR physicalist conjoins mental properties to physical properties and appeals to the causal power of the physical property to explain causal events. By conjoining mental property, *M*, to physical

property, P , it means that M is identical with, reduced to, or realized by P . The physicalist can then identify M 's causal task with P 's causal mechanics. In the case of functional reductions, the description of M is conjoined to P 's processes through systematic explanation. So, any description of M is always accompanied by an explanation of what is occurring with P .³² For instance, mental property, M_I , can be identified with or functionally reduced to physical property, P_I . In essence, lateral causation of this sort can be called “qua causation” because if M_I qua P_I , and P_I causes P_2 , and P_2 instantiates M_2 , then M_2 qua P_2 and M_{cause} qua P_{cause} .³³ That said, qua causation does not entail mental causal efficacy because all of the causal work is attributed to the physical.

The standard physicalist position is to give no more than a lesser causal role to mental properties. To the physicalist, mental properties do not have the power to affect physical states and merely play an explanatory role in the descriptions of events. In contrast to the standard position, I advocate for the viability of downward causation. To advocate for downward mental causation, I must address the standard physicalist argument against mental causal efficacy, namely the Causal Exclusion Argument.

Causal Exclusion

Jaegwon Kim explores the consequences of simultaneously holding four positions that property dualists operating under the constraints of physicalism typically maintain.

The four positions:

³² Kim 2005 writes, “[t]o reduce a property, say being a gene, [...], we must first ‘functionalize’ it; that is, we must define, or redefine, it in terms of the causal task the property is to perform. Thus, being a gene may be defined as being a mechanism that encodes and transmits genetic information. [...]. Next, we must find the ‘realizers’ of the functionally defined property. [...]. Third, we must have an explanatory theory that explains just how the realizers of the property being reduced manage to perform the causal task” (101).

³³ See Horgan 1989.

- I. Mental-physical property dualism: “the view that mental properties are irreducible to physical properties.”³⁴
- II. Mind-body supervenience: “Mental properties strongly supervene on physical/biological properties. That is, if any system *s* instantiates a mental property *M* at *t*, there necessarily exists a physical property *P* such that *s* instantiates *P* at *t*, and necessarily anything instantiating *P* at any time instantiates *M* at that time.”³⁵
- III. Causal closure of the physical: “If a physical event has a cause at *t*, then it has a physical cause at *t*.”³⁶
- IV. Principle of Causal Exclusion: “If an event *e* has a sufficient cause *c* at *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination).”³⁷

The first claim, metaphysical property dualism, takes mental realism seriously and recognizes the distinctive nature of mental properties. Without the first claim, there is no property dualism.

The second claim, mind-body supervenience, establishes that there is a physical relationship with mental properties. Though capable of alteration, the second claim is important to property dualism for it to be ultimately classified as a physicalist position. For instance, physicalism may survive some mental property anomalism, but the two aspects of reality could only diverge so much before physicalism is undermined. Physicalism must maintain some hold over mental properties in the form of a mental-physical relationship, or it will lose its authority to dictate the parameters that mental properties must abide by. For example, a mental-physical relationship helps to preclude the possibility of mental properties having extra-physical powers.

³⁴ Kim 2005, 22.

³⁵ Ibid, 33.

³⁶ Ibid, 15.

³⁷ Ibid, 17.

The third claim, the causal closure of the physical, is one of the cardinal tenets of physicalism. Causal closure ensures that all physical events are the result of physical causes. Popper and Eccles note that physicalists take a strong position on the causal closure of the physical, and their definition precludes the possibility of mental causes playing a role in physical effects. They write:

[...] physical processes can be explained and understood, and must be explained and understood, entirely in terms of physical theories. I call this the physicalist principle of the closedness of the [physical world]. It is of decisive importance, and I take it as the characteristic principle of physicalism or materialism.³⁸

Popper's and Eccles's understanding of causal closure posits that physical effects should be fully comprehended within the framework of physical theories. Based on this, one expects physicalism to strive for alignment with established physics. On the other hand, Kim's interpretation of causal closure does not dismiss the potential role of mental causes influencing physical effects. Nevertheless, he clearly emphasizes that physical effects must connect to physical causes. How causal closure is defined helps discern whether preceding physical causes are wholly sufficient or simply necessary. Many physicalists hold the third claim in high regard. Thus, it should, at a minimum, align with Kim's criteria, emphasizing the necessity of preceding physical causes. I delve deeper into the intricacies of causal closure in Chapters 3 and 4.

The fourth claim, the principle of causal exclusion, supplements the principle of causal closure. As noted, Kim's definition of causal closure does not preclude mental properties from having a causal role. The causal exclusion principle, however, effectively eliminates mental properties causing physical effects. Suppose physical effects have

³⁸ Popper and Eccles 1985, 51.

physical causes (i.e., causal closure), and one sufficient cause is required to explain physical effects (i.e., causal exclusion). In that case, distinct mental properties are superfluous and do not need to be referred to when explaining physical effects.

Chalmers's notes:

[t]he best evidence of contemporary science tells us that the physical world is more or less causally closed: for every physical event, there is a physical sufficient cause. If so, there is no room for a mental 'ghost in the machine' to do any extra causal work.³⁹

Since we have defined mental properties as distinct properties that are irreducible to hallmark physical properties, it appears that they would have no role to play in physical events.

From the four claims, Kim develops his Causal Exclusion Argument, a notable and convincing argument contending that all dualist views could not account for the causal efficacy of the mind.⁴⁰ The Causal Exclusion Argument is a four-premises dilemma that contends all four premises cannot be true. The four premises are:

1. the mental is not identical to the physical,
2. mental properties cause both physical and mental effects,
3. the causal closure of the physical—the physicalist tenet that every physical effect has a sufficient physical cause, and
4. causal overdetermination is rare and cannot justify large-scale systematic mental causation.⁴¹

³⁹ Chalmers 1996, 125.

⁴⁰ See Kim 1998. For a complete overview of his work on causal problems, see Kim 2005.

⁴¹ See Vision 2011, 95-7 for a clear overview of the causal exclusion argument and the various options one might select in response.

The argument is powerful because each premise is reasonable, and they combine to detract from the possibility of mental causal efficacy. However, combining claims does not eliminate the possibility of mental causal efficacy. One way the door is left open for mental causes is that the fourth claim notes that there can be more than one sufficient cause in cases of genuine causal overdetermination. This leaves proponents of mental causal efficacy an opening to argue for mental causal efficacy through causal overdetermination. Alternatively, proponents can also reject or revise one or more of the premises of the CEA. To briefly sketch what I address in the following chapters, consider the following options and consequences.

To reject premise 1 is to identify the mental with the physical. When the property dualist identifies mental states with physical states, they eliminate the need for mental properties to have causal powers of their own. When mental properties are identical to physical states, the causal work can be isolated to physical properties. As such, the mental state may only achieve causal relevance through quausation rather than causal efficacy. Rejecting premise 1 is a non-starter for proponents of mental causal efficacy.

To reject premise 2 is to deny that mental properties cause *both* mental and physical effects. Denying premise 2 leaves us with three options: either mental properties cause (a) only mental effects, (b) only physical effects, or (c) neither. To cause either (a) or (b) entails a disconnect between mental and physical effects, which calls into question the completeness of physicalism. To elaborate, consider the consequences of option (a). If M_1 causes M_2 but does not cause P_2 , we have a parallel but disconnected view of mental and physical events. In this scenario, reality has parallel tracks with no apparent causal connection. The major flaw of any parallel tracks view is the absence of a

physically (or naturally) grounded explanation for how the two parallel tracks remain synchronized. Typically, those who advocate for parallelism appeal to God as the governor who keeps both tracks aligned.

Conversely, option (b) affirms mental causal efficacy but requires a physically sound explanation for how mental causes bring about physical effects. Accomplishing option (b) in a way that respects physical theory is no small feat. In Chapter 4, I explain how higher-order mental states causally interact with underlying physical states by utilizing emergence and a framework I call “Dynamically Stable Causal Holism.”

Lastly, to opt for (c) entails epiphenomenalism. Option (c) outright rejects mental causal efficacy and, therefore, would not be an option for its proponents.

To reject premise 3 is to deny the cardinal tenet of physicalism. Most, if not all, physicalists find denying or diminishing causal closure to be too radical.⁴² That said, I find it important to address and clarify premise 3. For instance, causal closure is primarily based on the conservation of energy in closed systems, an idealized principle of physics.⁴³ In chapters 3 and 4, I examine this physicalist position.

To reject or modify premise 4 requires the mental realist to explain wide-scale causal overdetermination, which the physicalist maintains leads to a knot of metaphysical and epistemological difficulties. I agree with the physicalist that if the proponent of mental causal efficacy relies on causal overdetermination to vindicate mental causal efficacy, then the proponent is responsible for explaining wide-scale systematic causal

⁴² For instance, White writes, the “strategy of rejecting [causal closure of the physical] is, [...], typically viewed as simply too crazy to be taken seriously.” White 2017, 387.

⁴³ In Chapter 3 and 4 where I address this topic more thoroughly.

overdetermination. I respect this challenge in Chapter 4, where I posit my position on causal holism.

Suffice it to say that the various approaches to resolving the Causal Exclusion Argument led to different consequences deemed unpalatable to those sympathetic to mental causal efficacy and physicalism. That said, I do not find the Causal Exclusion Argument irreproachable. The intersection of mental realism and physicalism leads to various dilemmas, which has led to physicalists refining their position in several ways. I contend that there is a revision that should allow for mental causal efficacy. I dedicate the remaining chapters of this work to posit my argument in favor of mental causal efficacy.

Roadmap

This dissertation is divided into four distinct but interconnected chapters that build upon one another. Each chapter offers critical insights and contributes a deeper understanding of the physicalist perspective on mental phenomena.

In Chapter 1, I focused on contextualizing the challenge of mental properties in a physicalist framework. By building up mental realism and highlighting the limitations of physicalism, I set the foundation and tone for this exploration of mental properties and downward causation.

In Chapter 2, I propose a framework for evaluating the success of physicalism. I address leading physicalist positions on mental-physical relationships. The positions I focus on are identity, supervenience, and functional reductions. I argue that the reasons to accept each relationship type fall short of the claim itself, and ultimately, the positions do not improve our understanding of mental properties or mental-physical relationships. I

explain that these mental-physical relationships merely capitalize on the perceived connection between mental and physical properties. I argue that the connection between the mental and the physical cannot be taken as evidence in favor of physicalism unless one has preemptively resolved to prove physicalism. I contend that physicalists rely on their faith in physicalism to posit that mental properties must abide by a stringent formulation, but the grounds for this are dubious.

I accomplish this by showing that the mental-physical relationship types make bold claims with insufficient justification. For instance, these relationships do not explain the *how* and *why* of mental properties. I argue that MR physicalists must posit an explanation for the reality of mental properties. For mental-physical relationships to constrain mental properties, they must posit the origin of mental properties. I demonstrate that regardless of whether mental properties are generative or fundamental, they should serve as a major source of doubt that these mental-physical relationships can constrain mental properties.

In Chapter 3, I overview emergence and argue in its favor as a more fruitful and explanatory mental-physical relationship than those addressed in Chapter 2. I argue that emergence is a superior physicalist explanation because it (a) takes a stand on the origin of mental properties, (b) commits to a physicalist metaphysics, (c) allows for irreducible, unpredictable, and novel properties, and (d) does not limit the power of mental properties based on uncertain evidence or blind faith in physicalism.

Appealing to emergence also approaches the mental-physical relationship with candor about what we can claim to know. I argue that supervenience and identity overextend themselves by claiming more than we know about the relationship between

the mental and physical. In contrast, emergence takes a more tempered approach to the uncertain connection between mental properties and physical states. Despite the uncertainty, there is reason to accept emergence as a viable explanatory avenue. I conclude that emergence is the best explanation available, one capable of providing a theory that recognizes the connection between the mental and physical while embracing the explanatory gaps that we cannot close. Acknowledging the explanatory gap and tying it to nonlinear dynamics and indeterminacy in nature sets it apart as a superior relationship type. It is further supported by a layered conception of reality that, among other things, respects the autonomy of the special sciences and physics.

In Chapter 4, I note that causation is a mysterious process that is foundational to our experiences and explanations of reality. I propose that the Causal Exclusion Argument fails to preclude the possibility of mental causal efficacy and that emergent mental states play an essential role in human behaviors and actions. My argument hinges on the novel framework I introduce called Dynamically Stable Causal Holism.

Furthermore, I address the Causal Exclusion Argument head-on by explaining why emergence and Dynamically Stable Causal Holism allow mental causal efficacy without rejecting physicalism. Ultimately, I conclude that a multimodal approach to reality requires reductive and holistic approaches to analyzing complex systems. Through this multimodal approach, mental causal efficacy can be considered a physically compatible form of causation.

CHAPTER 2

MENTAL PROPERTIES UNDER PHYSICALISM

In this chapter, I have three objectives. First, I outline the parameters of physicalism to elucidate how mental realism can qualify as a form of physicalism. Second, I overview (1) type identity, (2) token identity, (3) supervenience, and (4) functional reductions. Third, I critique the mental-physical relationships overviewed in this chapter by arguing that these relationships fail in particular ways or must ignore certain puzzles that physicalists should address.

Notably, these leading relationships do not address the origin of mental properties. I argue that addressing their origin would be crucial to demonstrating how they align with the metaphysical and epistemological theses of physicalism. Ultimately, I argue that the mental-physical relationships that I address come up short. This chapter primarily serves as an overview of the leading mental-physical relationships in the field and sets the stage for my novel contributions in Chapters 3 and 4. In Chapter 3, I posit that emergence, a mental-physical relationship that takes a stand on the origin of mental properties, is the best path forward for mental realists sympathetic to physicalism.

Physicalism and Completeness

An idealized version of physicalism would hold that everything is physical *tout court*, and every fact is a physical fact derived from physical theory. However, non-reductive property dualism is not an idealized form of physicalism. The recognition that mental properties are real, distinguishable from physical properties, and irreducible requires greater nuance.

According to Hellman and Thompson, physicalism comprises two principles—ontological physicalism and physical determination.⁴⁴ If we take the principles at face value, physicalism maintains that (i) the universe comprises only physical things and (ii) physical facts determine all facts. These two principles reveal physicalism’s metaphysical and epistemological commitments, respectively. Even under a non-reductive theory of mind, such as property dualism, physicalism must aim to satisfy its metaphysical and epistemological theses. If a theory of mind does not satisfy these theses, then it is unclear on what grounds it can be considered a form of physicalism.

The major challenge to classifying non-reductive property dualism as a form of physicalism hinges on its central position. Property dualism is the view that mental properties are “distinct from and irreducible to physical properties, [...]”⁴⁵ If mental properties are distinct and irreducible, it would not be a leap to understand this to mean that because they are distinguishable from those things that are indisputably physical (e.g., oak trees, hydrogen atoms, and electrical currents), then they must be non-physical.

Mental properties do not neatly classify as physical things. Those things that are clearly physical are subject to repeat observation or detection. Moreover, when a physical thing is unobservable (e.g., gravity and universal inflation), it will still have a role in scientific theories and models.⁴⁶ Whether observable or unobservable, physical things are consistent with physical theory and scientific models. Lastly, a common physicalist

⁴⁴ Hellman and Thompson, 1977, 310-1.

⁴⁵ Lowe 2008, 1018.

⁴⁶ There are numerous ways that unobservable and undetectable things can be deemed valuable to physical theory. They can be justified by an explanation that would hold those unobservable physical things as logically consistent with physical theory and demonstrate predictive success. They can also be justified by entity realism or structural realism. For an overview of scientific realism’s take on unobservable physical entities, see Chakravartty 2017, 7-10.

position is that physical things can be reduced to fundamental physical things unless the physical thing in question is already fundamental.

Mental properties, such as tasting espresso, feeling pain, and holding a belief, are not observable and detectable. Rather, they are experienced directly. The experiencer may be able to report on the experience, but these reports will be vague because the intrinsic features of mental properties can be difficult to articulate.

The hardline physicalist position is that physical theories and models can function without mental properties because mental properties are causally inefficacious. Moreover, causally inefficacious properties would serve no purpose in a physical theory or scientific model. In turn, physicalists may go so far as to claim they are superfluous features of reality. Since physicalism is a scientific-minded worldview, mental properties could be excised or suppressed from physicalist theories on the grounds of parsimony. Put simply, since mental properties are subjective, unobservable, difficult to describe, and causally inefficacious, the hardline physicalist may view them as extraneous to their worldview.

However, the pervasiveness of mental properties and the central role they play in our understanding of reality is difficult to ignore. One way that physicalists may be able to address mental properties as non-extraneous features of reality is to treat them like higher-order physical things. The idea is that those higher-order physical things can theoretically be substituted with a lower-level item, event, or explanation when reduced adequately to or explained by bridge laws. However, the property dualist contends that mental properties cannot be replaced or substituted in this way. Lowe writes:

[i]t is hard to see how mental qualities such as the perceived redness of a rose or bitterness of a lemon could simply be physical properties of the brain or nervous system, [...]. There seems to be an unbridgeable “explanatory gap” between such sensory qualia and the neurophysiological activity that seems to be correlated with them, which prevents us from either simply identifying the former with the latter or even somehow “reducing” the former to the latter. The most that we seem to be able to say is that there is, as a matter of empirically confirmed fact, a causal correlation between such mental properties and certain physical properties of the brain – but one which it seems impossible for us to explain.⁴⁷

Lowes’s description suggests that the irreducibility of mental properties is partially due to the qualitative and private nature of experiencing mental properties. I consider these three aspects—private, qualitative, and irreducible—integral to mental properties. If any of the three aspects of mental properties is denied, then property dualism must be rejected.

Despite property dualism’s central position, the property dualist can identify as a physicalist. Vision notes that “[p]roponents acknowledge that there is at most one substance, but maintain that in addition to physical properties, that substance also has irreducible mental properties.”⁴⁸ The idea is that even though property dualists find mental properties distinct, one can still maintain physicalism because mental properties exist contingently upon physical substances. Vision notes that the consensus position among materialists is that “our reality bottoms out in [the] material world.”⁴⁹ Vision also notes that these physicalists hold that “[p]erhaps not everything is explicable; but to the extent that we have well-grounded explanations, they will contain at least traces of their

⁴⁷ Lowe 2008, 1019-20.

⁴⁸ Vision 2011, 25.

⁴⁹ Ibid.

physical origins.”⁵⁰ According to this position, property dualism can qualify as physicalism if mental properties exist contingently upon the physical.

To *exist contingently* obviously cannot mean reducible to the physical; instead, some other relationship must hold between mental properties and physical substance, or an essential aspect of mental properties is sacrificed (i.e., irreducibility), and property dualism must be rejected. Stoljar describes the completeness question as “[w]hat relation or relations must obtain between everything and the physical if physicalism is to be true?”⁵¹ He goes on to note:

[...] the completeness question holds fixed the issue of what it means for something to be physical, and asks instead what relation or relations obtain between everything and the physical if physicalism is true; in what sense, in other words, is physicalism a complete thesis, a thesis that applies to everything whatsoever.⁵²

Stoljar’s explanation of completeness shows that the property dualist must propose a relationship between mental properties and the rest of the physical world (i.e., physical substance and properties).⁵³ Therefore, without a well-defined relationship, property dualism is problematic on both metaphysical and epistemological grounds.

Regarding physicalism’s metaphysical thesis, the mental-physical relationship must tightly anchor mental properties to physical substances and properties. To accomplish this goal, the physicalist must categorize mental properties to distinguish them from physical properties but not ontologically distinct from physical substances. Furthermore, the mental-physical relationship must show that mental properties are

⁵⁰ Ibid, 25-6.

⁵¹ Stoljar 2021, 6.

⁵² Ibid.

⁵³ Schneider notes, “[i]t is fair to say that much of the literature on [non-reductive physicalism] concerns the viability of candidate psychophysical relations.” Schneider 2012b, 719.

contingent upon physical substances or states. If the mental-physical relationship cannot demonstrate that mental properties are contingent upon physical states, then the mental-physical relationship fails to satisfy the metaphysical thesis. If mental properties are not contingent, they would be on equal footing with physical substance. Mental properties that are distinct, irreducible, and not derivative of physical substance cannot satisfy the metaphysical thesis since all things would not be physical. A further consequence is that there would be no basis for the claim that reality bottoms out in the physical.

Regarding physicalism's metaphysical and epistemological theses, the mental-physical relationship must show that (i) the universe comprises only physical things and (ii) physical facts determine all facts. Should a mental-physical relationship fail to demonstrate these theses, distinct mental properties would have a legitimate claim to autonomy. Moreover, if neither the metaphysical nor epistemological theses are upheld, then physicalism would appear to fail in the face of mental realism. In this context, to be autonomous means that mental properties have at least partial independence from physical states, empowering them to behave in an extraphysical manner. Put differently, autonomous mental properties have the latitude to exist and operate in ways not clearly defined by physical theory. The existence of autonomous mental properties suggests that the totality of our mental lives cannot be fully captured even by a complete understanding of physical states.

Ultimately, the metaphysical and epistemological theses set a high bar for property dualism. Suppose mental realism is maintained and property dualism cannot satisfy physicalism's metaphysical and epistemological theses. In that case, there is little reason to accept the consequences of physicalism, including the undermining of mental

causal efficacy. Therefore, if property dualism is to be a form of physicalism, mental properties must be prescribed by a mental-physical relationship that appropriately limits the autonomy of mental properties.

Thus, the challenge for physical property dualism is demonstrating how distinctive mental properties are ontologically and epistemologically aligned with physicalism. From here, we explore the physicalists' best attempts at defining the relationship between the mental and physical.

Mental-Physical Relationships

Identity theory

Identity theories of mind aim to show that mental states are identical to physical states. Identity refers to something more robust than correlation, but it does not entail universal substitution. For instance, Smart writes:

[i]diomatically we do use 'She has a good mind' and 'She has a good brain' interchangeably but we would hardly say 'Her mind weighs fifty ounces'. Here I take identifying mind and brain as being a matter of identifying processes and perhaps states of the mind and brain.⁵⁴

Lewis makes a similar point when explaining identity theory. He writes:

[...] the identity theory does not imply that whatever is true of experiences as such is likewise true of neural states as such, nor conversely. [...] So it is pointless to exhibit various discrepancies between what is true of experiences as such and what is true of neural states as such. We can explain those discrepancies without denying psycho-physical identity and without admitting that it is somehow identity of a defective sort.⁵⁵

Lewis's explanation reveals that an identity relation need not entail complete sameness between mental and physical ascriptions. Instead, identity holds that mental properties

⁵⁴ Smart 2007, 1.

⁵⁵ Lewis 1966, 19.

and experiences have the same reference as physical states. Lewis writes, “experience-ascriptions have the same reference as certain neural-state-ascriptions: both alike refer to the neural states which are experiences. It does not say that these ascriptions have the same sense.”⁵⁶

Notably, type identity and token identity are two kinds of identity relationships. Type identity maintains that each mental state type (e.g., pain) is identified with a physical state type (e.g., c-fiber stimulation). Type identity theory is a reductive theory of mind because mental states, properties, and frameworks can be substituted with talk of physical states, properties, and theories. Jackson notes:

Kim [in *Mind and the Physical World*] rightly sees that a reduction should clarify how it is that the way things are as given in the terms of the reducing theory makes true the way things are as given in the terms of the reduced theory. A reduction should remove the ‘explanatory gap’, to borrow the famous phrase. He also rightly sees that a reduction should reduce in the sense of being ontologically parsimonious; it should make it clear that the entities and properties of the reduced theory are not entities and properties over and above those in the reducing theory.⁵⁷

When a mental state type is identified with a physical state type, the mental experience should be better understood because knowledge of the physical state will be well-defined thanks to the success of scientific investigations and physical theory. In theory, type identity could empower us to understand, initiate, alter, and end mental states we have identified with physical states. Put simply, type identity lessens the explanatory gap by claiming that mental experiences are not over and above physical states.⁵⁸

⁵⁶ Ibid.

⁵⁷ Jackson 2002, 644. I added the bracketed text for clarity.

⁵⁸ The explanatory gap is not fully closed because type identity does not resolve and explain every question about the relationship between mental and physical states. For example, the *why* of mental states remains in force.

Though type identity theory is an appealing way to think about the physical basis of mental experiences, it is problematic. The major challenge to type identity theory is the *multiple realizability* of mental states. Putnam describes the objection without using the term; he writes:

[the type identity theorist must] specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. [...] At the same time, it must not be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. [...] Thus if we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus (say “hungry”), but whose physical-chemical “correlate” is different in the two cases, [type identity] theory has collapsed.⁵⁹

Given that it is possible, if not true, that particular mental experience types (e.g., pain) can occur with different underpinning physical types (e.g., c-fiber stimulation or “x-fiber” stimulation), type identity fails.⁶⁰ Type physicalism falters when addressing multiple realizations, and its failure to account for multiple realizable mental properties is one of the major reasons that non-reductive physicalism has become a leading theory of mind.

Token identity theory does not fall victim to the multiple realizability argument and applies to non-reductive physicalism. Schneider writes:

[...], philosophers turned to [token identity] as a more plausible (and still physicalist) alternative to the type identity thesis. In the eyes of many, the phenomenon of multiple realizability motivated the conclusion that mental properties are not identical with physical properties.⁶¹

⁵⁹ Putnam 1967, 44. For clarity, I substituted brain state theory for type identity theory. Putnam is describing type identity theory.

⁶⁰ Multiple realizability, though pioneered by Putnam for the study of the mind, is now recognized in other disciplines as well. For instance, Batterman argues for multiple realizability in physics. See Batterman 2000.

⁶¹ Schneider 2012, 719-20.

Under token identity, the identity theorist does not identify types of physical states with types of mental states; instead, fine-grained identities are invoked. The token identity theorist maintains that each pain token (i.e., a particular instance of pain) is identical to a physical state token (i.e., the exact physical state that the pain experiencer is in when the pain token was experienced).

Token identity, a form of non-reductive physicalism, avoids the multiple realizability objection because the view maintains that each pain token is identical to a physical state token.⁶² Since each mental token is only identical to a particular physical token, multiple realizability is not a viable objection. However, this specificity also entails that physical properties and statements cannot subsume mental properties and statements. The fine-grained nature of token identity does not allow for the generality needed for reduction. In other words, because we are dealing with identical sets of tokens, we cannot draw generally applicable conclusions about all mental properties and physical states.

A challenge to token identity theory is its inability to provide strong evidence for the identity relationship its claiming. Since each mental state (or token) is proposed to be identical to a specific physical state (or token), confirming this identity through observation or experimentation is impossible. In other words, the uniqueness and non-repetitive nature of tokens hinder the ability to verify token identities.

Moreover, even when there is a strong correlation between types of mental tokens (e.g., pain) and types of physical tokens (e.g., variations on c-fibers firing), there is built-

⁶² Vision notes that token identity is “probably the most popular version of materialism on offer.” Vision 2011, 203.

in ambiguity. This ambiguity would include identifying irrelevant aspects of the token physical state with the mental property. For instance, M_1 at t_1 is identical to P_1 , and M_2 at t_2 is identical to P_2 , but even though P_1 and P_2 both have c-fibers firing, P_1 also has x-fibers firing, and P_2 has y-fibers firing. Token identity appears to have no way of confirming that M_1 and M_2 are token identical only with c-fibers firing. Token identity reveals that M_1 is token identical with both c-fibers and x-fibers, and M_2 is token identical with c-fibers and y-fibers. Even if we have good reason to believe x- and y-fibers have nothing to do with the experience of each token of pain, token identity does not offer a way to narrow it down to c-fibers.

The token identity theorist provides a mental-physical relationship that reveals little about the relationship between the mental and physical. Notably, it might seem that my objections to token identity might apply equally to emergence, the mental-physical relationship I advocate for in Chapter 3. Unlike token identity, which seeks precise one-to-one correspondence and struggles with ambiguity, emergence embraces the complexity and vagueness of the relationship between the mental and physical. Emergence views indeterminacy not as a flaw but as an integral aspect of higher-level phenomena arising from complex interactions. Therefore, the distinction is that the foundation of token identity is threatened by ambiguity. In contrast, emergence has ambiguity built-in and recognizes our complex reality's nuanced and non-deterministic nature. Rather than critique identity further, I provide a general objection to token identity and supervenience after outlining supervenience in the following subsection.

Supervenience

Supervenience is a mental-physical relationship that physicalist property dualists commonly hold. Vision notes:

[...] a leading reason why supervenience has thoroughly captivated the imaginations of writers in hotly contested areas is that it appears to supply a kind of dependence of apparently problematic areas on settled ones while remaining neutral on the further reduction to or identity of supervenients with their bases.⁶³

Supervenience is commonly held among physicalist property dualists because it is thought capable of constraining mental properties while avoiding some of the more serious troubles faced by stronger relationship types. Supervenience, in its various forms, recognizes mental properties while fixing those mental properties as dependent on physical properties or states. By codifying the dependency of mental properties on the physical, supervenience helps alleviate any concern that mental properties could be independent of the physical, which would be an untenable outcome for physicalism.⁶⁴

Though supervenience maintains the dependency of mental properties on physical states, the relationship does not claim that mental properties must be reduced to physical properties or states. Kim writes, “supervenient properties are in some substantive sense dependent on, or determined by, their subvenient base properties and yet, it is hoped, they are not reducible to them.”⁶⁵ Also, supervenience by itself is not a form of identity. Supervenience is more general and less restrictive than identity, but it is also core to identity. Vision writes, “supervenience subsumes identity. Identity following from

⁶³ Vision 2011, 36.

⁶⁴ Vision writes, “[...], for those who, like emergentists, regard further reduction as unpromising, but who hold that properties of a certain class are not free floating, supervenience has seemed to supply enough dependence to satisfy them.” Vision 2011, 37.

⁶⁵ Kim [1993] 2010, 275. Kim also presses the non-reducibility of supervenience in Kim 1990.

supervenience has been a common step in arguments for reductions.”⁶⁶ Mental-physical supervenience maintains that mental properties depend on subvenient physical states, but mental properties need not be identical to the subvenient physical states.

Since supervenience takes a weaker stand on the relationship between mental properties and physical properties, it is immune to some of the more damaging problems applicable to reductive and identity relationships. As a weaker position, supervenience has been the fallback mental-physical relationship for physicalist property dualists.

Supervenience is typically held as the baseline physicalist position. Montero writes:

[t]he supervenience of mental properties on fundamental physical properties is taken as a necessary condition for physicalism because the failure of such supervenience is thought to render mental properties nonphysical; and if there is something nonphysical, then physicalism, which holds that everything is physical, is false. Although many of those engaged in the debate over the mind-body problem object to various aspects of these standard antiphysicalist arguments, most, if not all, agree that if physicalism is true, then mental properties must supervene on fundamental physical properties.⁶⁷

Typically speaking, Supervenience is a relationship type where mental properties are in a reflexive, transitive, and non-symmetrical relationship with physical properties or states.⁶⁸ Supervenience is reflexive because any change in mental properties or states requires a change in the underlying physical property or state. McLaughlin and Bennet note that “there cannot be A-differences without B-differences” is supervenience in slogan form.⁶⁹ Essentially, the alteration or instantiation of a mental property (i.e., A-differences) requires a change in the physical subvenient base (i.e., B-differences).

⁶⁶ Vision 2011, 37.

⁶⁷ Montero 2013, 93.

⁶⁸ Kim 1984, 166.

⁶⁹ McLaughlin and Bennet 2018, 3.

Though supervenience relationships are reflexive, they are also non-symmetrical. The relationship is non-symmetrical because physical states can change without necessarily altering mental states. McLaughlin and Bennet write, “while the mental may supervene on the physical, the physical does not supervene on the mental. There can be physical differences without mental differences.”⁷⁰ According to the non-symmetry of supervenience, a mental property, such as pain, can perdure while the subvenient base changes. For example, one can imagine the nervous system being in non-identical states at t_1 and t_2 while the same mental property is experienced at t_1 and t_2 . The non-symmetry qualification of supervenience can accommodate multiple realizations because two physical states can instantiate or maintain the same mental property.⁷¹

The transitivity of supervenient properties tells us that if mental property M_1 supervenes on brain state B_1 and B_1 supervenes on a physical state P_1 , then M_1 transitively supervenes on P_1 . The transitivity aspect of supervenience means that any instantiation or alteration of higher-level properties depends on alterations among all subvenient states.

Lastly, supervenience is generally found to be a non-reductive mental-physical relationship. While the supervenience relationship establishes a connection between several properties and states, it is not the case that supervenient properties reduce to subvenient states. Kim writes:

[f]irst, supervenience is to be a relation of *dependence*: that which is supervenient is dependent on that on which it supervenes. Second, it is to

⁷⁰ McLaughlin and Bennet 2018, 9-10.

⁷¹ McLaughlin and Bennet provide the following example: “radically physically different things—a washing machine and a paper bag, say—can be mentally just alike in virtue of lacking mental properties altogether.” McLaughlin and Bennet 2018, 10.

be a *nonreductive* relation: supervenient dependency is not to entail the reducibility of the supervenient to its subvenient base.⁷²

With the parameters of supervenience laid out, I turn to an example to illustrate the entailments of supervenience's reflexivity, transitivity, non-symmetry, and irreducibility.

Reflexivity tells us that when M_2 succeeds mental property M_1 , there is a difference in the physical subvenient base so that we can assume M_1 to M_2 means that P_1 becomes P_2 . Supervenience also maintains non-symmetry, so it follows that the subvenient base can change without altering mental properties. So, even though M_1 to M_2 requires P_1 and P_2 , M_2 could be maintained by P_3 and P_4 . Though supervenience allows for this scenario, there must be limits to the extent of physical change before it affects the otherwise perduring mental property.

Consider, for example, a person experiencing a lingering headache over a few hours. We will call the initial experience of headache pain H at t_1 . If supervenience is obtained, then H at t_1 supervenes the physical state of the headache sufferer. Presumably, the headache sufferer's neural state would be the nearest physical state subvening the headache, so H supervenes on N at t_1 . The headache lasts throughout the day, so H is also experienced at t_2 . However, the neural state at t_2 is not identical to the neural state at t_1 . At t_2 , there is variation in the neurons firing as the person transitions from one task to the next, such as reading the newspaper, riding their bicycle, and composing an email. The refined picture is that H supervenes N_1 at t_1 and H on N_2 at t_2 . Supervenience can easily account for H supervening on different N states because of its non-symmetry feature.

⁷² Kim 1990, 8.

One question to ask is whether H perdures or if each succeeding N state instantiates a mental property that is like H (i.e., an H token). Either we have (1) H supervenes on N_1 at t_1 , and H supervenes on N_2 at t_2 , or (2) H_1 supervenes on N_1 at t_1 and H_2 supervenes on N_2 at t_2 . If (1) H perdures while supervening on different physical states (i.e., N states), then we must ask what explains the discontinuation of H among succeeding N states. Option (1) requires a proposal for how H perdures so that the theory explains why H eventually ceases to supervene on succeeding N states. I see two viable proposals.

Proposal (A) suggests that mental property H perdures so long as the neural states fall within a certain acceptable range. Since H can be multiply realized, it can supervene various N states. So long as the N state stays within a certain range of fluctuation, H will perdure. If the N states surpass the acceptable range, H will cease. I will return to this proposal after outlining Proposal (B).

Proposal (B) is a more specific approach that suggests H perdures because it supervenes an essential aspect within the global neural state of N. We know that the neural state changes from moment to moment, but there must be an unchanged essential (or local) aspect explaining H perduring. This approach is meant to establish a direct connection between the mental state and the physical state it supervenes on. For instance, Proposal (B) tells us that H supervenes on N_1 at t_1 and N_2 at t_2 but not on N_3 at t_3 . Proposal (B) can explain this by pointing to the firing of c-fibers, which I refer to as C. Thus, when we have N_1 and N_2 , we also have C, but those c-fibers did not fire in N_3 . In essence, Proposal (B) aligns itself more closely with identity theory because it explicitly

ties one state's existence to another. Proposal (B) establishes a one-to-one correspondence between the mental state and a particular physical state.

Proposal (A) raises questions about the range of physical states that can occur before H discontinues. Still, the supervenience physicalist is not compelled to determine the range or threshold. Supervenience is a non-technical and non-reductive relationship that does not get into specifics by design. This inability of Proposal (A) to define the "acceptable range" for any given mental state appears as a theoretical shortcoming within supervenience. While proponents of supervenience might accept the non-specific nature of this relationship, the problem lies in its assertion of a one-to-one dependence between mental and physical states without explaining the intricacies of this connection.

In contrast, a strong emergentist stance also does not offer a precise mapping between an emergent phenomenon (e.g., H) and its base components. However, the rationale behind emergence differs significantly from supervenience. My position on emergence recognizes reality to be nonlinear, complex, layered, and having built-in indeterminacy. So, while emergence also does not provide a technical mapping, it avoids the pitfalls of overdelivering, leading to the dilemma of certainty and ambiguity inherent in supervenience. In other words, by acknowledging reality's inherent complexity, emergence provides a more coherent framework for understanding higher-level phenomena without requiring strict one-to-one technical mapping. I provide greater detail on emergence in Chapter 3.

Proposal (B), on the other hand, attempts to lessen the explanatory gap by tying H to C (i.e., c-fibers firing), a particular aspect of N. While this proposal may appear like a refined version of supervenience, it aligns closely with identity theory and requires

relitigating the problem of multiple realizations. Proposal (B) asserts that H is definitively linked to C. This definitive link is problematic because it is plausible for H to be instantiated without C. Thus, the multiple realizability of H suggests that Proposal (B) oversimplifies the complex nature of mental states and how they are realized. Proposal (B) narrows its scope to resolve the ambiguity of Proposal (A), but by doing so, it inadvertently aligns with identity theory and a reductive framework. The problems associated with identity theory and reductive physicalism are not insignificant. Since supervenience is meant to avoid the pitfalls associated with identity and reductionism, it follows that Proposal (B) is not viable.

Both proposals are problematic. Proposal (A) reveals little about the relationship between mental and physical properties and requires a dogmatic approach, while Proposal (B) inadvertently aligns with identity theory and thereby inherits its problems. I contend that the issues in Proposals (A) and (B) reveal a theoretical flaw in supervenience. Supervenience insists that a definitive dependence relationship exists between one type of state and another more fundamental state. Still, the relationship must be ambiguous to avoid succumbing to the same problems as identity theory. I find the certainty of supervenience and the presence of ambiguity contradictory.

Since Proposal (B) is akin to identity theory, it seems that Option (1) must follow Proposal (A). However, endorsing Proposal (A) requires a dogmatic approach that accepts supervenience despite the dilemma of certainty juxtaposed with ambiguity. With the shortcomings of Option (1) outlined, let us explore Option (2).

Under Option (2), token instances of H instantiate in each moment so that we have H_1 supervening on N_1 at t_1 and H_2 supervening on N_2 at t_2 . Because of non-

symmetry, reflexivity, and mental realism, different neural states can instantiate a similar mental token. For example, H_1 supervening on N_1 at t_1 , H_2 supervening on N_2 at t_2 , and H_3 supervening on N_3 at t_3 . Option (2) aligns with the parameters of supervenience and its slogan, but it extends beyond simple supervenience into token identity. Therefore, Option (2) is subject to the same critiques as token identity.

Despite the precision of every token being tied to a token physical state, we cannot verify this connection. Since each instance of H is a unique token, empirically verifying that that token supervenes a particular mental state is problematic. The uniqueness and non-repetitive nature of tokens hinder the ability to verify that each instance of H supervenes on each N state.⁷³

Along similar lines, ambiguity makes directly tying H_1 to N_1 problematic as well. As noted earlier when discussing token identity, the built-in ambiguity of this relationship means that H_1 could also supervene irrelevant aspects of the underlying state. This means that, even if the supervenience advocate strongly believes that H_1 supervenes on N_1 , other substates will co-occur with N_1 that are also positioned to be supervenient bases to H_1 .

Option (2) fails to concretely establish the specificity of the supervenience relationship promised. This leaves us with a definite, yet unverifiable and uncertain, supervenience relationship between mental tokens and physical states. Ultimately, this highlights the key limitation of Option (2) and token identity theories.

⁷³ For instance, consider a real-time experiment conducted in an fMRI machine where a subject is poked with a needle repeatedly and must report on the pain. The indeterminate and qualitative nature of each instance of P (i.e., needle poke pain) introduces several uncertainties into the reporting about P by the subject in the fMRI machine.

Typically, the supervenience relationship does not explain mental property instantiations. Otherwise, as shown above, supervenience becomes a form of identity theory. In other words, supervenience does not propose *how* or *why* particular mental properties are present in certain physical circumstances. Rather, the relationship leans into physicalist dogma by requiring mental properties to depend on physical states. However, supervenience does not prove that dependence; it merely posits it. Since mental properties depend on physical states, it is natural to infer that they are metaphysically derivative of the physical properties responsible for the physical state. Ultimately, Options (1) and (2) reveal that supervenience reveals nothing about the *how* and *why* of mental properties. Supervenience merely posits that mental properties supervene physical states, and its rationale is physicalist dogma and mental-physical correlations. I do not find this a strong reason to be a supervenience physicalist.

On those demerits alone, I find supervenience to be an inadequate mental-physical relationship, but there are other issues. For instance, even as a non-technical view, supervenience must support physicalism's ontological and epistemological theses. As a non-technical view, however, there is no way to determine how supervenience accounts for mental properties in such a way that satisfies both the metaphysical and epistemological theses of physicalism.

To satisfy the metaphysical thesis, supervenience must demonstrate that mental properties are, in fact, physical. However, supervenience only tells us that a distinct and irreducible mental property supervenes the physical. On its own, supervenience does not demonstrate that mental properties are metaphysically physical. Supervenience merely shows that mental properties and physical states are linked.

To satisfy the epistemological thesis, supervenience must show that physical facts determine mental facts. Supervenience does not explain *how* physical facts determine mental facts. In its most basic state, supervenience only asserts that a difference in a mental property must mean a difference in a physical property (i.e., no A-differences without B-differences).

Minimal supervenience needs to be supplemented with a fine-grained mental-physical relationship (e.g., an identity relationship) if it is to satisfy the epistemological thesis. The issue is that supervenience takes no stand on the technical relationship between the mental and physical other than the metaphysical claim that the mental is derivative of the physical. Based on the implications of supervenience, I find it difficult to grant that supervenience satisfies the metaphysical or epistemological theses.

On the Origin of Mental Properties

Supervenience is silent on the origin of mental properties, their instantiation, and what accounts for perduring or repeating tokens across different physical states. When we take these shortcomings seriously, nothing negates the possibility of mental and physical states existing coextensively so that neither is dependent on the other. Indeed, this is the point Churchland makes when he writes:

[a] property dualist is not absolutely bound to insist on both claims. He could let go of the thesis of evolutionary emergence, and claim that mental properties are fundamental properties of reality, properties that have been here since the universe's inception, properties on a par with such properties as length, mass, charge, time, and other fundamental properties.⁷⁴

⁷⁴ Churchland [1984] 2013, 20.

The anti-physicalist has a basis for such an argument, pointing to mental properties repeating or perduring despite physical state changes. If one puts aside physicalist dogma, then a new perspective might suggest the repeating or perduring mental properties are independent of physical states as suggested by their continuance despite physical state changes. Recognizing the non-technical relationship of supervenience and the multiple realizability of mental properties, it follows that physical facts do not determine all facts. How can such a claim be made when the supervenience physicalist takes no stand on the nuanced relationship between repeating or perduring mental properties?

Without addressing the origin of mental properties, supervenience physicalism cannot satisfy the metaphysical and epistemological theses. The issue is that supervenience is not a productive explanation for the relationship between mental properties and physical states. It is merely a limiting rule for mental properties based on physicalist expectations. So, even if mental properties supervene physical states, there is insufficient reason to find this evidence of physicalism.

It is generally held that all versions of non-reductive physicalism maintain supervenience.⁷⁵ Notably, supervenience is posited as the second premise of the Causal Exclusion Argument. The importance of supervenience to non-reductive physicalism implies that any non-reductive theory that champions mental causal efficacy must uphold supervenience or provide sufficient reason to deny it or modify it to remain consistent. I believe there are reasons to doubt the necessity of supervenience, and there is potential

⁷⁵ Moore and Campbell writes, “[...] all non-reductive physicalists accept some form of supervenience (otherwise mental properties would not be *determined* by physical properties and we would have full-blown dualism rather than non-reductive physicalism) [...]” Moore and Campbell 2010, 437.

for it to be replaced by other nuanced views. I address supervenience again in my discussion of emergence in Chapters 3 and 4.

Reflecting on Mental-Physical Relationships

In the above sections, I addressed how mental realism and physicalism lead to an uneasy connection between irreducible and distinct mental properties and that which is neatly defined by physical theory. In those subsections, I reveal that non-reductive physicalism is an unstable position. The following look into mental causes will reveal even more instability.

If one accepts the Causal Exclusion Argument without qualification, it follows that non-reductive mental properties cannot be causally efficacious under physicalism. Moore and Campbell claim that “the point of adopting non-reductive physicalism is to preserve the significance and autonomy of the mental [...]”⁷⁶ They ask, “if mental properties lack causal efficacy exactly how much autonomy and significance could they really have?”⁷⁷ I share Moore’s and Campbell’s intuition about the status of mental properties and wonder what mental properties can amount to under physicalism. Are mental properties merely concepts, or are they significant enough to surpass the conceptual threshold? Respecting physicalism—and monism in general—requires that mental properties cannot be built up to the point that they become a secondary substance. Substance dualism is a rejection of physicalism, but it is also a hasty move for the mental realist sympathetic to mental causal efficacy because it would result in the same old metaphysical problems.

⁷⁶ Moore and Campbell 2010, 435.

⁷⁷ Ibid.

Kim suggests that “[m]ental properties are real properties of objects and events; they are not merely useful aids in making predictions or fictitious manners of speech.”⁷⁸ Kim’s definition prompts us to ask, what precisely about mental properties makes them real? There are several options worth considering. Either mental properties are real because they are (i) reducible to the physical but explanatorily valuable in their own right, (ii) irreducible to complex physical states and therefore ineliminable, or (iii) causally efficacious. I have already addressed (i) reducibility and do not see it as a viable option. I find (ii) and (iii) to be viable options that mental realists will want to hold simultaneously.

The ideal proof for mental realism would be (iii) causal efficacy. Alexander’s Dictum tells us that “to be is to be causal,” and though the Dictum is not a law of nature or logic, it is intuitive, and there would be little motivation to argue against the reality of causally efficacious things.⁷⁹ Take, for instance, Fodor’s hyperbolic but sincere comment about causally inefficacious mental properties; he writes:

I’m not really convinced that it matters very much whether the mental is physical; still less that it matters very much whether we can prove that it is. Whereas, if 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 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.*⁸⁰

Kim also writes, “[t]he reality of the mental is closely tied to the possibility of mental causation, and anyone who takes a realist attitude toward the mental must be prepared

⁷⁸ Kim 1993b, 198.

⁷⁹ Cargile prefers “To exist is to have causal powers.” Cargile 2003, 143. See Cargile 2003 for an overview of Alexander’s Dictum.

⁸⁰ Fodor 1989, 77. Italics added for emphasis.

with an account of how mental causation is possible.”⁸¹ These quotes from Fodor and Kim note the value of causal efficacy and the desire for mental properties to be causally efficacious. Yet, it seems that the Causal Exclusion Argument championed by Kim requires the property dualist to either accept epiphenomenalism or a lesser form of causal relevancy through a tight association between mental properties and physical processes (e.g., functional reductions).

In contrast, if a mental realist were to opt only for option (ii) as their proof for mental realism, then they fall into the trap of epiphenomenalism. Since epiphenomenal properties cannot affect the world, we have reason to question their nature and, ultimately, their existence. As Samuel Alexander said:

[...] we are entitled summarily to dismiss the conception that mind is but an inert accompaniment of neural process, a kind of aura which surrounds that process but plays no effective part of its own: the doctrine that mind is an epiphenomenon of nervous process, which nervous process would continue to work equally well if mind were absent. The doctrine is not simply to be rejected because it supposes something to exist in nature which has nothing to do, no purpose to serve, a species of noblesse which depends on the work of its inferiors, but is kept for show and might as well, and undoubtedly would in time be abolished.⁸²

Mental properties that are not causally efficacious serve no functional role. Properties that serve no function are inherently mysterious and easy targets to be classified as mistaken ideas. Epiphenomenalism is an odd stance that imperils mental realism.

Interestingly, epiphenomenalism appears unable to account for how mental properties fit within physicalism. Instead, mental properties are treated as metaphysical danglers that cannot be ignored but should not be given too much attention either.

⁸¹ Kim 1993a, 285-6.

⁸² Alexander 1927, 8.

Epiphenomenalism is an equally uncomfortable position for the mental realist as it is for the physicalist. Regarding the latter, epiphenomenal mental properties cannot satisfy the metaphysical thesis. For instance, it is unclear how something ineffectual can play any kind of role in physical theory, even if it is merely an aftereffect. Notably, many physicalists, including Kim, object to epiphenomenalism.⁸³

The preferred option for most non-reductive physicalists is to seek causal relevancy. The physicalist attempts to save mental properties from causal irrelevance by closely associating the perceived function with correlating physical processes. The idea is that if mental properties and their causal roles can *piggyback* on the causal efficacy of physical properties, mental properties can be *saved* from epiphenomenalism.

Given that identity theories are problematic and that supervenience tells us very little about mental properties, let us turn to Kim's attempt at resolving the tension between the mental and physical. Kim finds that "any physicalist who wants to save the causal efficacy of mentality must be prepared to embrace physical reductionism."⁸⁴

Functional Reductions

The type of *saving reduction* that Kim proposes is a functional reduction. Kim finds that mental properties that seem to have causal roles, such as beliefs, desires, and intentions, are candidates to be functionally reduced.⁸⁵ Kim explains that there are three steps to functionally reducing a mental property. First, the mental property in question must be *functionalized* or reinterpreted, focusing on what causal role the mental property

⁸³ See Moore and Campbell 2010, *passim*, for a discussion of Kim's objection to epiphenomenalism as well as their criticism of Kim's proposed solution.

⁸⁴ Kim 2002a, 640. Kim also notes that "if we want robust mental causation, we had better be prepared to taken reductionism seriously, whether we like it or not." Kim 2005, 22.

⁸⁵ Kim, 2005, 161-68

is supposed to perform. Second, through physical examination and scientific discovery, the physical realizers of that causal role must be identified in the population of interest (e.g., humans). Third, the causal work at the reductive physical level must be explained in terms of how the subvenient base or underlying physical processes perform the causal work in question.⁸⁶

Kim's often-used example of a functional reduction is genes being functionally reduced to DNA molecules.⁸⁷ Genes can be functionalized because they have an identifiable causal role: the transmission and inheritance of genetic factors. The study of molecular biology reveals that DNA molecules demonstrate and explain how inheritable traits are passed to offspring. As such, genes can be functionally reduced to their physical realizer, DNA molecules.⁸⁸

Functional reductions are attractive primarily because they avoid one aspect of multiple realizability. By noting that a functional reduction must identify the "population of interest," Kim and other advocates of functional reductions acknowledge that a higher-order property might be functionally reducible to different subvenient bases. For example, though human genes functionally reduce to DNA molecules, Keplerians (i.e., an extraterrestrial species) may have a different physical realizer of genetic inheritance. Therefore, their genes would functionally reduce to a different physical realizer. Webbiens (i.e., another extraterrestrial species) also have inheritable traits, but their physical realizers of genetic inheritance have yet to be discovered. In the case of the

⁸⁶ Kim, 2005, 101-2, 164.

⁸⁷ Kim 2005, writes "[c]onsider the gene and how it has been reduced in molecular biology. The concept of a gene is the concept of a mechanism in an organism that encodes and transmits genetic information. [...] In short, the concept of a gene is a 'functional' concept, and the property of being a gene is a functional property defined by a 'job description' (163).

⁸⁸ Kim overviews functionally reducing genes to DNA molecules in Kim 2005, 163.

Webbians, their genes are potentially functionally reducible but cannot be functionally reduced until the physical realizers are discovered.⁸⁹

Kim believes that functionally reducing a mental property can save mental causal efficacy by equating mental properties' functions with subvenient causal processes. Kim's functional reductions espouse a form of causal inheritance, where mental properties have no causal powers but are tightly associated with their physical realizers.⁹⁰ For the physicalist, a functional reduction allows one to recognize the legitimacy of mental properties while respecting the Causal Exclusion Argument. Kim attempts to codify mental causal efficacy through quausation, where M_1 can be functionally reduced to P_1 so that if P_1 causes P_2 , then M_1 qua P_1 causes P_2 .⁹¹

Notably, Kim's functional reductions are not true reductions. The way that functional reductions are reductive is not akin to ontological or theoretical reductions. Instead, functional reductions work by determining the function of a higher-order property and then associating that function with a lower-order physical process. In ontological reductions, a mental property is identified with lower-order physical properties. Furthermore, in theory reductions, one theory is reduced to another through bridge laws that link the terms of one theory to those of the other. Whereas functional reductions do not explain how a mental theory translates to a physical theory through bridge laws (i.e., theory reduction).

⁸⁹ Kim concisely overviews how multiple realizability does not negatively impact functional reductions. Kim 2005, 163-4.

⁹⁰ Kim 1998, 54.

⁹¹ See Horgan 1989.

Functional reductions do not provide any detailed mapping between mental and physical properties. Rather, functional reductions merely posit that the perceived causal role of any given mental property can be equated to the subvening physical process believed to play a role in the events in question. Since functional reductions are neither ontological nor theoretical reductions, functional reductions do not run afoul of my earlier claim that mental properties have three fundamental aspects—private, qualitative, and *irreducible*.

Kim is forthright that not all mental properties can be functionally reduced. According to him, mental properties marked by their intrinsic qualitative nature (i.e., qualia) cannot be functionally reduced. For example, pain, which one might presume functionally leads to retreating from further injury or initiating a self-care plan, cannot be functionally reduced. Even if functional explanations for pain may be intuitive, Kim writes, “[p]ain may be associated with causal tasks, but these tasks do not define or constitute pain. Pain as a sensory quale is not a functional property. In general, qualia are not functional properties.”⁹² Pain and qualia in general are not functionally reducible because the experience of a quale cannot be captured by physical processes alone and go beyond any functional role.

Kim is a physicalist that recognizes the challenge of mental realism. Functional reductions aim to explain the reality of mental properties in a way that abides by physicalism while being metaphysically sound. To accomplish this, Kim knows that he must reconcile the appearance of mental causal efficacy. By positing that mental properties can be functionalized, Kim believes he saves mental properties from

⁹² Kim 2005, 169.

epiphenomenalism while respecting physicalism's metaphysical and epistemological theses. However, Kim's theory of mind only complicates the mental-physical ontology.

In his theory of mind, Kim gives us two classes of mental properties: those that can and those that cannot be functionally reduced. These two classes of mental properties set up a troubling dynamic. If the primary motivation for functionally reducing mental properties is to demonstrate how they can be causally relevant so that physicalism can prevail, then the second class of property becomes a quagmire for physicalists. As an MR Physicalist, Kim needs mental properties to fit neatly within physicalism. Specifically, his metaphysics requires that mental properties be causally relevant without contradicting the Causal Exclusion Argument.

The dilemma is that Kim may provide a reason to find functionally reducible mental properties to be real, but he simultaneously undermines the reality of irreducible qualia. In other words, if Alexander's Dictum is the metaphysical means to determine the reality of a property, and functional reductions are the way to find mental properties causally relevant, then Kim's method entails that qualia have no metaphysical claim to existence since they cannot be functionally reduced.

Most mental realists, including Kim, should reject any conclusion about the unreality of qualia. Qualia are among the most pervasive mental properties, and they are precisely the type of mental properties that made mental realism a commonly held position among philosophers. Notably, Kim recognizes the problem with qualia remains, leading him to describe his view as physicalism manqué. He writes:

[t]he position is, as we might say, a slightly defective physicalism—physicalism manqué but not by much. I believe that this is as much physicalism as we can have, and that there is no credible alternative to

physicalism as a general worldview. Physicalism is not the whole truth, but it is the truth near enough, and near enough should be good enough.⁹³

Ultimately, physicalism manqué represents Kim's best efforts to stabilize physicalism with mental realism while preserving the Causal Exclusion Argument.

I find functional reductions to be overly problematic. For instance, according to proponents of functional reductions, for a mental property M to be causally relevant, it must be linked to a specific physical process P. This only works, however, when M has a functional aspect that can be associated with a physical process. Despite Kim's best efforts, his approach does not resolve the problems introduced by qualia. Given that qualia are not functionally reducible, they highlight a tension in Kim's theory of mind and with physicalism more generally. While physicalism manqué purports to be a worldview based on our best evidence, it marginalizes those aspects of mental life that it cannot incorporate. This tendency pushes physicalism ever closer to a dogma requiring conformity than a comprehensive and inclusive worldview informed by the relevant evidence. Thus, until all mental properties, including qualia, can be properly integrated, this approach to physicalism remains problematic. Put simply, the strategy of discounting troublesome mental aspects calls into question the logic of adopting functional reductions in the first place.

In addition to my objection to the logic and stance of functional reductions, I find the view is systemically problematic. It is metaphysically insufficient to merely align M's defined role with P's causal mechanics. A compelling metaphysical justification is still required. Without a substantive mental-physical relationship posited, functional

⁹³ Kim 2005, 174.

reductions risk either negating mental realism by portraying mental properties as non-causal conceptual placeholders or destabilizing the core tenets of physicalism by upholding mental realism without an adequate metaphysical anchoring of mental properties to physical states.

It seems that Kim's functional reductions require identity as the mental-physical relationship that allows for causal relevancy and metaphysical grounding in the physical. To elaborate, *M* can be functionally reduced when we define its causal role. I will refer to *M*'s functional definition (i.e., causal role) as *H*. It follows that *H* is the functional property of *M*. So, under Kim's model, *M* identifies with *H*, and *H* identifies with *P*. Through the transitivity of identity, it follows that *M* and *P* are identical. Kim writes:

[t]his grounds the identification of *M* with *P*. *M* is the property of having some property that meets specification *H*, and *P* is the property that meets *H*. So *M* is the property of having *P*. But in general the property of having property *Q*= property *Q*. It follows then that *M* is *P*.⁹⁴

Thus, according to Kim's functional reductions, forging an identity between *M* and *P* entails that *M*'s causal powers *are* *P*'s causal powers.

There is an oddness in this model that Kim recognizes. He writes, "how could *M* be identical with *P*? Isn't it incoherent to think that a property could be both first order and second order, both extrinsic/relational and intrinsic, both a role and its occupier?"⁹⁵ Moreover, the multiple realizability of functionally reducible mental properties leads Kim to posit a disjunctive model of functional reductions. For example, if Smith can realize functional property *M* with functional role *H* in either state *P*₁ or *P*₂, then *M* is multiply realized. Since functional reductions utilize identity relationships, we know that *M* = *H*

⁹⁴ Kim 1998, 98-9.

⁹⁵ Kim 1998, 103.

and $H = P_1$ or P_2 . Because of the transitivity of identity, it also follows that $M = P_1$ or P_2 . Thus, M identifies with multiple P states and does not coalesce in the way one would expect of identity. The disjunctive model has systemic issues because it treats identity across innumerable physical variables.

Based on the above concerns of the identity model of functional reductions, Kim explores a model of functional reductions where mental properties are understood as concepts. This view aligns with his oft-used example of functionally reducing genes to DNA. Genes are conceptual categorizations of inheritable traits, and molecular biology finds that genes function through DNA molecules transferring genetic information. However, it is difficult to see how mere concepts can have causal or even quausal powers attributed to them.

Claiming mental properties are mere concepts because of the disjunctive model serves as a rebuke of Kim's description of mental properties, namely that they are "[...] real properties of objects and events; they are not merely useful aids in making predictions or fictitious manners of speech."⁹⁶ The disjunctive model revealing the need for mental properties to be concepts reveals how far afield the project has gone from its original focus of finding agreement between mental realism and physicalism.

It is troubling to think that functional reductions that rely on conceptual mental properties reopen the door to anti-realist arguments about mental properties. The argument is that if mental properties are merely concepts that can be reduced to empirical operators, such as neural pathways, chemical imbalances, and so on, it stands to reason that the concept adds very little, if anything, beyond those operators and, therefore, may

⁹⁶ Kim 1993b, 198.

be eliminable. In essence, under the conceptual model of functional reductions, nothing would be missed in a complete physical analysis of reality other than the *fictitious manners of speech* serving as conceptual placeholders for empirical operators. Mental realists will reject this conclusion, which means that mental properties as concepts must also be rejected. Ultimately, Kim has expressed his own dilemma over whether he can endorse the conceptual model of functional reductions.⁹⁷

The functional model runs into difficulty because one must either endorse the disjunctive model or the conceptual model. The disjunctive model is ontologically unstable because it relies on an identity relationship to grant causal relevancy. On the other hand, the conceptual model is ontologically impoverished because there's no well-defined reason to be a mental realist, yet mental realism prevails precisely because mental properties are ubiquitous, pervasive, and essential to understanding reality.

I find both the disjunctive identity and conceptual models overly problematic, and I see enough reason to question the suitability of functional reductions to describe mental causal efficacy. Both the disjunctive and conceptual models reveal that functionally reducing mental properties to physical properties is more closely akin to equivocation than to an ontologically and epistemologically satisfactory mental-physical relationship. A functional reduction may be able to identify a conceptual causal task that lies between M and P, but it does not provide a robust view of M nor its causal efficacy.

Put simply, the physicalism manqué championed by Kim provides no stance on the status of irreducible mental properties (e.g., qualia), and it provides a troubling stance

⁹⁷ Kim writes, "moreover, I have been torn between the conceptual approach recommended in my book and the disjunction approach also discussed in the book." Kim 2002b, 678.

on the status of functionally reducible mental properties. The point of functional reductions was to shore up the place of mental properties under physicalism, but physicalism *manqué* reveals considerable disorder in the theory. Additionally, the causal relevancy that physicalism *manqué* posits is not an attempt to understand mental properties and is not equivalent to mental causal efficacy. In other words, quausion does not preserve mental causation; it merely gives a different name to reductive physical causation. Ultimately, functional reductions and causal relevancy are conceptual equivocations that draw upon correlations between properties with affinitive states.

Conclusion

Let's assume that mental realism, as Kim defines it, is true. One way that the above mental-physical relationships failed is that they did not provide mental properties with the same credence as physical properties. Instead, these inquiries often start from the assumption that mental properties are, at best, a lesser type of property. The physicalist's preconceived notion about mental properties inadvertently undermines a full exploration of mental phenomena. Any physicalist who accepts the reality of mental properties should begin their exploration by affording them the same initial considerations as physical properties. In other words, assuming mental properties to be real and genuine provides a more appropriate starting point to examine whether non-reductive mental properties can coexist within the physicalist's framework, specifically by fulfilling the metaphysical and epistemological requirements of physicalism. The status of mental properties must be given the benefit of the doubt if mental realism and physicalism are to

be seriously considered. Undermining mental properties at the onset evades the complexities inherent in reconciling non-reductive mental properties with physicalism.

The perspective that mental properties are lesser properties than physical properties or mere concepts is often held as the starting point of any physicalist exploration claiming to take mental realism seriously. The stance is due to a firm commitment to stringent physicalism and an unwillingness to adjust this stance to accommodate mental properties. Yet, this standpoint compromises the premise of mental realism and the grounds for considering non-reductive physicalism as a worthwhile theory of mind.

Attempts to identify mental properties are one such path to making mental properties subsidiary. Supervenience and functional reductions start from the same assumption. Type identity falls short because of its inability to account for multiple realizations. The primary shortcoming of token identity is its failure to provide a productive explanation for the connection between token mental properties and particular mental states. Supervenience remains consistent, but it is vague and does not take a stand on the origin of mental properties. Lastly, functional reductions yield a convoluted ontology.

There are several ways that the reality of a property might be posited, including through empirical evidence, coherence with a worldview, agreement among observers, predictability, and causal efficacy. I believe the key to determining the reality of mental properties and determining their coherence with physicalism is to take a stand on the origin of mental properties. If a physical origin can be established, there will be inroads

to determining the reality of mental properties through one of the methods mentioned above.

If property dualism is true, then there will be an affinity between mental properties and physical processes. However, quausation is not equivalent to mental causal efficacy because the view does not find mental properties to be a locus of causation. Under quausation, the physicalist attributes all the causal work to physical processes, meaning mental properties are only causally *relevant*.

The assertion that mental properties exist contingently on physical states and that physical facts determine mental facts are underdetermined. The irreducible and distinct nature of mental properties complicates satisfying the metaphysical and epistemological theses of physicalism. By positing the origin of mental properties, there may be a way to bridge the gap between mental properties and physical facts.

There are two broad options for the origin of mental properties. Either mental properties are fundamental (or proto-fundamental) or generated. Fundamental or proto-fundamental properties are taken seriously in the literature, but I will bypass addressing this option because it is further afield from traditional physicalism than I would like to tread. For instance, fundamental mental properties raise entirely new complications for the physicalists' assertions that the universe comprises only physical things and that physical facts determine mental facts. Conversely, if mental properties are generated by physical activity, there is a clearer path toward addressing physicalism's metaphysical and epistemological theses.

The non-reductive property dualist physicalist who takes a stand on the origin of mental properties must demonstrate that the generation of mental properties does not

contradict physicalism. In Chapter 3, I champion emergence as a viable explanation for generating mental properties and the best mental-physical relationship to reconcile mental realism and physicalism.

CHAPTER 3

EMERGENT MENTAL PROPERTIES

Chapter 3 focuses on emergence as a theory of mind with advantages not possessed by the mental-physical relationships addressed in Chapter 2. I argue that emergence presents an explanatory view of mental life largely aligned with physicalism. My argument builds on and diverges from Vision's view of emergence. Vision's endorsement of strong emergence as an ontological, rather than epistemological, distinction is foundational to my view. I present a view of strong emergence that endorses its novel properties and draws from *ceteris paribus* laws as a driver of ontologically distinct emergent phenomena.

I contend that emergence satisfies physicalism's metaphysical thesis and does so without diminishing the role of mental properties. Regarding physicalism's epistemological thesis, emergence provides an explanation for the relationship between mental and physical facts, but it does not claim that physical facts reveal specifics about mental facts. I propose a view of emergentism that posits a tight relationship between physical and mental facts, meaning that emergentism requires a revised epistemological thesis to qualify as physicalism.

This chapter is pivotal because it champions emergence as a physically sound explanation for mental properties. Though my position is primarily based on Vision's study of emergence, I uniquely synthesize several concepts from the philosophy of science and the special sciences to offer new details and understanding of emergence. Ultimately, I provide a unique perspective on emergence, differentiating it from other emergentist theories and setting the stage for the subsequent discussion on mental

causation. Notably, I advance emergence as a mental-physical relationship that could lead to a more inclusive version of physicalism. This supports my views in Chapter 4 that emergence and Dynamically Stable Causal Holism are physically compatible.

I believe physicalism, as our prevailing worldview, should be the initial framework within which we attempt to reconcile emergent phenomena. This methodological approach reflects a pragmatic stance, seeking to adapt and refine our understanding of physicalism to accommodate the facts of emergence rather than immediately opting for alternative worldviews.⁹⁸ By aligning with physicalism, emergentism is strategically positioned within the prevailing worldview. This fact alone gives emergence particular advantages over other theories of mind that accept mental realism but do not abide by physicalism, such as substance dualism.

Secondly, part of my rationale favoring strong emergence focuses on a rebuttal of scientific idealism. By incorporating notions from the philosophy of science, including a rejection of the unity of science and an endorsement of *ceteris paribus* laws, I develop a theory of reality that is not as linear as the one presented by physicalists focused on supervenience and causal drainage. Through greater distinguishment between reality's layers, I set the stage for an alternative causal theory, allowing for top-down causation and defeasibility at lower levels.

In the remainder of this introduction, I will overview three reasons why emergence is an intuitive and appealing mental-physical relationship and outline its challenges.

⁹⁸ For instance, if I did not bind myself to this methodological approach, I would be inclined to pursue neutral monism as an alternative to physicalism.

The first advantage of emergentism is that the theory takes mental realism seriously in its own right. As a theory of mind, emergence affirms the reality of mental properties and considers those properties to be as real as physical properties. Unlike strict physicalism, which tends to sideline and severely limit mental phenomena based on an idealized physicalism, emergence recognizes mental life without summarily diminishing its significance. This is an advantageous position because the emergentist can approach mental realism as it is experienced. In contrast, stringent physicalists tend to minimize mental properties despite their impact on us and our understanding of reality. In other words, emergentists can embrace mental realism and mentality's role in comprehending reality.

Second, the metaphysical and epistemological assumptions that emergence makes are more metaphysically founded than the assumptions of competing physicalist theories of mind. Consider, for instance, the two options Churchland provides for mental realists, namely that mental properties must either be fundamental or generated. He writes:

[a] property dualist is not absolutely bound to insist on both claims [i.e., the simultaneous claim of evolutionary emergence and irreducibility]. He could let go of the thesis of evolutionary emergence, and claim that mental properties are fundamental properties of reality, properties that have been here since the universe's inception, properties on a par with such properties as length, mass, charge, time, and other fundamental properties.⁹⁹

Churchland notes that mental properties could be fundamental, having always existed alongside physical properties, even before sentient organisms existed. Conversely, if mental properties are generated, there is a mechanism in nature that allows for the

⁹⁹ Churchland [1984] 2013, 20. The bracketed text is added for clarity but consists of a quote from Churchland's preceding paragraph.

formation of mental properties. Typically, physicalist mental realists buck this dichotomy and opt for a third approach. The dominant physicalist views, including supervenience, identity, and reductive functionalism, acknowledge the existence of mental properties and then govern them without considering their origin. Though physicalists purport physicalism to be the most grounded and evidentiary-based worldview, this approach of granting mental realism and ignoring other pertinent questions about these properties leaves much to be desired.

A key advantage of emergence over other theories of mind is that it asserts the origin of mental properties. This origin is couched in the dominant worldview (i.e., physicalism), thereby avoiding the pitfalls of the contrasting position, namely that mental properties are fundamental. There is a vast epistemological and metaphysical difference between the position that mental properties are generated by physical complexity versus that mental properties exist fundamentally. Notably, if mental properties are fundamental, physicalism would not be able to accommodate that idea, revealing a serious flaw in our leading worldview.

By taking a stand on the origin of mental properties, emergence provides ontological grounding for mental properties, whereas supervenience, identity, and functionalism do not. I argue that the grounded metaphysics of emergence should appeal more to physicalists than the mental-physical relationships discussed in Chapter 2 because of its ontological commitment to understanding the nature and existence of mental properties.

Third, emergence is a physicalist-aligned theory of mind that may allow for downward mental causation. If emergence offers a viable explanation for downward

mental causation, then the direct experience of the mind seemingly affecting bodily actions would be justifiable. Moreover, it is worthwhile to note that emergence is studied in disciplines outside of philosophy as well. In other disciplines, emergent systems—typically complex systems that lead to unexpected higher-order outcomes—can influence and affect lower-level constituent elements. Thus, if it is the case that mental properties form an emergent system, then it is judicious to argue that the mind—as an emergent whole—can affect its constituent parts (e.g., parts of the brain or nervous system). Ultimately, if true, the mind would have hegemony over some of our actions and behaviors.

The possibility of downward causation puts emergence in a unique and potentially superior position to other mental-physical relationships. If emergence can accommodate downward causation while remaining aligned with physicalism, it would have a solid claim to being the preeminent theory of mind. Emergence would do this by killing two birds with one stone. It would resolve the ontological status and causal efficacy of mental properties by employing the preferred metaphysical method—Alexander’s Dictum.¹⁰⁰ To elaborate, emergence would demonstrate its superiority by achieving a dual feat: (i) it would validate the status of mental properties, and (ii) grant mental properties a legitimate causal role instead of the ceremonial roles offered by other physicalist-endorsed mental-physical relationships. Moreover, it would accomplish this dual feat while remaining grounded in physicalism. In effect, emergence would bridge the divide between mental realism and physicalism, combining the strengths of both, all while

¹⁰⁰ To be is to be causal.

serving as the foundation of a potential explanation for the apparent influence of the mind on our actions.

The three reasons overviewed make a compelling case for emergentism, but there are also several points of contention. I find there to be four general objections to emergence.

First, the belief that the mechanism of emergence is inherently unknowable presents itself to critics as an obfuscation of the explanatory gap problem. This would make emergence no more explanatorily valuable than the other mental-physical relationships that fall short of explaining mental properties in one or more ways.

Second, emergence does not submit to proper scientific scrutiny because it appeals to a brute relationship between emergent phenomena and their bases. This objection is rooted in the belief that any proper scientific explanation must be grounded in precise, understandable mechanisms, not inexplicable brute facts. I will address these first two objections in this chapter.

Third, emergence cannot be classified as a form of physicalism because it conflicts with physicalism's core tenets. To Kim, the core tenets are that physicalism must be (i) reductive, (ii) causally closed, and (iii) not overdetermined. I lay the groundwork for how emergence qualifies as a form of physicalism in this chapter, and I directly address Kim's physicalism in Chapter 4.

Fourth, Kim contends that emergence leads to either epiphenomenalism or causal overdetermination. The core of this objection is the Causal Exclusion Argument and the general notion that for mental properties to have downward causal powers, one must either (i) reject supervenience, (ii) reject causal closure, or (iii) embrace widespread

causal overdetermination. If (i), (ii), or (iii) cannot be rejected, one must (iv) embrace epiphenomenalism. Kim does not seriously entertain the possibility of (i) or (ii), which leaves only options (iii) and (iv) for him. This fourth objection to emergence is addressed in its entirety in Chapter 4.

To address the advantages and criticisms associated with emergence, this chapter will be structured as follows. I describe emergence and solidify it as a robust theory. I do so, in part, by drawing from diverse fields that embrace emergence. After introducing the concept, I argue that emergence is a viable theory of mind that takes mental realism seriously. Next, I argue that the mysteriousness of emergence is not as dire as its opponents claim and that the mysteries native to the other mental-physical relationships are equally troubling. I then address the major criticisms of emergence and conclude that emergence's explanatory advantages outweigh its faults. The third advantage and fourth criticism, which focus on downward mental causation, are the focus of Chapter 4.

Emergence

In its most general form, emergence is the concept that some higher-order phenomena and properties *emerge* from their underlying parts' complex arrangement and interactions. Consider the description of emergence by one of the British Emergentists. C.

Lloyd Morgan writes:

[w]hen there is increased complexity, with variety and multiplicity of the items of stuff, and increased richness [...] in the substance (or substantial unity) in virtue of which the entity is an indivisible whole, there is evolutionary advance. [...] [W]ith each step or jump there are in evidence new qualities and properties which characterise the new entity which leaps into being without loss of some at least of its old traits. This step-like

advance, this coming of new characters with a leap—not necessarily a big one but still saltatory—is what I mean by emergent evolution.¹⁰¹

Morgan indicates that in certain cases of complexity, novel qualities “leap into being.”¹⁰²

Notably, complexity and novelty are two key characteristics of emergence.

Emergent phenomena are often described as exhibiting novel properties and behaviors that arise from the complex interactions of simpler components. These emergent properties are also irreducible to, and unpredictable from, the properties of individual components. Kim writes:

[...] as systems acquire increasingly higher degrees of organizational complexity they begin to exhibit novel properties that in some sense transcend the properties of their constituent parts, and behave in ways that cannot be predicted on the basis of the laws governing simpler systems.¹⁰³

We typically refer to the constituent parts and the various considerations accompanying the underlying parts, including the organization and relation of those parts and their properties, as the emergent phenomena’s *base*. In supposed cases of emergence, that which is emergent is not *present* in the underlying parts.

Bedau explains that “[m]acro properties are traditionally classified into two kinds: genuine emergent properties and mere ‘resultant properties,’ where resultant properties could be viewed as those that can be predicted and explained from the properties of the components.”¹⁰⁴ Bedau’s explanation illuminates the contrast between emergence and resultant properties. Consider, for instance, the distinction drawn by McGivern and Rueger. They write:

¹⁰¹ Morgan 1925, 70-1. Please note that Morgan uses the term evolution idiosyncratically. He writes, “the word ‘evolution’ has several different meanings, one of them has to be chosen. In that which I accept it names the orderly advance of natural events, without restriction to these events or to those.” Ibid, 70.

¹⁰² Notably, Morgan’s description of emergent evolution responds to a skeptic—Flora I. MacKinnon—critiquing the value of emergence. See MacKinnon 1924.

¹⁰³ Kim 1999, 3

¹⁰⁴ Bedau 2010, 50.

[...] the core criteria we find associated with emergence are those of *non-reducibility* and *novelty*: emergent phenomena are in some way irreducible to and novel with respect to their base phenomena, whereas ‘resultant’ phenomena are reducible and/or non-novel. In fact, we see these two criteria as two sides of the same coin: emergent phenomena are typically taken to be not only novel but in some way ‘qualitatively’ novel, and talk of irreducibility often seems intended to capture just this distinctive feature.¹⁰⁵

It is beneficial to distinguish between emergent and resultant phenomena because it allows us to understand and categorize complex and diverse phenomena. Resultant phenomena can be reduced to or predicted from the properties and interactions of their constituent parts. Resultant phenomena are the *expected results* of these interactions, and as such, they do not exhibit novelty or unpredictability. Resultant phenomena align well with reductionism, the widely adopted methodology that seeks to explain systems and phenomena in terms of what underlies the phenomena in question.

Emergent phenomena, on the other hand, are characterized by irreducibility, unpredictability, and novelty. These properties or behaviors cannot be explained by their underlying components or predicted by examining their underlying parts. The emergent phenomena will also display qualitatively new abilities or properties not found in the underlying parts.

The concept of emergence, in one form or another, stretches back to at least the Greek philosophers.¹⁰⁶ For example, the roots of emergence can be seen in Aristotle’s concept of hylomorphism—the compound of form and matter—that emerges from

¹⁰⁵ McGivern and Rueger 2010, 215.

¹⁰⁶ Vision traces early notions of emergence back to Anaximander; he writes “Anaximander (fl. 571 BCE) held that living creatures arose from moisture evaporated by the sun.” Vision 2011, 3.

various arrangements of the four material substances.¹⁰⁷ Emergence re-entered the modern philosophical era with the British Emergentists—notably, J.S. Mill, Samuel Alexander, and C.D. Broad—but was relegated to novelty status by subsequent scientific discoveries that were decisively reductive.¹⁰⁸

Now that we have traced the roots of emergence, it is worth noting how modern scientific paradigms have evolved with these ideas. For instance, as interest in emergence waned, there was a rise in the belief in a unity of science.¹⁰⁹ Oppenheim and Putnam describe the unity of science as a combination of accepting one or more of three broad categories. They write:

[f]irst, Unity of Science in the weakest sense is attained to the extent to which all the terms of science are reduced to the terms of some one discipline (e.g., physics, or psychology). [...] Second, Unity of Science [...] is represented by Unity of Laws. It is attained to the extent to which the laws of science become reduced to the laws of some one discipline. [...]. Third, Unity of Science in the strongest sense is realized if the laws of science are not only reduced to the laws of some one discipline, but the laws of that discipline are in some intuitive sense "unified" or "connected."¹¹⁰

For several reasons, the unity of science became a widely adopted and prevailing idea.

Among them was the push to eliminate as much metaphysics as possible in favor of observable and verifiable facts. Also closely tied to that effort was the seemingly unmarred success of reductionism as a methodology for tying what seemed like different

¹⁰⁷ O'Connor writes, "Aristotle's form/matter compound conception of material substances is consonant with the standard emergentist stance between substance dualism and reductionism." O'Connor 2021, 4.

¹⁰⁸ O'Connor writes, "[...] the emergentist vision waned beginning shortly after Broad's writing, with important scientific developments eroding the boundaries between adjacent 'levels' (most notably quantum chemistry and molecular biochemistry [...])." O'Connor 2021, 6-7.

¹⁰⁹ Notably seen in the early debates between Max Planck and Ernst Mach about the best way to unify science. For an overview of Planck's and Mach's debate, see Toulmin 1970, 1-52. Debates continued through the logical positivist movement, notably through the work of Rudolph Carnap and Otto Neurath. See Cat 2017 for a more detailed overview.

¹¹⁰ Oppenheim and Putnam 1958, 3-4.

layers of reality together and explaining those various levels in simpler, more fundamental terms.

While the unity of science offers a comprehensive framework for understanding reality, reductionism takes a more granular approach. Reductionism is often hailed for its simplicity and explanatory power. The idea is that once we know a system's fundamental components, we will grasp the entire system's behavior. Subscribing to reductionism as the absolute method entails believing in the power of reductions to solve every puzzle. I refer to the idea that reductionism can solve all puzzles and, therefore, has dominion over other methodologies as "absolute reductionism." Conversely, emergence is naturally in conflict with absolute reductionism because it suggests that some phenomena and properties cannot be thoroughly explained by analyzing the parts of the whole.

The preference for reduction or a form of holism has swung like a pendulum throughout the Western philosophical tradition. For example, Aristotle, when positing hylomorphism and teleology as explanations for the structure and behavior of reality, is essentially recognizing that the whole, in and of itself, plays a crucial descriptive role that the parts—or even the sum of the parts—cannot capture on their own. Fast forwarding through the centuries, as modern science advanced, starting with Newton and continuing until today, reductive explanations gained a strong foothold in science and philosophy.

Though emergence waned shortly after Broad's sophisticated treatment, the concept has been revived in the philosophy of mind and various scientific disciplines. Its resurgence in philosophy can be largely attributed to mental properties and phenomena being irreducible to physical theory. Moreover, counterarguments to reductive physicalism, such as the oft-cited explanatory gap, led to physicalists' broader acceptance

of mental realism. The drive to reconcile physicalism with mental realism demands a different approach to explaining the existence of mental properties, as I have hopefully elucidated in this work.

Another key reason emergence has gained traction in the literature is that the special sciences, including biology, climatology, and ecology, now utilize the concept.

For instance, Cartwright writes:

[t]here is a tendency to think that all facts must belong to one grand scheme, and moreover that this is a scheme in which the facts in the first category have a special and privileged status. They are exemplary of the way nature is supposed to work. The others must be made to conform to them. This is the kind of fundamentalist doctrine that I think we must resist. Biologists are clearly already doing so on behalf of their own special items of knowledge. Reductionism has long been out of fashion in biology and now emergentism is again a real possibility.¹¹¹

This quoted text is laden with important ideas, including an implied skepticism about absolute reductionism and the unity of science, but presently, I wish to note Cartwright's recognition of biology's reconsideration of emergentism. Physicalism is ultimately rooted in physics, but a deep appreciation for the natural sciences also prevails among physicalists, making biology's embrace of emergence noteworthy.¹¹² Furthermore, emergence is not only a tool of the special sciences; it is also increasingly referred to by physicists. For example, one such physicist studying emergence, Lars Q. English, explains:

[a] reductive view of science usually leads to categorizing the various sciences in terms of a strict hierarchy of significance. At the top are the social sciences, [...] [and] physics at the bottom. [...] Emergence opposes that hierarchical view [...]. [...] Meaning is found in the effective theories

¹¹¹ Cartwright 1999, 25.

¹¹² For example, see Okasha 2012 for his exploration of natural selection across difference levels of biological organization.

or models developed to explain a particular phenomenon at its appropriate level of description. The sciences are not separate but make contact [along] lines of continuity. A phenomenon at a larger scale is best described at that scale, but it is supported by processes at shorter lengths. These processes are typically not in a one-to-one relationship with the emergent phenomenon and often can't be used to derive it, but they participate in its appearance nonetheless.¹¹³

Like Cartwright, English expresses skepticism in absolute reductionism and the unity of science in favor of emergent explanations. Ultimately, the use of the concept in disciplines that physicalists hold in high esteem should legitimize the exploration and use of emergence in philosophical discourse.

Reductionism is the premier methodology of the sciences, so we can assume that reductions are generally successful. This fact might mean that most known macro properties and phenomena are resultants. Based on this, we might assume that emergent phenomena may be rare compared to resultants. If that were the case, then emergence may only occur under unique circumstances, making it, in a sense, a special feature of reality. By distinguishing emergent and resultant phenomena, we can better identify and explore the limitations of reductionism, appreciate the complexity and diversity of reality, and potentially develop new methodologies and theories that can explain the existence and behavior of complex, higher-order systems.

The interplay of emergence and reductionism serve as lenses through which we perceive and understand the complexity of reality. Reductionism, with its emphasis on simplicity and explanatory power, asserts that understanding the fundamental components can elucidate the behavior of the entire system. Emergence asks us to

¹¹³ English 2017, 191-2.

appreciate the novel properties and behaviors that arise from the complex interaction of simpler parts while emphasizing that the whole can be more than the sum of its parts.

The Layered Conception of Reality

Emergentism characterizes reality as carved up into hierarchical *levels* of ascending complexity. We can call this the layered conception of reality. To make sense of this, consider Rabin's description; he writes:

[I]et's use the phrase "the layered conception of reality" ("the layered conception" for short) as a label for the general idea that reality is layered in a hierarchy structured by relations of dependence. We can add a claim about fundamentality to the layered conception: the lower tiers of the layering are more fundamental than the higher tiers.¹¹⁴

The layered conception asserts that there is a distinction between the various levels and the entities that populate each level of reality. Intuitively, we categorize these levels along the lines of the special sciences and the objects each examines. For example, quantum physics examines subatomic particles, chemistry examines molecules, biology examines living organisms, ecology examines environments, etc. For physicalist-emergentists, the fundamental or foundational level of this hierarchy would be the quantum level consisting of subatomic particles or quantum fields. This level underlies all levels, objects, phenomena, and properties.

Notably, reducing levels and the objects that populate each level is also helpful when distinguishing the layered conception. Intertheoretical reduction posits that the theories and principles governing higher-level phenomena can be reduced to the theories and principles of a more fundamental level. For example, the reproduction studied by biologists can, in theory, be reduced to chemistry. The ability to reduce special sciences

¹¹⁴ Rabin 2018, 37.

and their target phenomena to fundamental sciences and their target phenomena is one method for distinguishing levels and the objects populating each level.

As noted, we can define each layer by how we understand said layer and the concepts that populate that understanding. For instance, chemistry is a more practical science than quantum physics when dealing with molecules. When dealing with wave-particle duality, quantum physics is the science required. Even if, in theory, quantum physics could tell us everything we need to know about molecules, chemical principles are easier to work with than quantum calculations. Furthermore, the concept of a molecule is native to chemistry, while it is non-native to quantum physics. For these reasons, the chemical layer is intuitively carved out, and the other layers follow the same rationale, albeit focused on their own scientific methodology, objects, and principles.

Herein lies one of the reasons that bridge laws are necessary to link the concepts and principles of different levels of reality. We intuitively understand each level as a level unto itself, consisting of its own objects, principles, and methodologies. Some may accept that reality is ultimately just fluctuations in quantum fields, but few believe we could understand reality as well as we do if we only utilized quantum physics. In other words, even if reality is just one big hunk, we carve layers to analyze and better understand the whole of reality.

Categorizing levels by the special sciences and by way of reduction is intuitive, but what do the special sciences and reduction tell us about the layered conception? In other words, is reality layered in an ontologically substantive way, or is each layer merely a conceptual device marking an intellectual categorization? Carving reality at the joints of the special sciences and bolstering those levels through talk of intertheoretical reductions

does not decisively reveal whether the layered conception is substantive or intellectual; however, it does point toward an epistemological approach to answering the question. The epistemological approach asks us to determine the status of each layer based on the best evidence we have available. This is opposed to a metaphysical approach, which could begin from an ontological assumption and determine if available evidence supports or contradicts that assumption.

Evidence suggests that higher-order phenomena metaphysically depend on lower orders, at least in terms of metaphysical priority. A deep examination of higher-level phenomena often reveals lower-level phenomena within. For example, as we delve into a speck of skin, our investigations uncover underlying structures such as cells, DNA molecules, and chemical elements. This suggests that higher-order phenomena are as real as their underlying makeup. To deny the reality of the speck of skin because it is revealed to be a collection of cells does not make much sense unless one also denies the existence of the cells, DNA, and so on to grant reality only to the fundamental level and its subatomic particles.

I find it questionable to sacrifice the parts of reality nearest our awareness in favor of the parts farthest from our awareness.¹¹⁵ After all, our understanding of the fundamental level stems from knowledge borne out of our awareness of higher levels. So though lower levels may have ontological priority, higher levels have a form of epistemological priority. For example, a speck of skin is more familiar to us than the

¹¹⁵ For example, take the metaphysics of Democritus. The following text is often attributed to him: “By convention sweet and by convention bitter, by convention hot, by convention cold, by convention colour, but in reality atoms and the void” (DK 68 B9). See Long 1999, 190.

quantum states that underly the speck. We know the speck (and other macro phenomena, principles, and so on) before we can know the underlying layers.

What we uncover about lower-level phenomena provides crucial information and contextualization that forms foundational knowledge of higher-level phenomena. In other words, the chemical elements within a speck of skin contextualize the DNA molecules, which in turn help to contextualize the cells and so on. Yet, the attempt to reduce a speck of skin to quantum terms fails to capture its full reality, illustrating the absurdity of suggesting that a quantum-level description could be equivalent to the macroscopic experience and knowledge of the speck of skin.

The idea that lower levels metaphysically constitute and epistemologically contextualize higher ones is consistent with physicalism and emergence. However, emergentism maintains that emergent phenomena metaphysically depend on—without being determined by—their base. The distinction between metaphysical priority and determination sets it apart from the standard physicalist understanding of levels with its emphasis on reductionism. Under reductionism, lower levels are thought to fully determine the ontological character of higher-level entities. Emergence, in contrast, recognizes new entities and properties at various levels, suggesting that there is ontological novelty at different layers or levels of reality. This marks the major ontological distinction between the layers according to emergence; some higher-level phenomena possess unique properties and characteristics that cannot be reduced to those at lower levels. This idea is significant because it suggests that emergent phenomena have legitimate ontological status due to their irreducibility, unpredictability, and novelty.

Both physicalism and emergentism agree that higher-level phenomena metaphysically depend on lower-level phenomena (i.e., there is metaphysical priority), but emergent phenomena resist reducing to lower-level phenomena. If absolute reductionism were true, then reductive physicalism would be the prevailing worldview. However, the core issue is that at least some higher-level phenomena cannot be reduced. Ultimately, under emergentism, reality's layers signify more than mere conceptual devices; they denote actual ontological differences.

To elaborate, if the character of a higher-level phenomenon (c.f., a higher-level concept) cannot be reductively explained without sacrificing key elements of the phenomena, we must recognize and maintain the ontological status of the higher-level entity. To do otherwise would be to disregard a unique and possibly essential aspect of reality simply out of preference for, or blind faith in, absolute reductionism. While higher-level entities are composed of lower-level ones, the properties at each level may exhibit ontological novelty absent at lower levels. This departure from reductionism suggests an ontological autonomy, highlighting unique properties and characteristics that resist reduction.

Notably, the existence of irreducible higher-level phenomena does not negate the power of reductions. In other words, ontological reductions, intertheoretical reductions, and bridge laws would continue to serve as useful tools that connect reality's levels. However, so long as there are irreducible ontological distinctions, emergentism will remain another compelling tool for explaining the complex features of reality.

Ultimately, the case for emergentism hinges on the existence of features of reality that possess novel properties that resist reducing to lower levels. As long as reductions remain elusive, the best evidence points to emergentism over traditional reductive physicalism. That said, it is generally held that lower-level entities constitute those at higher levels, and I do not wish to challenge that idea. A key component of emergentism is that higher-level phenomena *depend* on lower levels to exist. The notion of dependence can be formulated in various ways, but there are two conceptions related to emergence that I find necessary to address—grounding and realization.

Grounding is the idea that higher-level phenomena necessarily depend on lower-level phenomena. Bryant writes:

[g]rounding is a metaphysical dependence relation that tracks the ‘in virtue of’ locution [...], so that whenever we say that *x* exists or obtains in virtue of *y*, we may also say that *y* grounds *x*. The relation is generally thought to involve necessitation [...]. The nature of the dependence is also sometimes thought to be productive in some sense. [...] At any rate, the thing doing the grounding is thought to be more fundamental ontologically than the thing being grounded. Since some prominent grounding theorists [...] take grounding to be a primitive term, examples often play a key role in its elucidation.¹¹⁶

Grounding establishes a metaphysical hierarchy where higher-level phenomena owe their existence to lower-level phenomena. Grounding establishes that lower levels determine higher-level phenomena and their features.

Conversely, I have characterized emergent phenomena as metaphysically dependent on lower levels but not entirely determined by or reducible to those levels. Consider Vision’s description of how the grounding principles break down in the

¹¹⁶ Please note that in this quote, Bryant refers to Bennett 2011, 90–92, Bennett 2017, Fine 2012, and Skiles 2015.

presence of higher-level properties. To make his point, he shares an example of a lump of clay, *L*, used to form a statue of Pericles. He writes:

Pericles sports a look of triumph, but does *L*? Moreover, recall that constitution and realization are non-reflexive; whereas Pericles is constituted by and realized in *L*, *L* isn't constituted or realized in *L*. [...] Still, it is uncertain what lesson we are meant to draw from [the grounding principle]. Is it that, despite appearances, *L* possesses all the elements of the persistence conditions for Pericles? That Pericles doesn't possess them? [...] The grounding principle by itself is problematic enough to permit us to set aside further inquiries about its role in the larger argument.¹¹⁷

Vision's exploration of the grounding principle and the relationship between the lump of clay and Pericles illustrates a potential limitation and fault in metaphysical grounding. Vision's example focuses on the nature of this dependence, mainly when dealing with higher-level properties. This highlights the need for careful consideration and further examination of grounding's role, especially when applied to emergence.

A strict reading of grounding establishes metaphysical determination between lower and higher levels. Vision's example clearly illustrates that a statue's properties and the properties of the lump of clay it is composed of are not as neatly aligned as grounding suggests. For instance, Pericles sans an arm arguably remains a statue of Pericles, but the lump of clay would be substantially changed. Specifically, the mass of the lump of clay would be an essential characteristic of the lump, but the lump minus the mass of Pericles's arm suggests that Pericles persists while the lump does not.¹¹⁸ Without stronger evidence favoring determinacy between the lower-level and the emergent phenomena, I do not see metaphysical grounding as a viable explanation for the relationship between

¹¹⁷ Vision 2011, 74.

¹¹⁸ This scenario is also borrowed from Vision. Vision 2011, 66-70.

emergent phenomena and their bases. By establishing vertical determination, grounding treads closely to reductionism, which would lead us to rehash the failure of reductive mental-physical relationships.

Realization is another type of relationship often invoked to describe emergence. Realization concerns how a higher-level phenomenon is instantiated or implemented by a lower-level phenomenon, but it does not imply the same strict dependence as grounding. The flexibility of realization allows for a more nuanced understanding of the relationship between emergent phenomena and their realizing base.

Under realization, an underlying physical state realizes higher-level phenomena or properties, but all higher-level phenomena are not entirely determined by nor reducible to the realizing state. Regarding emergence, emergent phenomena metaphysically depend on the realizing state, but its novel features are not determined by and cannot be reduced to or predicted from the realizing state. The unpredictability of novel properties further distinguishes realization from grounding. As such, the realized phenomenon is not the inevitable result of the realizing state, nor does the realizing state entirely capture it.

Furthermore, the realized and realizer are not in a one-to-one correspondence relationship, so emergent phenomena can be multiply realized, and the layered conception can be understood as a rich complexity. For example, the mental state of anxiety can be realized by different realizing states. Anxiety can be realized in the same human at various times through different neural patterns. Anxiety could also be realized in structurally similar (yet different) organisms like dogs and humans and potentially even among entities with entirely different structures. Given the importance multiple

realizability plays in our discussions of mental realism and realization, its allowance for multiple realizability provides an added level of robustness to emergence.

Moreover, realization allows for a dynamic and potentially interactive relationship between reality's layers. Notably, realized higher-level phenomena may be able to interact with its realizing lower-level phenomena, allowing for downward causation. The relational uniqueness of realization underscores why it is more suitable than grounding for understanding emergent phenomena.

While grounding provides a deterministic framework that aligns closely with reductionism, its strong formulation is too rigid to capture the relationship between lower-level and emergent phenomena. On the other hand, realization better suits the multifaceted nature of emergence and the general understanding of mental properties. Notably, realization allows for a layered conception that recognizes the ontological dependence of higher-level phenomena on lower levels without reducing them to lower levels or making them resultant features of reality. Realization's multifaceted approach provides a more robust framework that recognizes complexity, multiple realizability, and potentially even the interactivity of levels.

Interpreting Emergence

There are two ways of interpreting emergence. The first interpretation suggests that real novel properties and phenomena come into being (i.e., emerge) from a natural but unknowable process. The second interpretation is more tentative, claiming that

emergent phenomena are unexpected, presumably due to an existing knowledge gap or limitation.¹¹⁹

The first interpretation denotes strong emergence, maintaining that emergent properties are genuinely novel and irreducible due to their ontological realization (i.e., in virtue of their coming into existence). Underlying properties and events can generate novel properties, phenomena, and events, but they are not reducible to them. This interpretation maintains that the process of emergence is opaque to us, meaning that emergent properties are not predictable or explicable from the natural laws and principles of the base. Chalmers describes strong emergence as follows:

[...] a high-level phenomenon is *strongly emergent* with respect to a low-level domain when the high-level phenomenon arises (in some sense) from the low-level domain, but truths concerning that phenomenon are not *deducible* even in principle from truths in the low-level domain.¹²⁰

Strong emergence is the idea that the activity of a complex system generates higher-level properties, and those higher-level properties cannot be predicted from or located in the parts of the complex system. Strong emergentists hold that we do not know how higher-level properties come to be generated, nor can we know.

The second and more tentative interpretation denotes weak emergence, focusing on our inability to deduce or predict emergent phenomena. Chalmers defines weak emergence as:

[...] a high-level phenomenon is weakly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are *unexpected* given the principles governing the low-level domain.¹²¹

¹¹⁹ It is worth noting that Chalmers prefers “unexpected” to “unpredictable.” Chalmers 2008, 245.

¹²⁰ Chalmers 2008, 244

¹²¹ Chalmers 2008, 244-5

Weak emergence is a more reserved claim about the nature of emergent phenomena. By claiming that emergent phenomena are merely *unexpected*, weak emergence is distinguished from strong emergence, which finds that emergent phenomena are not deducible even *in principle*.

To summarize, under the strong interpretation, emergent phenomena are novel and irreducible because they are, in fact, novel and irreducible. Strong emergence denotes a metaphysical approach by positing that emergent properties are inherently novel and irreducible. In contrast, weak emergence utilizes an epistemological approach that focuses on our inability to predict emergent phenomena. Under the epistemological interpretation, emergent phenomena are novel and irreducible because we do not know how to reduce said phenomena. Weak emergentists may believe that our current understanding or methods may be insufficient for predicting or deducing emergent phenomena. Though currently deficient, the possibility remains that discoveries and refined methodologies may reveal how the underlying system generates so-called emergent phenomena.

Under the weak interpretation, emergent phenomena are akin to resultants that have yet to be determined. This form of emergence is often utilized in fields outside of philosophy. For instance, consider the definition provided by computer scientists Vinerbi, Bondavalli, and Lollini. They write:

[i]n Computer Science the concept of emergence is strictly related to the level of knowledge [...] of a specific system. Since computers are deterministic machines, the overall system's behavior can be seen as a function of the components' behavior (subfunctions) composing it; ideally, having the full knowledge of this function, [...], would allow to fully predict the overall system's behavior, [...]. However such kind of knowledge is unrealistic in most of the current complex information

systems, called *emergent*, for which it is impossible to foresee the overall system's behavior from the composition of its single sub-functions due to the lack of knowledge on some aspects of the system, and this is the source of the emergence. Accordingly, we define *emergent phenomenon* as an unexpected system behavior not directly traceable to the system's components' behaviors, and *emergence property* as a system quality that cannot be deduced by the properties of the system's constituent parts.¹²²

This description of emergence describes a notion that is considerably like weak emergence.¹²³ Emergent phenomena are tied to complexity in the system, leading to an unexpected outcome because of a lack of necessary knowledge.

Regardless of the extent to which each interpretation is willing to go, both maintain that emergent phenomena resist reductive explanations. The strong interpretation entails a more sweeping claim about the nature of reality. In contrast, the weak interpretation is more tentative in its claims about reality, focusing instead on our knowledge and understanding of complex systems.

For illustrative purposes, let us compare the weak interpretation with an anti-emergentist stance. The anti-emergentist may concede that we cannot calculate so-called emergent phenomena due to knowledge or computational limitations, but they are likely to claim that, in principle, those predictions are possible. At least some anti-emergentists would hold that something like Laplace's demon would have no difficulty predicting so-called emergent phenomena with certainty, and a supernatural entity would arguably be far more intellect than is required.¹²⁴ In describing Laplace's demon, Collier writes:

¹²² Vinerbi, Bondavalli, and Lollini 2010, 134.

¹²³ I utilize this long quotation because the authors' description illuminates the similarities and differences between philosophy's and another discipline's view of emergence.

¹²⁴ Please note that Laplace does not use the term, "demon," but the implications remain the same. Laplace writes, "[w]e ought then to consider the present state of the universe as the effect of its previous state and as the cause of that which is to follow. An intelligence that, at a given instant, could comprehend all the forces by which nature is animated and the respective situation of the beings that make it up, if moreover it were

[...] Laplace's methodology contains two assumptions, locality and predictability, [...]. These assumptions are central to reductionist methodology guiding most of science until the last few decades. If we give up these assumptions, the door is open to holism and emergence, the former eschewing locality, and the latter adopting nonreducibility, unpredictability and novelty of emergent properties [...].¹²⁵

Laplace's demon illustrates an anti-emergentist position. Laplace's universe is wholly deterministic, allowing the demon to predict future states of the universe with certainty. The demon's ability contrasts sharply with the notions of unpredictability and irreducibility central to strong emergence.

The strong emergentist would deny that Laplace's demon could calculate emergent phenomena and that knowledge of the microphysical facts and non-physical conceptual truths would reveal all truths, including any truths about emergent phenomena. According to the strong emergentist, emergent phenomena are ontologically novel, distinct, and cannot be deduced even *in principle*.

A deterministic worldview is a powerful concept, but it can be criticized in several ways. For one, quantum mechanics, our prevailing physics, is probabilistic. If quantum mechanics is our best look into physical reality, then its probabilistic nature suggests that reality is not deterministic. One such reason to believe this conclusion is the uncertainty principle, which states that one cannot simultaneously know the exact position and momentum of a particle. If the uncertainty principle accurately captures reality, then Laplace's demon would not even, in principle, be able to possess the knowledge necessary to make certain predictions.

vast enough to submit these data to analysis, would encompass in the same formula the movements of the greatest bodies of the universe and those of the lightest atoms. For such an intelligence nothing would be uncertain, and the future, like the past, would be open to its eyes." Laplace [1825] 1995, 2.

¹²⁵ Collier 2011, 231.

While strong emergence posits that certain phenomena are fundamentally unknowable, unpredictable, and irreducible, weak emergence is open to the possibility of emergent phenomena eventually being understood, predictable, and reducible. This distinction highlights a foundational difference in how the two approaches view the nature of reality and what is knowable.

Weak emergentists acknowledge our inability to predict or deduce emergent phenomena but take a provisional stance, maintaining that that inability may be temporary and due to incomplete knowledge. In this sense, weak emergence could be seen as waiting for scientific or philosophical breakthroughs to bridge the gap between emergent phenomena and their underlying causes. In blunt terms, weak emergentism is like reductionism in waiting.

In contrast, strong emergentists maintain that certain phenomena are irreducibly novel and cannot be understood solely in terms of their underlying parts, even in principle. Strong emergentists assert that knowledge of emergence as a process is intrinsically beyond our reach. Advances in technology, methodology, or knowledge of the underlying parts will not reveal the details of strong emergence.

This distinction between strong and weak emergence reflects differing views on the nature of reality, the limits of human understanding, and the potential for future discovery. While the strong interpretation posits inherent and insurmountable barriers to understanding emergent phenomena, the weak interpretation sees these barriers as possibly surmountable.

Generally, I try to approach philosophy from a place of epistemological humility and try not to deal with absolutes. One might interpret that to mean I am a weak

emergentist; however, I take a nuanced position. Our epistemological limitations and shortcomings demand that we stay open to new methodologies, technologies, and knowledge, but they also suggest we need to act on the best evidence available.

Our inability to reduce, predict, and understand emergent phenomena, evidence of emergent phenomena, and our best physics pointing toward uncertainty all favor strong emergentism. Strong emergentism better aligns with our intuitive experiences of irreducible mental properties as well as with other complex systems that behave in ways that are irreducible to their parts. To be a weak emergentist is to distrust the evidence we have and proceed without committing to how we understand reality. Instead, the weak emergentist commits to an idealized future worldview that may or may not come.¹²⁶ The weak emergentist position extends too far by believing that a reductive explanation lies in waiting. In other words, the weak emergentist is operating on blind faith, committed to a reductive explanation without evidence to support it.

I believe that as long as the evidence supports strong emergentism, we ought to commit and proceed as strong emergentists. Strong emergence recognizes irreducible properties and phenomena and aligns better with the lesson taught by Hempel's dilemma. It does not rely on an unproven faith in future reductive explanations and instead recognizes both complexity and novelty as it is presented to us. If evidence undermines strong emergentism or technological advances reveal the truth of emergent phenomena, then the dilemma will not be between strong and weak emergentism but between emergentism and reductionism. At that point, the strong emergentist would need to

¹²⁶ Essentially, I am setting up a revised version of Hempel's dilemma here. To rehash briefly, Hempel's dilemma challenges us to decide whether to work with our current physics or to hold out our judgments based on a future, complete physics.

reassess and decide to either cling to the old paradigm or accept the new evidence and paradigm.

Emergent Characteristics Examined

With the layered conception of reality, realization, and strong emergence now addressed, we can examine the characteristics of emergence. The goal is to distinguish emergent phenomena from other types of phenomena and to understand emergence better. Notably, the details and rationale that I present help to distinguish my view of emergence from Vision's view.

Irreducibility

Intertheoretical reductions suggest that higher-level phenomena, governed by their theories and principles, can be boiled down to the theories and principles of a lower, more fundamental level. Yet, while these reductions are theoretically possible across the board, practically speaking, it is case-by-case and often a messy affair. Intertheoretical reductions are most successful when they deal with sets of known quantities. The more complete the knowledge of a higher-level phenomenon and the lower-level concepts undergirding it, the easier it is to deconstruct the higher-level concepts and link them to lower-level concepts. In these cases, intertheoretical reduction is akin to translating well-understood higher-level concepts into more fundamental and equally or better-understood lower-level concepts.

The above description traces a nearly idealized look at intertheoretical reductions, but the reality can be far more perplexing, inaccurate, and error laden. For one, since the success of reduction often depends on existing knowledge of higher- and lower-level

phenomena, any shortcoming in our knowledge will negatively impact our ability to reduce with confidence. In cases where our knowledge is lacking, bridge laws may be imprecise and informal, making them more closely resemble simple correspondence.

Typically, lower-quality bridge laws can be refined and enhanced as more knowledge is uncovered, but uncovering additional knowledge is not guaranteed. Also, even with the uncovering of knowledge, bridge laws do not necessarily improve. For example, our understanding of neuroscience has increased dramatically, but attempts to reduce psychological theories continue to fall flat. Undoubtedly, we have a better understanding of specific neural processes (e.g., how neurons transmit signals) and mapping these processes to higher-order psychological constructs (e.g., emotions, personality types, and states of mind). However, despite these advances, a reductive relationship between psychology and neuroscience remains elusive.

Second, intertheoretical reduction is often complex because of the multifaceted nature of the two systems being compared and the intricate interplay between different levels. For example, the theory of evolution should be reducible to genetics.¹²⁷ Yet, genetics will not provide the whole picture since many other factors affect evolution. For instance, genetics does not address a studied population's ecosystem, but the ecosystem will play a key role in how a population evolves. Specifically, climate changes, the extinction of a rival species, and the abundance or scarcity of a food source will affect how the population adapts. Put simply, the theory of evolution cannot be fully explained

¹²⁷ Assuming evolution can be understood as a change in the frequency of genes in a particular species over time. Species is intuitive here to non-biologists like me, but a more accurate description may be a "particular population," since new species might arise due to evolution.

by or reduced to genetics. In support of those points, consider the following explanation from an evolutionary biologist about the adaptations of Darwin's finches. Reid writes:

Darwin's finches are birds; not disembodied beaks, far less genes for beaks. When they first arrived in the Galapagos, their avian adaptabilities enabled them to deal with dry or wet, hot or cold conditions, to eat and digest seeds, fruit, insects, or snails. Founder effect and hormonal destabilization may have strongly influenced epigenetic homeorhesis in the small population of invaders, restoring a primitive plasticity to beak formation until populations filled out and selection took charge, putting a premium on birds with beaks that could keep up with the competition. [...] In contrast to conventional adaptive radiation, Darwin's finches illustrate the constraining effects of competition, and the speed at which adaptability and diversifying experimentation can act in its absence.¹²⁸

Reid's account depicts evolution as a tapestry of interwoven factors beyond pure genetics. Darwin's finches are not just carriers of genetic information but complex organisms interacting dynamically with their ecosystem and many other influences that affect evolution. So, while genetic factors play a crucial role in evolution, they are only part of the complexity. Consequently, reducing the theory of evolution to genetics will fall short because it fails to account for various evolutionary factors.

It is important to note that in this case of two widely accepted and nearly comprehensively understood scientific concepts, solely focusing on intertheoretical reduction does more harm than good. In such cases, reduction oversimplifies the complex reality of reducing one layer of reality to a more fundamental layer. In other words, reduction is a powerful tool, but it has its place, and overdependence on it can overlook important aspects of complex systems. The above example suggests that we must also consider holistic or systems-wide approaches to understand the intricacies of complex systems.

¹²⁸ Reid 2007, 415-6.

A counterargument may be that the issue addressed oversimplifies a complex problem by moving from one special science to another. Rather, if we were to move from the special science directly to the fundamental science (i.e., quantum physics), the complete picture would be available since the quantum level supposedly undergirds everything. The flippant, but I think, fair response to that argument is to encourage the proponent of such a reduction to proceed with reducing the theory of evolution to quantum physics. It has never been done and arguably cannot be done due to its vast complexity and inherent indeterminism.

An intertheoretical reduction is a powerful method for understanding reality because it turns complexity into simpler and more definitive terms. Intertheoretical reductions cannot reduce all higher-level phenomena, but perhaps another type could. Ontological reductions attempt to reduce all phenomena to their fundamental physical elements. This type of reduction holds considerable allure, particularly for physicalists, but they also fail to capture some higher-level phenomena.

Ontological reduction is a tidy view that renders all phenomena as constructs of physical particles. For example, an ice cube is ontologically reducible to water molecules locked in a crystalline lattice structure with low kinetic energy keeping the molecules from breaking free from the structure. Ice cubes can be stacked atop one another because they have the property of solidity, which is not identifiable in the underlying water molecules when in a liquid state, but it can be ontologically reduced to the crystalline lattice structure of the water molecules.

The tidiness of ontological reductions is appealing, especially because it breaks a seemingly complex reality down into easier-to-analyze pieces. However, ontological

reductions struggle to explain the properties and features of all higher-level phenomena. The uniqueness of the higher-level phenomena (e.g., mental properties) or if it is a complex phenomenon that builds upon or is affected by many underlying aspects (e.g., the stock market).

Let us consider pain qualia as an example. Pain qualia defy reducing to c-fibers firing because the uniqueness of the qualitative experience of pain cannot be dissected and attributed to c-fibers firing. Attempts to ontologically reduce qualia is more like dissolving the higher-level phenomena, fundamentally losing key characteristics, and then being unable to find those same key characteristics anywhere within the underlying structure.

The resistance of mental properties and other complex higher-level phenomena to reduce presents as an unavoidable conundrum for physicalists. This ever-persistent conundrum reminds us that we must recognize the complexity of reality, treat its phenomena fairly and in good faith, and attempt to categorize it properly without bias.

Though mental properties, complexity, and emergence are problematic for physicalism because it is entrenched in reductionism, it is important to recognize that reduction is not an outright failure and that there are multiple ways of categorizing reality. Concepts like reduction and emergence each provide insight into the variety of ways that different levels of reality may be interconnected. These concepts guide us to embrace the complexity of reality, allow us to see beyond absolute reductionism, and help us to appreciate the multifaceted and interdependent nature of reality.

Unpredictability

A successful reduction occurs when we confidently replace higher-level phenomena with lower-level phenomena while not excluding or losing essential aspects of the higher level. Reductionism is closely related to predictability—the ability to deduce the unknown from what is known.

Predictability is naturally related to reductionism. If higher-level phenomena can be analyzed and understood in terms of their underlying physical parts, then the various levels of reality are close-knit. Given the relationship between the levels, we should also be able to move from lower to higher levels. Predictability is the idea that we can begin with known objects, principles, and measurements, and from those, we can predict unformed or unknown phenomena.

For example, assume that the formation of a hurricane can be reduced to ocean temperature, humidity levels, atmospheric pressure, and other analyzable physical factors.¹²⁹ When we understand the elements of a hurricane, we should be able to predict the formation of a hurricane by tracking its key parts. Hurricanes may arguably qualify as reducible and predictable in a limited sense, but not all complex phenomena lend themselves to reductionism and prediction.¹³⁰ An essential characteristic of emergent phenomena is unpredictability, but emergent unpredictability is often misunderstood and overextended. Emergent unpredictability does not entail absolute unpredictability where emergent phenomena are complete surprises; instead, it applies predominantly to initial

¹²⁹ For a more comprehensive description of hurricanes and how they are formed, please see Maslin 2013, 48-51.

¹³⁰ Please note that we do not predict hurricanes with one hundred percent accuracy. See Cangialosi 2022 for a report on the National Hurricane Center's predictive model and forecasting success.

occurrences, determining specifics about emergent phenomena and the level of certainty we can have about how emergent phenomena will present themselves.

For example, while we may be able to assume that subsequent generations of humans will experience pain, this so-called prediction is not derived from a deep understanding of the connection between human physiology and the sensation of pain. This type of prediction is more of an assumption based on inductive reasoning. The disconnect in this example is due to the lack of a reductive theoretical framework that would more or less guarantee the presence of pain qualia in structurally similar future generations of humans. To illustrate this point, imagine that we can create a synthetic creature that mimics human physiology and functions. It may be possible for the synthetic human to experience pain qualia, but predicting those mental properties in the synthetic human would not be based on any reductive theoretical framework that connects pain qualia to c-fibers (or synthetic c-fibers). Any prediction must be based on the presumption of similarity (or dissimilarity).¹³¹

Even ignoring the difficulty associated with reduction, deducing higher-level facts is problematic. In these cases, we must venture from what is known to the unknown. Consider, for example, locating an extraterrestrial device that consists of simple mechanical parts that we have recovered in a disassembled but undamaged state. Individually, we may understand the components quite well—they are gears, levers, and so on. The extraterrestrial device may ultimately be as familiar to us as a clock, but one

¹³¹ There would, of course, be individuals who take both sides. Some would claim that the synthetic human feels pain, while others would claim that the synthetic human is more similar to a philosopher's zombie than to a human.

that reflects time in increments other than hours, minutes, and seconds. The device could also be entirely unknown, such as an alien astrological horoscope.

The analogy of trying to reassemble an unknown extraterrestrial machine from a collection of unfixed components helps to illustrate this challenge. Our knowledge of individual components does not guarantee correct assembly. The complexity of reassembly, especially with intricate components, accentuates this problem. Considering that this analogy utilizes simple mechanical parts, hopefully, this helps illuminate how daunting it would be to predict complex phenomena.

Prediction can also be tied back to the idea of a unified science. It is possible to argue that the special sciences are partial glimpses into the unified whole of reality. Proponents of unified science believe that we will develop an all-encompassing physical theory that will eliminate the need for specialized sciences. If a unified science is possible, it should empower us to deduce truths native to all levels based on known facts at any level. Historically, the idea of a unified science has been alluring but elusive.

Consider Kuhn's exposition on progress in the sciences. He writes:

[t]hough science surely grows in depth, it may not grow in breadth as well. If it does so, that breadth is manifest mainly in the proliferation of scientific specialties, not in the scope of any single specialty alone. [...] Perhaps they indicate that scientific progress is not quite what we had taken it to be. [...] We are all deeply accustomed to seeing science as the one enterprise that draws constantly nearer to some goal set by nature in advance. But need there be any such goal? [...] Does it really help to imagine that there is some one full, objective true account of nature that the proper measure of scientific achievement is the extent to which it brings us closer to that ultimate goal?¹³²

¹³² Kuhn [1962] 2012, 169-70. For additional remarks on scientific progress, special sciences, theory replacement, and other notes of interest regarding progress toward a unified science, see Kuhn 2012, passim.

Kuhn, and other philosophers of science, argue that scientific progress does not necessarily mean we are getting closer to a single, objective truth about reality. If Kuhn's discussion reflects some greater truth of reality, then a unified science may never be discovered.

Beyond Kuhn's landmark exploration of revolutions in the sciences, several issues complicate the possibility of a unified science. For one, deducing higher-level facts from lower-level facts requires increasingly complex calculations. Take, for instance, Feynman's explanation of the difficulty of doing complex quantum calculations with a classical computer. He writes:

[...] let's suppose that a particle has a probability $P(x, t)$ to be at x at a time t . [...] Now we could discretize t and x and perhaps even the probability itself and solve this differential equation like we solve any old field equation, and make an algorithm for it, making it exact by discretization. First there'd be a problem about discretizing probability. [...] If we had many particles, we have R particles, for example, in a system, then we would have to describe the probability of a circumstance by giving the probability to find these particles at points x_1, x_2, \dots, x_R at the time t . That would be a description of the probability of the system. And therefore, you'd need a k -digit number for every configuration of the system, for every arrangement of the R values of x . And therefore if there are N points in space, we'd need N^R configurations. Actually, from our point of view that at each point in space there is information like electric fields and so on, R will be of the same order as N if the number of information bits is the same as the number of points in space, and therefore you'd have to have something like N^N configurations to be described to get the probability out, and that's too big for our computer to hold if the size of the computer is of order N .¹³³

As Feynman explains, the complexity of these calculations is exceedingly great, and there is a high risk of error that reduces predictive confidence. This complexity only increases

¹³³ Feynman 1982, 471-2. Please note that though Feynman was skeptical about the ability of classical computers to handle this complexity, he was optimistic about quantum computers being able to handle this complexity. That said, forty years have passed since this publication, and quantum computers are still nascent and unable to perform these calculations.

as one attempts to calculate higher-level probabilities. Put simply, the higher one goes, the greater the complexity. As complexity grows, the possibility of error increases, and predictive confidence must decrease proportionally. One might reasonably postulate that the complexity involved in moving from quantum information to meteorology would be so complex that even precise quantum information could lead to uncertain meteorological predictions.

Additionally, inherent uncertainty built into reality, notably at the quantum level, confounds our ability to deduce all higher-level truths. This uncertainty, as characterized by Heisenberg's uncertainty principle, states that the position and momentum of a particle cannot both be precisely measured. Polkinghorne explains uncertainty as follows:

[t]here is an inescapable trade-off between the increasing accuracy of position measurement and the decreasing accuracy of knowledge of momentum. This fact is the basis of the uncertainty principle: it is not possible simultaneously to have perfect knowledge of both position and momentum. [...] This demi-knowledge is a quantum characteristic. Observables come in pairs that epistemologically exclude each other.¹³⁴

Put simply, the more accurately we know a particle's position, the less accurately we can know its momentum, and vice versa. In more colorful language, Polkinghorne notes that "one can know where an electron is, but not what it is doing; or one can know what it is doing, but not know where it is."¹³⁵ This introduces inherent unpredictability in the calculations required to move from lower to higher levels.

Lastly, there is a greater dependence on using approximations and probabilities when dealing with fundamental sciences and when performing interlevel calculations.

We are limited in our ability to know exact details, such as the location or momentum of

¹³⁴ Polkinghorne 2002, 33.

¹³⁵ Ibid.

particles in a mixture of liquids, so we must rely on approximations and discretization when performing these calculations. Utilizing approximations undermines the confidence we could have when deducing higher-level truths because the metrics used to make the prediction are imprecise. As we progress from fundamental principles to complex phenomena, these approximations can compound, introducing wide-ranging inaccuracies in our predictions. Consequently, this makes our attempts at deducing higher-level truths from fundamental principles unreliable and uncertain.

Turning now to the emergent mind, we widely maintain that it emerges from the brain. The brain is a vast complexity of 100 billion neurons and around 100 trillion connections.¹³⁶ Given the magnitude of this complexity, one can easily see the potential for unpredictability based on Feynman's explanation. Moreover, one can also conceive of the apparent irreducibility between the mind and the brain, considering not only because of complexity but also because of the possibility of indeterminacy found in such a vast complexity.

However, the complex layered conception that makes up reality does not undermine the possibility that reality is a unified whole. For instance, consider Neurath's writing on the relationship between a unified reality and a unified science. He writes:

[...] we do not arrive at 'one' system of science that could take the place of the 'real world' so to speak; everything remains ambiguous and in many ways uncertain. 'The' system is the great scientific lie. Not even as an anticipated goal is it a useful guiding thought as it takes us close to Laplace's spirit which, it is thought, with knowledge of all the equations of the sciences, constantly makes correct predictions: an assumption that

¹³⁶ O'Shea writes, "your brain—an organ of just 1.2 kg, containing one hundred billion nerve cells, none of which alone has any idea who or what you are. In fact the very idea that a cell can have an idea seems silly. [...] However, conscious awareness of one's self comes from just that: nerve cells communicating with one another by a hundred trillion interconnections." O'Shea 2005, 1.

serves no prediction, an assumption that cannot be verified in any way; [...]. Multiplicity and uncertainty are essential. [...]; the multiplicity of predicting cannot be excluded by any method; no degree of systematic procedure can alter this. [...] The progress of science consists, as it were, in constantly changing the machine and in advancing on the basis of new decisions. Still, the result in fact is far-reaching unity that can not be deduced logically.¹³⁷

I interpret Neurath to mean that despite our inability to unify science, likely for the reasons I list above, our best understanding of reality, informed by many specialized sciences, suggests there is a unity to reality. Said unity cannot be addressed in a single science, but we should not be discouraged by the necessity to carve up reality into layers and to understand it through various methodologies. The layered conception only precludes the possibility that we can understand the unity of nature in a single methodology because uncertainty is inherent to the system, and our abilities to approach interlevel understanding are impeded by the complexity I address above.

Arguably, the issues addressed above may point toward an epistemological limitation that we have and not any metaphysical differences between the layers. An epistemological understanding of layers helps to explain why deductive success is higher when the behaviors or properties of higher-level phenomena strictly conform to and are accounted for by lower-level principles. Since our views on abutting levels are conceptually near one another, this interlevel work is more precise. When there is greater complexity between levels, whether adjoining or distant, there is a wider gap between levels, introducing conceptual drift and the necessity of relying on approximations.

I agree that the complications discussed could be due to an epistemological limitation; however, the appearance of an epistemological limitation could also be due to

¹³⁷ Neurath 1983, 116.

the metaphysical starkness between layers. Put simply, if there were metaphysical distinctions, we would need different methodologies to understand metaphysically distinct phenomena and properties. I believe that emergence is not a case of a failed reduction due to epistemological limitations.

Emergence occurs when there is enough complexity (and possibly indeterminism) inherent in a system, conjoined with the presence of novel properties. If emergent characteristics and novel traits are irreducible, there is reason to find emergent phenomena metaphysically significant and distinguishable from their underlying parts or levels. Moreover, I believe it is possible for this to be the case without metaphysically separating emergent phenomena from their bases. In other words, without rejecting physicalism's metaphysical thesis. In the following subsection, I discuss the significance of novelty.

Ultimately, we do not possess a unified theory of everything and have no reason to believe in its imminent discovery. From a metaphysical perspective, one might be committed to reality being a unified whole, but epistemologically, the unity of science is challenged by complexity and uncertainty. Ultimately, when it comes to the unity of science, it does not matter whether that complexity and uncertainty result from epistemological limitations or because they are metaphysically embedded features of reality. The layered conception is how we best understand reality, and our best evidence points to irreducibility and unpredictability being baked into the system; at least, that is the case with phenomena described as emergent. That said, there is metaphysical significance to emergent phenomena that I cover in the following subsections.

Novelty

A key characteristic of emergent phenomena is the presence of novel properties that are irreducible and unpredictable from their underlying components. Possessing irreducible and unpredictable novel properties would mean that emergent phenomena are distinct from their underlying base. Emergent phenomena are often called novel because they display characteristics or properties that cannot be found or derived from where they emerged.

Although rooted in the complexity of their underlying parts, these properties exhibit new behaviors and traits unexplained by underlying physical parts and principles. For example, the mental properties that emerge from complex neural activity cannot be reduced to, predicted from, or described by the physical interactions underlying mental activity. To utilize the emergent slogan, the whole is greater than the sum of its parts.¹³⁸

The emergence of novel properties extends beyond unpredictability and complexity. To acknowledge that higher-level emergent properties are more than a complex arrangement of lower-level phenomena (i.e., greater than the sum) is to recognize the existence of something new. In other words, if emergent phenomena present new properties and traits that cannot be reduced, we have reason to find emergent phenomena to be metaphysically distinct.

Emergent phenomena with irreducible, unpredictable, and novel characteristics are more than epistemologically different; they are also ontologically distinct. It follows

¹³⁸ This phrase is often attributed to Aristotle. In *Metaphysics* VIII, he writes, “In the case of all things which have several parts and in which the whole is not, as it were, a mere heap, but the totality is something besides the parts, there is a cause of unity; for as regards material things contact is the cause in some cases, and in others viscosity or some other such quality” (*Metaphysics* VIII, 1045a8-1045a12). Aristotle (1984) 1995, 1650.

that emergent phenomena do not represent mere conceptual distinction nor an epistemological complexity that we cannot untangle. Instead, emergent phenomena represent a new order of reality that cannot be fully explained or predicted. Ontologically distinct emergent phenomena severely harm absolute reductionism and the quest for a unified science. In its place, I suggest we need a multimodal approach that takes on multiple perspectives.

One major challenge to novel and ontologically distinct emergent properties is the idea that novel emergent properties may be latent properties of the underlying physical structure. For instance, Shoemaker writes:

[c]ertainly the [emergent] properties of macro-entities will not be predictable on the basis of, and resultant relative to, the micro-facts that constitute the instantiation of the physical micro-structural properties. But they will be realized in, and resultant relative to, the micro-facts that constitute the instantiation [...]. They will not be predictable on the basis of these micro-facts, prior to the micro-entities being combined in emergence engendering ways, simply because these facts cannot be known prior to our observing the effects of combining the micro-entities in such ways. But it is as much true on the emergentist view as it is on the view of an anti-emergentist physicalist that the micro-facts fix the macro-facts; the only difference is that on the emergentist view the micro-facts include the instantiation of micro-latent powers.¹³⁹

Under this view, novel emergent properties exist in the underlying structural components but are otherwise hidden. According to this view, these latent abilities manifest only under complex configurations of the underlying physical structure. The idea of latent powers allows for unpredictability and irreducibility, but it also dilutes emergentism. In essence, latent powers bring emergence closer to physicalism by eliminating novel properties in favor of identifying those higher-level properties with latent powers. By

¹³⁹ Shoemaker 2007, 75.

eliminating novel properties, the metaphysical distinction of emergent phenomena becomes suspect.

I find latent powers to be a flawed position for several reasons. First, the idea lacks explanatory power. The idea that there are latent powers puts these powers in conflict with regular or manifest powers. There is an evident ambiguity delineating manifest and latent powers. For instance, what explains some powers' manifestation and others' latency? Such a position demands a clear explanation for *why* certain powers remain hidden but also *how* latent powers become manifest when necessary conditions are met. Without satisfactory answers to these questions, latent powers appear to be a theory with a biased agenda that dogmatically supports reductive physicalism.

Second, as a way of galvanizing physicalism, the view inadvertently undermines the scientific enterprise that physicalism is supposed to champion. Suggesting that novel powers and abilities are just activated powers that were latent in the underlying structure is to propose an explanation divorced from both observational and coherent confirmation. There is no way to observe latent powers because they would be non-causal in their latent form. Since they are non-causal while latent, they cannot play a coherent role in explanations at the level where they exist in their latent form. These so-called latent powers would only be observed and coherent when they are activated at a higher level. Such a view also begs where latent powers should be metaphysically attributed. For instance, activated latent powers could reasonably fit at the higher level, yet the claim that they are latent attempts to attribute them to the lower level metaphysically. This strategy appears ill-founded.

Latent powers appear to me as a rebranding of emergent properties based on nothing more than a preference for a physicalist and reductively motivated worldview. Given that the powers are latent, there is no empirical rationale to find the powers to exist within the underlying structure other than a preference for the powers to be in the underlying structure. Moreover, such a position offers no additional explanatory value, only a clear bias for physicalism and reductionism.

Furthermore, the position ultimately makes both the higher and underlying levels unstable. Our understanding of the underlying layer, including its objects and principles, is rendered doubtful because there are undetectable latent abilities supposedly waiting to manifest. On the other hand, knowledge of higher-level complex systems appears to be arbitrary knowledge. According to the idea of latent powers, the novel features of higher-level phenomena are part of an underlying physical structure hidden from us. Furthermore, because the higher-level phenomena are irreducible to the underlying latent powers, the ability to connect the layers becomes even more troubled.

The concept of latent powers within microphysical structures presents an intriguing idea that would play a key role in undermining the metaphysical significance of emergent phenomena. Ultimately, my position is that latent powers are a biased effort to undermine emergentism and support reductionism rather than a genuine attempt to explain novel properties. Latent powers face significant conceptual and practical challenges. Given the absence of direct evidence for them and the fact that higher-level powers cannot be reduced to these latent powers, this view appears inherently problematic.

So long as emergent phenomena have irreducible novel properties, they have a strong claim to being metaphysically significant and something beyond a mere epistemological distinction. Notably, those novel properties could be causally efficacious, granting metaphysical significance by way of Alexander's Dictum. Emergence reveals a reality that is layered rather than monolithic. The irreducibility and unpredictability typically associated with emergent phenomena suggest more than just epistemological gaps; they suggest a genuine metaphysical distinction between emergent phenomena and their underlying base. The novel characteristics of these phenomena reinforce this distinction. In the following subsection, I will further elaborate on the significance of this ontological differentiation and why it warrants serious consideration.

Downward Causation

In the previous subsections, I argued that emergent phenomena are metaphysically distinct because they are irreducible, unpredictable, and novel. Metaphysically distinct emergent features suggest that novel properties may be causally efficacious, making emergent features as metaphysically significant as solidity is to ice cubes. In this subsection, I discuss several examples that attempt to illuminate the possibility of emergent phenomena having downward causal abilities. To ease skeptics into the possibility of downward causation, none of the following examples deal with mental properties.¹⁴⁰

Extreme weather patterns, such as hurricanes and tornadoes, are sometimes considered emergent phenomena or at least serve as useful analogs of emergent

¹⁴⁰ The downside of this approach is that there are few things, if any, in nature that closely resemble mental properties and their supposed downward causal abilities. That said, I recognize the potential for disanalogy but one has to start somewhere.

phenomena. Under this interpretation, these weather events are irreducible, unpredictable (in the way I have described earlier), and possess novel features. For instance, a key characteristic of a hurricane is the rapid intensification of wind speeds. This rapid intensification is not seen under weather conditions where only some of the ingredients of a hurricane are activated. The irreducible nature of a hurricane's novel features emphasizes that they cannot be fully explained or broken down into lower-level components. Hurricane winds have real causal power. Since the phenomena themselves are irreducible, any causal effects attributed to the emergent phenomena will not be identifiable at a lower level, and, therefore, their causal power cannot be drained away.

Hurricanes form under specific weather conditions, including atmospheric pressure conditions. As a hurricane forms and grows in intensity, the eye lowers in pressure while pressure increases around the outer perimeter of the hurricane. A fully formed hurricane will affect its underlying parts, including moving wind particles around in ways they otherwise would not have moved. Additionally, as the hurricane travels, it will continue manipulating atmospheric pressure to sustain itself, lowering pressure systems throughout its path.¹⁴¹ In effect, the hurricane has both horizontal and downward causal abilities. It affects both atmospheric conditions and its own underlying ingredients, respectively.

A country's economy is another example that can be invoked when discussing emergence. The behavior of any given economy and our evaluations of the state of that economy cannot be wholly reduced to the components that make up the economy being examined. Those components include monetary inflation, government policy on tariffs

¹⁴¹ See Dunlop 2017, 45-51 for a clear explanation of the formation and dissolution of hurricanes.

and taxes, consumer spending power, and job reports. The economy is also unpredictable in a variety of ways. One such way is that positive news in any one component area does not necessarily result in an overall improvement in the economy. Lastly, the state of the economy arguably has causal power, such as leading companies to lay off employees or expand their workforce, shrink the size of products, or expand their business in various ways.

Many components come together to form an emergent economy, and that emergent economy arguably has downward causal effects. One example is that the economy is partly a function of prime interest rates. As the economy evolves into something promising or unpromising, it can downwardly affect interest rates. Sometimes, manipulating interest rates will affect the economy in the intended way, but occasionally, it will not do so as expected or take longer than anticipated. The two-way street of upward and downward causation continues as the economy evolves. Another example is that the economy is partly a function of the unemployment rate. The unemployment rate will contribute to the economy, but the economy will also have a downward causal effect on the unemployment rate, causing some companies to freeze, reduce, or increase hiring.

As a final example, consider network science and how interconnected computing nodes lead to behaviors and properties of the networked whole that are not seen at each node. For example, it is highly improbable, if not impossible, to predict the dynamics of the Internet based on understanding various network nodes or individual human-computer interactions. The interactions among the lower-level entities (i.e., individual human-computer interactions) lead to higher-level phenomena (i.e., Internet dynamics). To be more specific, the behaviors engendered on a particular social media platform are

generally not behaviors that we would see in an individual who is not on a social media platform.

Social media platforms depend on many lower-level parts, and the higher-level phenomena—the social media platform and its dynamics—directly impact the underlying structure. In other words, the unique culture that emerges on the platform will affect how individuals behave, and individual behavior is what makes up the social media network. Once again, we see a dynamic of upward and downward causation. The human-computer nodes act jointly and independently to form a social media culture, which affects individuals in profound and often strange ways. Interestingly, the downward causal effects of the social media platform will have cascading effects in the real world as well. The prevalence of rampant conspiracy theorists, flat earth theorists, election deniers, and so on, despite decisive evidence to the contrary, can be more clearly attributed to a social media platform than to any individual parts.

I have asked readers to consider all these examples to be cases of emergent phenomena with downward causal abilities.¹⁴² I recognize that many may find this an adverse position to take and may contend that these so-called emergent phenomena can be reduced to underlying physical processes. For example, they may prefer to believe that hurricanes, while complex, can be fully understood in terms of their underlying factors and a powerful enough supercomputer running the correct formulas. Should that be the case, then the supposed causal efficacy of these higher-level phenomena would be

¹⁴² Moreover, in addition to emergent phenomena being able to downward affect its own components, the several examples I shared also shared the dynamic causal interaction that occurs in both directions when dealing with emergent phenomena. In Chapter 4, I address dynamic causal holism in more detail and explain why it is an essential component of the layered conception of reality.

derived from the fundamental physical interactions of its underlying factors. Furthermore, the claim that these features lead to causal relevance might be seen as a misunderstanding of the nature of these underlying interactions. Lastly, critics may insist that any description of irreducibility is actually due to epistemological limitations and not metaphysical distinctions between the underlying factors and the emergent phenomena.

These counterpoints imply that even if we see complexity, emergent phenomena, and downward causation, all higher-level phenomena are reducible to the fundamental level, so all of these examples can (eventually) be explained at the fundamental level. However, to make that argument is to reject the difficulties listed in earlier subsections with no additional rationale as to why we should reject emergence. To reject that open-mindedness and those complications without new reasons to do so must be based on a dogmatic belief in reductionism or the eventual unity of science. But even if there is the desire to reduce, doing so remains elusive and potentially impossible.

Currently, we have no unified science, only the desire for one. Also, there seem to be several irreducible and unpredictable phenomena with novel features that are characteristically like emergent phenomena, such as hurricanes, economies, and the dynamics of social media platforms. Given the difficulty of reducing and finding a unified science, the existence of higher-level phenomena with emergent traits (i.e., irreducible, unpredictable, and novel) provides an excellent reason to keep an open mind about emergence and downward causation. However, for those who plant their flag starkly behind absolute reductionism, I do not see a path toward convincing them to believe in emergent phenomena and downward causation. The remainder of this work can only address those open to emergence and downward causation.

Emergent Mental Properties

We can summarize the minimum characteristics of emergent phenomena from the above descriptions. Emergent phenomena are (i) irreducible, (ii) unpredictable, and (iii) novel. Mental properties meet the three criteria mentioned above and are a key example of emergent phenomena. While the first two chapters addressed the irreducibility of mental properties, I will briefly recap with a focus on emergence.

To be irreducible means that specific properties or phenomena cannot be reduced to or satisfactorily explained by their physical underpinnings. In the case of mental properties, we cannot reduce or explain these properties solely in terms of the physical. For example, neuroscience may be capable of identifying which brain areas are active during mental experiences, but neuroscience falls considerably short of explaining the mental state itself through reductive means. For one, the subjective quality of mental experiences (i.e., qualia) cannot be reduced to physical states. Second, even if we observe repeatable patterns of brain activity associated with a particular mental state, such observations provide limited insight into the nature of these mental states. In addition to these two considerations, the multiple realizability of mental states needs to be addressed (see Chapter 2) as well as the issues associated with complexity discussed earlier in this chapter further complicates the relationship between emergent mental properties and physical states.

In the context of emergence, unpredictability means that the emergent mental phenomena cannot be deduced from a close understanding of the underlying physical processes. Another way of looking at unpredictability is that emergent phenomena are not merely equivalent to their underlying physical processes. Since the whole is more than

the sum of its parts, emergent phenomena cannot be deduced (i.e., predicted) from their base components. That means that even a comprehensive understanding of neurological activity and an idealized grasp of physical principles at all levels would not result in predictable mental experiences.

The inherent unpredictability of emergent phenomena works particularly well with mental properties because it explains why there is an explanatory gap problem. There is an explanatory gap because there is, in fact, a metaphysical gap between the underlying physical base and the emergent phenomena. Emergent mental experiences are more than a different conceptualization of an underlying physical event or set of properties. They are distinctive because of their irreducibility, unpredictability, and novel characteristics. Where the underlying base can be objective and public, emergent mental properties are understood subjectively and privately. As such, they do not reduce to or identify with underlying states (e.g., brain states); they are more than the sum of those parts.

As metaphysically distinct emergent phenomena, mental properties would possess novel features. For emergentists, novelty refers to new properties, traits, or abilities that are not in (or locatable in) the underlying physical processes. Once again, the qualitative nature of mental properties speaks to the novelty of these properties. Consider the often-used example of pain qualia. A comprehensive understanding of the nervous system cannot reveal the inherent features of pain qualia. Whether the experience of pain qualia is nagging or agonizing, bodily or psychological, the experience of pain is unique and innately novel to the experiencer. To the experiencer, pain is categorically unlike c-fiber stimulation. A pain quale is not equivalent to a chemical-electric signal or a specific type

of physical damage; it is a private experience marked by its qualitative and personal features.

As a final illustration of the emergent features found in mental states, consider the intentionality of mental states. Intentionality is an irreducible, unpredictable, and novel ability of mental states that allows one's thoughts to be focused on something beyond the mental state itself. When a mental state attends to an idea or thing outside of the mental state itself, the intentional mental state fixates on something that leads to qualitative experiences and could put one on a productive path toward achieving an objective. For example, when thinking of espresso, one's thoughts are focused on something outside of them, and those thoughts may lead to seeking espresso. If strong emergentism is onto something, it could mean that emergent mental phenomena have downward causal powers.

Critiquing Emergence

The major criticisms of emergence tend to focus on its mysteriousness. There are two general categories of mysteriousness. Emergence is mysterious because (i) emergent phenomena are irreducible and unpredictable, which conflicts with scientific idealism, or that reality is uniform and discoverable throughout, and (ii) emergence is an unknowable process and must rely on brute facts to explain the existence of emergent phenomena.

Scientific Idealism

The first category of criticism has already been partially addressed. Emergent phenomena are irreducible, unpredictable, and novel by their very nature. Finding fault in these characteristics challenges emergence based on conflicting beliefs about reality,

namely that reality is reductive and discoverable throughout. However, disagreements about the essence of reality must be based on evidence, not merely faith in one's worldview or blind faith in scientific progress that should eventually vindicate one's worldview.

At this point, Hempel's dilemma must be invoked to challenge the idea that reductive physicalism will eventually be vindicated.¹⁴³ Reductionists and reductionist-motivated thinkers wish to reduce emergent phenomena and neatly organize reality. However, the best evidence we have suggests that there are indeed phenomena that are irreducible, unpredictable, and novel.

So long as there are phenomena with the characteristic traits of emergence, and the existence of said phenomena is not in question, then the denial of emergence would be based on faith in future breakthroughs that will allow for a proper reduction. This position banks on new scientific theories, innovative technologies capable of revealing unobserved information, or improved computing power that would enable us to unravel the complexity of nature. Alternatively, some may clutch to a metaphysical belief that reality is fundamentally integrated. Individuals who hold this metaphysical position may admit that we cannot unravel the complexity but remain convinced of their metaphysical position regardless.

Denying our best evidence in favor of a future breakthrough or on a hardline metaphysical stance is to do so on a matter of faith and not an evidence-based view of reality. Taking one of these two anti-emergentist stances can be seen as a refusal to acknowledge the evidence that shapes our understanding of reality. In essence, it is a

¹⁴³ Hempel 1969.

denial of the evidence that leads us to these worldviews. Rather than address this argument borrowed from Hempel any further, I will offer additional considerations for why we should take emergence seriously.

One can contextualize the place of emergence in our scientific understanding of reality by placing it within Kuhn's stages of scientific revolution.¹⁴⁴ The first stage consists of normal science, where scientists and practitioners have reached a consensus agreement about how to describe the nature of reality. Eventually, cases or phenomena that are inconsistent with normal science appear. These anomalies present themselves as problems that must be solved. Typically, proposed solutions fall within the context of normal science, but revisions to normal science may be required. If the anomalies accumulate and cannot be accounted for in a revised version of normal science, confidence erodes in the existing scientific system. New theories and models may be proposed to account for the anomalies. Typically, the old guard resists radical theories, making wide-scale adoption of the new theory an uphill climb requiring considerable evidence and explanatory power. If the new theory eventually succeeds, it becomes the new normal science.

Our reigning scientific paradigm, or normal science, is dominated by the methodology of reductionism. We have reason to believe in the existence of irreducible, unpredictable, novel phenomena that challenge reductionism. These phenomena are anomalies that our normal science cannot account for, and our best attempts to lightly revise normal science have also failed to reduce these anomalies. Emergence presents itself as an alternative to reductionism, suggesting that some phenomena are more than

¹⁴⁴ Kuhn 2012, passim.

the sum of their components and that there are metaphysical distinctions among emergent phenomena.

When looking at emergence through the lens of the Kuhnian cycle, it seems that we are in the anomaly accumulating stage. As more special sciences begin invoking emergence to explain complex irreducible phenomena, emergence will continue to demand attention. However, even if emergence gains broad-based acceptance, skeptics will seek to discredit it, clinging to reductionism. The difficulty for skeptics is that any alternative to emergence based on reductionism must dissolve the so-called mysteries of emergence and allow for the reduction of emergent phenomena. So long as emergent phenomena remain irreducible, emergence remains the best explanation available.

Without a reductive resolution, the reductionist must grapple with our repurposed use of Hempel's dilemma. Either accept the evidence available or convert the basis of their beliefs about reality to a matter of faith.

Emergence is a challenge to traditional scientific views that operate entirely on reductionism. Emergence offers a complex perspective on the nature of reality, accepting the value of reduction but denying its universality. Resisting emergence may be more of a reflection of the desire for order and predictability than it is of a legitimate, evidence-based critique. By acknowledging the irreducibility, unpredictability, and novelty of certain phenomena, we accept that reality is too complex for a single methodology or worldview to explain.

Ceteris Paribus Laws

The indeterminacy of emergent phenomena could be difficult for reductionist-minded philosophers to accept. However, the indeterminacy of emergent phenomena can be explained by a concept from the philosophy of science known as *ceteris paribus* laws. Cartwright, a philosopher of science, is arguably the leading figure discussing *ceteris paribus* laws. She writes:

[...] true of all our laws, whether we take them to be iron—the typical attitude towards the laws of physics—or of a more flexible material, as in biology, economics or psychology. J.J.C. Smart has urged that biology, economics, psychology and the like are not real sciences. That is because they do not have real laws. Their laws are *ceteris paribus* laws, and a *ceteris paribus* law is no law at all. The only real laws are, presumably, down there in fundamental physics. I put an entirely different interpretation on the phenomena Smart describes. [...] If the topic is laws in the traditional empiricist sense of claims about necessary patterns of regular association, we have *ceteris paribus* laws all the way down: laws hold only relative to the chance set-ups that generate them.¹⁴⁵

At its core, the concept of *ceteris paribus* laws suggests that absolute laws of nature that apply universally do not exist. The Latin phrase *ceteris paribus* means “other things being equal.”¹⁴⁶ I further interpret the *ceteris paribus* in *ceteris paribus* laws to imply that the so-called laws of nature are not universal rules that drive events and their outcomes; rather, there are relevant variables and conditions that are crucial to an event’s outcome. In other words, we are not dealing with laws but the complex interplay of conditions and variables (i.e., situational stipulations) that influence outcomes.

If my interpretation of *ceteris paribus* laws is well-founded, then the regularity and predictability we observe in nature can be attributed to the essential underlying

¹⁴⁵ Cartwright 1999, 175-6.

¹⁴⁶ Kidd and Wade 2003, 3.

conditions and variables remaining consistent across instances. Put simply, observed regularity would be due to the essential variables and conditions tied to an event being largely maintained, allowing for recurrent patterns to manifest. Assuming Cartwright is correct in her assertion that *ceteris paribus* laws go all the way down to fundamental physics, then emergent phenomena could be closely tied to *ceteris paribus* laws.

As we know, emergence is powered by the complex interplay of variables and conditions. We have limited insight into the conditions and variables that are essential to any given emergent phenomenon, but we have little insight into the extent of the role played by each essential variable, and we have no understanding of how the variables lead to emergent phenomena.

There is an inextricable relationship between complexity and emergence. Because of this, identifying all the variables that shape and influence emergent phenomena is incredibly challenging, if not impossible. Furthermore, as the underlying variables change, emergent phenomena can be affected in unknown ways. This would ultimately entail unpredictability in the emergent phenomena while requiring an essential—but irreducible—connection to the underlying variables. The likelihood of maintaining consistent conditions and variables that underlie the emergent phenomena would be low. Moreover, the likelihood of an expected outcome would decrease with the amount and rate of variation.

The notion that we cannot reduce a complex system's properties to its essential parts challenges conventional scientific belief. Traditional science often views reality's layers as conceptual and reducible to essential elements. Emergence points to a more intricate reality than previously thought. It implies that reductionism alone cannot capture

the full picture. Reduction as a methodology is incredibly important because of its explanatory power. However, it fails to explain certain higher-order phenomena, such as the mind and its mental properties. Emergence explains irreducible higher-order phenomena by appealing to the complexity of reality. The complexity of reality, further supported by the layered conception, can point to various variables that affect outcomes.

Ceteris paribus laws also help us to conceptualize unpredictability and irreducibility further. If there is less complexity between two layers, then fewer conditions need to match between various instances, making outcomes more predictable. If there is great complexity, there is great potential for uneven conditions between instances, leading to unpredictability. Moreover, adding indeterminacy into the equation would significantly undermine predictive confidence. Thus, *ceteris paribus* laws entail that we may be able to reduce and predict some things confidently while others may be impossible to predict or reduce.

One can be reductionist-minded while recognizing the shortcomings of absolute reductionism. Such individuals should find emergence appealing. While emergence is not reductive, it has reductive elements. For instance, emergence's views about the essential nature of underlying variables, and the higher-level effects that occur due to changes in those variables should appeal to reductive-minded individuals. Moreover, emergence should be appealing because to believe in the phenomena is not to reject reduction altogether. Emergentism is not meant to replace reductions as a methodology; rather, it supplements reductionism in cases where phenomena are irreducible, unpredictable, and novel.

The widespread success of reductionism and the evidence and persistence of irreducible, unpredictable, and novel properties demand something more than absolute reductionism. As methodologies, reduction and emergence are not mutually exclusive, they are both necessary for understanding our complex reality. Reduction and emergence work in tandem as a hybrid approach that embraces reduction as the default methodology but understands that it cannot account for the whole picture. When reductionism fails because it runs into irreducible, unpredictable, and novel phenomena, we can turn to emergence.

The moral of the story is that when reductions fail without progress toward an eventual reductive explanation, the responsible approach is not to become further entrenched in absolute reductionism but rather to explore other ways of explaining irreducible phenomena. Entrenching oneself in reductionism leads to unclear explanations, such as the types of mental-physical relationships examined in Chapter 2. Those relationships must ignore our apparent experiences of mental properties and their capacities, forcing mental properties to comply with reductionism. This path results in mental-physical relationships that fail to approach mental realism in good faith and instead treat it as an inconvenience that must be addressed to keep the reductive physicalist agenda on track.

Physicalists often lean heavily into reductionism, and the mental-physical relationships they offer reveal that bias. As I have argued in Chapters 1 and 2, physicalists typically find mental properties to be either conceptual, identifiable with physical phenomena, or in some way metaphysically lesser than that which is self-evidently physical. In other words, the mental-physical relationships they provide do not

seek to give metaphysically robust explanations for mental properties. Rather, they seek to couch mental properties into a reductive physicalist framework.

The uncompromisingly reductive approach taken by many physicalists would not be the appropriate approach to any known phenomenon that does not reduce. Such phenomena must eventually be addressed in good faith. When reductionism cannot bridge the gap, we must explore other ways to explain said phenomena. Given that emergence has gained traction in various scientific disciplines, it should not be hastily dismissed by reductionist-minded philosophers. In other words, there are enough anomalies, evidence, and traction to consider mental properties from a perspective beyond reductionism, and emergence is the next best suitor.

To recognize the complex interplay of emergence, reductionism, and *ceteris paribus* laws is to take a nuanced view of the nature of reality and our understanding of it. *Ceteris paribus* laws are part of an explanatory framework that indicates why emergence is unpredictable. Namely, there is a vast complexity of variables that affect emergent phenomena in unknown ways. Complexity, indeterminacy, and *ceteris paribus* laws challenge absolute reductionism.

In contrast to the absolute reductionist view, I propose a hybrid approach that recognizes the widespread success of reduction as a methodology without denying irreducible, unpredictable, and novel phenomena. Ultimately, I find that reduction, emergence, and *ceteris paribus* laws are intertwined, and we have good reason to resist absolute reductionism. In its place, we must accept the complexity of reality and take a more comprehensive approach to understanding it.

Mysteriousness

One of the key characteristics of emergence is the absence of a technical mapping or detailed explanation that elucidates how a property emerges from complexity. Strong emergentists, believing in the emergence of ontologically novel phenomena, can point to one of two naturalistic explanations. Either (a) the process of emergence is opaque (i.e., there is a process of emergence, but it is naturally obscured or incomprehensible to us), or (b) the relationship between the base and the emergent phenomena is metaphysically brute. Whether (a) or (b) is a more accurate descriptor, the outcome remains the same: irreducible, unpredictable, and novel properties have found a legitimate place in our naturalistic and scientific understanding of reality.

While these two explanations are acceptable to emergentists, reductionists find them unpalatable. In place of opacity and bruteness, reductionists will instead opt to explain emergence through a temporary mystery that will eventually be resolved. Moreover, for many reductionists, accepting opacity or bruteness in cases of so-called emergence would represent a failure of the reductive scientific enterprise.

Emergentists, on the other hand, highlight the utility of emergence for describing phenomena with no reductive explanations. Emergentists can counter that banking on scientific discovery to unravel emergence is to tread upon Hempel's dilemma. Instead, they maintain that emergence must be taken at face value. Regarding bruteness, Crane notes that the hard problem for nonreductive physicalists is a brute fact for emergentists. He writes a "brute fact that must be swallowed with what Broad called the 'philosophic jam' of 'natural piety.'" This is an aspect of what we can think of as the *epistemological*

attitude of emergentism, not its *metaphysical content*.”¹⁴⁷ He goes on to note that “we would do well to consider favorably the epistemological attitude of emergentism as one of the available options for ‘bridging the explanatory gap.’”¹⁴⁸

To further understand the debate between mysteriousness and the emergentist perspective, consider Vision’s explanation of bruteness. He writes:

[it] has been suggested that resting emergentism on the absence of explanations transforms a putatively metaphysical issue into an epistemological one. The inclusion of the ontological conception of explanation and its gloss in terms of the brute nature of the supervenient dependence are inserted precisely to prevent this misunderstanding. But even an ontological conception of explanation should come with the proviso that the explanatory state of affairs be understandable in principle. To complete the case, it should be added that the intended lack of an explanation is the result of (a) the world’s not containing one rather than the result of (b) limitations on human cognitive capacities preventing our species and its successors from being able to grasp a genuine explanation, [...]. It is the world rather than the class of rational beings that lacks the wherewithal for explanation.¹⁴⁹

Vision posits that the mysteriousness of emergence might be misconstrued. He suggests that emergentism is not something that we do not yet understand (i.e., an epistemological concern), but it could point to real, inherent characteristics of reality (i.e., a metaphysical distinction). Vision addresses the “ontological conception of explanation” as actually referring to the nature of reality and not simply our understanding of it. Vision ultimately lands on emergence being a metaphysical issue, where there is a genuine inexplicability built into emergent phenomena.

Building on Vision’s ideas, I agree that the ontological conception of explanation is the best approach to understanding emergence. However, my rationale diverges

¹⁴⁷ Crane 2001, 210.

¹⁴⁸ Ibid.

¹⁴⁹ Vision 2011, 52.

slightly. If rigorous attempts to reduce emergent phenomena fall short, then this suggests the necessity of revising our approach. In my analysis of the issue, I find that we should either (i) doubt reductionism, (ii) doubt our ability to thoroughly investigate nature, or (iii) find emergence to be brute. We have no reason to doubt the ability of reductions to elucidate many truths about reality, but the failure to reduce emergent phenomena might mean reductionism is incapable of revealing all truths.

Though there are philosophers who support option (ii) and make claims that we are cognitively closed to certain truths, the approach of simply accepting option (ii) is a dangerous path to tread.¹⁵⁰ While it is reasonable to doubt our infallibility and the completeness of our knowledge, we need to approach our investigations with the right amount of epistemological confidence. To be overly confident in our abilities will assuredly lead us down false paths, but to have too little epistemological confidence can undermine too many of our beliefs. In other words, some methodological skepticism is a healthy approach to tempering our claims and pushes us toward discovering multiple proofs for our beliefs. However, adopting a mindset of cognitive closure tells us that we are inherently limited, and our ability to understand reality is epistemologically fraught. I worry this view would have a detrimental undermining effect on our explorations of our experiences and understanding of reality.

Therefore, options (i) and (ii) are too extreme, so option (iii) bruteness is the next logical step.

I think it important to state the point I made above more explicitly. The failure of reductions to elucidate emergence is not a call to abandon reductive efforts. Instead, it is

¹⁵⁰ The idea of cognitive closure is most famously tied to McGinn. See McGinn 1989.

a recognition of the complexity of reality. A testament to the layered conception, which reveals to us that a neat and tidy reduction might not always succeed because of nature's built-in complexity and indeterminacy.

For context, let us consider consciousness. Despite the success and strides of neuroscience, consciousness still eludes reductive efforts and, therefore, it continues to be best understood as an emergent phenomenon. Vision writes:

[t]he leading point is that for conscious mental states the possible moves have been fairly well plotted and the structure of the problem facing investigators seems clear. In addition, the methods are largely in place, though we can expect them to be enhanced by more sensitive instruments. In sum, the direction in which future inquiry will move is as thoroughly understood as we have any reason to expect in a mature science. [...]

If we cannot expect to find anything more than this with controlled animal experiments and the diagnoses of mental deficits due to brain trauma or genetic mishaps in humans, then [...] we have good reason to claim to know that there is nothing more to be explained about the supervenience of the conscious on the neural.¹⁵¹

Vision's quote stands on its own, but I will continue with a slightly related point.

Regardless of the advances we may make in neuroscience (or science in general), we ought to hold that consciousness (and various mental states) is a metaphysically brute emergent phenomenon until said advances can reduce it. However, given that our empirical methods are firmly set, future discoveries seem limited to more precise measurements and readings. There are no clear avenues toward breakthroughs that will result in the reduction of consciousness. Thus, I agree with Vision when he writes:

[...] there is little more to be said than that material configurations can give rise to uncharted consequences. The result I arrive at is that conscious properties and states are *emergent*, and that, although they depend for their existence on their material bases (a dependence summed up by

¹⁵¹ Vision 2011, 54-5.

supervenience or realization relations to the material), there are no further details to explain that dependence.¹⁵²

Vision also notes that “many will be unsettled by the fact that this leaves us with a primitive, brute relation between the conscious and the physical.”¹⁵³

Based on the above discussion and given the ontological conception of explanation, I contend that emergentists should opt for (b) bruteness over (a) opaqueness. If the mysteriousness of emergence were tied to opacity, then that mystery would be directly tied to an epistemological shortcoming. Emergence is not about the limits of our knowledge; it is about the nature of the phenomena itself. Ultimately, emergence challenges absolute reductionism as a paradigm methodology while reminding us that reality is complex. That said, emergence might seem contradictory to reductive practices, but they can coexist.

Strong emergence emphasizes the existence of irreducible novel properties that are fundamentally rooted in our naturalistic understanding of reality. While critics may level a claim of obfuscation, on the contrary, emergence provides a reason for the appearance of an explanatory gap problem. The ontological conception of explanation, as explained by Vision and expanded upon in this chapter, posits that strong emergence is metaphysically brute. It follows that strongly emergent phenomena are metaphysically inexplicable and not representative of an epistemological limitation. Thus, the appearance of an explanatory gap is a testament to the inherent complexity of reality and the bruteness of emergence. Under this light, both reductionism and emergentism, though initially appearing contradictory, both contribute to our understanding of reality.

¹⁵² Vision 2011, ix.

¹⁵³ Vision 2011, xi.

Summary

In the preceding subsections, I explored the nature of emergent phenomena, focusing on their intrinsic mysteriousness, resistance to reduction, and alignment with *ceteris paribus* laws. These aspects illuminate the metaphysical bruteness of emergent phenomena and challenge the scientific idealism underpinning absolute reductionism.

I rely heavily on Vision's work on emergence and his conclusion that the lack of a detailed explanation is due to the inherent nature of emergence. I agree with Vision's assessment but add that we should expand our investigative methods when studying emergent phenomena, especially regarding the mind and mental properties. Despite the ontological brute nature of emergence, a multi-disciplinary and multimodal approach could enrich our understanding of emergent phenomena.

The inherent complexity and indeterminacy of reality alongside the metaphysical bruteness of emergent phenomena defy the possibility of a singular methodological approach. Instead, a multimodal approach that integrates various scientific and philosophical methods offers a more holistic perspective. For instance, while neuroscience has made significant strides it cannot bridge the gap between mental facts and physical facts. However, neuroscience supported by philosophy and psychology can enrich our overall understanding of the mind even if a knowledge gap will remain due to the metaphysical bruteness of emergence.

The current scientific paradigm, heavily influenced by reductionism, is challenged by anomalies like emergent phenomena. Proponents of emergent phenomena, such as myself, find that those particular anomalies should not be dismissed. Instead, they should be integrated into our explorations and understanding of reality. Through a multimodal

approach, emergence can exist in supplement to reductionism. Put simply, embracing the metaphysical bruteness of emergent phenomena does not indicate a dead end in our investigation of reality as we experience it; rather, it calls for a holistic and multimodal approach to inquiry. Through various disciplines, we can pave the way for a richer and more nuanced understanding of emergent phenomena.

Physicalism's epistemological thesis states that physical facts determine all mental facts. This thesis can still stand so long as it is recontextualized. First, it must be understood that mental facts are themselves physical facts. Second, the word "determine" cannot be understood as reductive determinism. Instead, a holistic conception that accepts indeterminacy and the considerations of *ceteris paribus* laws must be adopted. For instance, a better word would be "informed" so that physical facts inform all mental facts.

Thus, so long as one accepts that emergent mental facts are physical facts and that physical facts are not just reductive facts, physicalism's epistemological thesis is not threatened, and therefore, physicalism is not threatened either. The following sections further explain how physicalism's theses remain intact under my theory of emergent mental properties.

On the Origin of Mental Properties

Now that emergence is presented as a viable theory capable of picking up where reductionism falters, let us focus again on mental properties. Once mental realism is accepted, the alternatives to emergence appear flawed. For one, recall Churchland's challenge that physicalist property dualists must either find that mental properties result

from evolutionary emergence (i.e., be generated) or be fundamental properties of reality.¹⁵⁴

Physicalists' leading mental-physical relationships each fail to provide explanations for the origin of mental properties. For example, physicalists who subscribe to token identity theory or functional reductions do not consider whether mental properties are fundamental or emergent. These physicalists merely acknowledge mental properties and attempt to reckon with them without trying to understand them. As noted, I find this stance problematic and untenable.

By avoiding addressing the nature and genesis of mental properties, the issue is merely pushed aside. However, the issue is crucial to our understanding of the nature of reality. I contend that physicalist mental realists adopting a stance that does not address the origin of mental properties is metaphysically and epistemologically deficient. The stance is metaphysically deficient because it fails to take a position on the status of mental properties while ostensibly accepting their existence. To acknowledge their existence and take no stand on their metaphysical status is to treat them as metaphysical outliers. This approach does not inspire confidence that the physicalist is embarking on a good-faith exploration of mental properties and mental realism.

Furthermore, the position is epistemologically deficient because, without a stance on the metaphysical status of mental properties, all attempts to regulate mental properties must be based on either assumptions or ignorance. To be ignorant about the status of mental properties or to simply hold assumptions about them both leads to the creation of theories that prop up existing knowledge structures at the expense of a genuine

¹⁵⁴ Churchland [1984] 2013, 20.

exploration of apparent phenomena and facts. Once again, such an approach cannot be said to be a genuine attempt at enhancing our understanding of reality and a good-faith exploration of mental properties. These constructs result from dogmatic thinking that hampers exploration and understanding, ultimately perpetuating the notion that the emergence of mental properties is mysterious.

Criticizing emergence as mysterious does not provide solid grounds for dismissing the explanation. For one, the bruteness of emergence is not a true mystery. Bruteness is, in fact, a metaphysically sound explanation for the relationship between some phenomena. Second, even if emergence feels vague or obscure, it does amount to our current understanding of certain phenomena (e.g., mental phenomena). Given that it is our best explanation for certain phenomena, we ought to embrace the so-called mystery in favor of progress and move to investigate the nature of reality and mental properties through it as our most promising framework.

Ultimately, in the philosophy of mind, the issue becomes less about avoiding mystery and more about choosing the most promising framework for capturing reality as we understand it. Moreover, it underscores the need for a more flexible and inclusive scientific methodology capable of accounting for phenomena that do not neatly fit into a reductionist framework. It is within this framework of epistemological humility and openness that we can fully explore the richness and depth of mental properties.

We must acknowledge that the current methodological approach of reductionism, while largely effective, may not be suitable for understanding mental phenomena. Mental realism demands that physicalists recognize the need to expand their methodological toolkits. As it stands, emergence is the most promising alternative framework available.

Importantly, incorporating emergence does not mean abandoning the reductive methodology. A broader approach is needed to address the complexity and uniqueness of mental phenomena. We should embrace a pluralistic approach to the philosophy of mind and mental realism, where reductionism and emergence are utilized to understand mental phenomena and their various cognate considerations. Moreover, given its widespread application in various scientific disciplines, physicalists should be open to evaluating mental properties through the lens of emergence.

As we have seen in earlier chapters, the common physicalist attempts to explain mental properties yield unsatisfying results. The mental-physical relationships offered treat mental properties like metaphysical danglers that cannot be ignored but must be closely managed. The physicalist management of these phenomena leaves much to be desired largely because they do not attempt to substantiate them metaphysically. Therein lies the fundamental dilemma of the mental realist physicalist—once they grant the existence of mental properties, they must either ignore their metaphysical status and origin, or they can accept that emergence is what metaphysically substantiates mental properties.¹⁵⁵

Conclusion

As discussed in earlier chapters, the mind and mental properties have resisted reduction, and reductive avenues do not look promising. At the heart of reductive physicalism is the belief that mental states are nothing more than physical states. Unfortunately, there are several shortcomings to reductive physicalism that were discussed in Chapter 2.

¹⁵⁵ See Aulisio 2022 for a longer discussion on the origin of mental properties.

Reductive physicalism asks us to adhere to identity based on insufficient evidence. For instance, if pain is identical to a certain neural firing pattern, then every instance of that neural firing pattern should cause pain. However, there is no definitive evidence favoring this kind of one-to-one correspondence between mental and physical states, and there is reason to reject this universal pairing. The primary evidence for reductive physicalism is correlations between reported mental experiences and physical facts, but this evidence does not rise to proof of identity. The secondary evidence for reduction and type identity is its alignment with dogmatic adherence to generally held physicalist principles.

As an alternative to reduction, several non-reductive mental-physical relationships have taken up the physicalist cause of explaining and limiting mental properties. These mental-physical relationships also leave a lot to be desired. For one, these non-reductive relationships tend to be supported by their generality. While mental states are realized by physical states, none of the non-reductive positions offered a detailed account of how this realization occurs. The most detailed of the non-reductive physicalist relationships is supervenience, which, at its basic level, maintains that any changes in mental properties require a change in the underlying physical states. Yet, supervenience is also compatible with forms of emergentism, and I have argued that it is not robust enough to stand on its own.

The generality of non-reductive physicalism can be considered a strength as it avoids contradictions and allows for flexibility should future scientific discoveries reveal new understanding about the relationship between mind and brain. However, this flexibility requires vagueness, making non-reductive physicalism minimally explanatory.

I find the most troubling explanatory shortcoming to be the lack of commitment to the origin of mental properties and other emergent phenomena. Without an origin, mental properties become metaphysical dangles that “float” above physical states. The problem of floating mental properties leads to questionable entailments, such as epiphenomenalism. Put simply, there is an important gap in the account of how physical activity gives rise to different mental states.

The idea of emergent phenomena being realized by their complex bases allows us to simplify the metaphysical model without dismissing the origin of higher-order properties. In other words, emergence accepts the reality of higher-order properties, such as mental properties, and proposes an origin for those mental properties. Emergent phenomena are generated by complex physical activity and may stem from multiple sources that work in combination. The emergent phenomena will correlate with its contributing factors, but identifying a reductive relationship is impossible because emergence is metaphysically brute.

Emergence is an attractive theory of mind because it offers an explanation for mental properties. Emergence addresses *how* mental properties come to be through a physically anchored, metaphysically brute feature of reality and explains *why* they are distinct from physical properties through the ontological conception of explanation.¹⁵⁶ Using the mind as an example, emergence claims that out of the complexity of the brain and its numerous interconnected processes, the mind and its mental properties emerge.

¹⁵⁶ A future avenue of research would be an exploration of emergent mental properties as I have presented them considering substratum and bundle theories. This would further clarify the substantive distinction of the mind and its relationship with physicalism.

Furthermore, emergence lays the groundwork for understanding mental properties, their purpose, and their causal role. For instance, if mental properties can be shown to serve a necessary role in more significant structural changes, then it is conceivable to see them serving a causal role that could otherwise be denied to mental properties. In other words, if it is plausible that emergent mental properties play a necessary role in changing larger structures or systems, such as human behavior or society at large, then mental properties have real causal power or influence. Mental properties would not be byproducts or illusions but would genuinely shape physical events.

Notably, emergence builds in an explanation for mental-physical correlations because emergent phenomena necessarily depend on their bases. By being dependent, emergentism commits to fitting into our broader understanding of reality. Emergent mental phenomena will, at minimum, have their metaphysical roots in the physical world. It follows that mental properties are not rogue existents and must play by the rules of reality. As such, emergent properties satisfy physicalism's metaphysical thesis. This work primarily focuses on causally efficacious mental properties under physicalism. Given that focus, I will explore how causally efficacious emergent phenomena could be classified as a form of physicalism primarily by addressing the Causal Exclusion Argument in Chapter 4.

CHAPTER 4

CAUSALLY EFFICACIOUS MENTAL PROPERTIES

Chapter 4 demonstrates the possibility of downward mental causation. To accomplish this objective, I focus mainly on Kim's Causal Exclusion Argument (henceforth CEA). My refutation of the CEA hinges on addressing the causal closure, supervenience, and overdetermination theses. Ultimately, I show that downward mental causation is necessary under emergentism. I accomplish this by introducing a framework that I call Dynamically Stable Causal Holism.

In my theory, I emphasize the nonlinear aspects of reality to explain emergent phenomena, arguing that they do not align with supervenience, which assumes a linear and predictable relationship between different levels of reality. This perspective empowers the exploration of downward mental causation within emergent systems. The essence of emergence lies in the idea that higher-level properties and phenomena arise from lower-level parts, properties, and processes while remaining irreducible to the underlying structure. Downward causation, conversely, requires higher-level emergent phenomena to causally influence the lower-level substrates from which they emerged.

Vision describes three theses of emergent properties, they are:

- (1) E [i.e., an emergent phenomenon] is *dependent* on different sorts of a non-emergent base in a way made manifest by E's *supervenience* (or *realization*) on those same properties.
- (2) There is *no* further (minimal) *explanation* of why E is supervenient on (or dependent on, or realized by) that non-emergent base, viz., the relationship is brute.

(3) E is a cause (of both mental and physical aspects) in ways in which there is no sufficient cause in context at the levels of E's non-emergent base(s).¹⁵⁷

I refer to Vision's three theses as (1) realization, (2) bruteness, and (3) causal efficacy. Of the three, causal efficacy encounters the greatest resistance, as it challenges the physicalist position that all causes are physical causes and highlights the novelty of emergent phenomena.

Physicalism motivated by strong supervenience or absolute reductionism ostensibly allows for two causal possibilities: vertical determination (i.e., bottom-up causation) and horizontal causation. Under horizontal causation, one physical thing causally affects another physical thing on the same level. For example, a baseball breaking a glass window is an everyday example of horizontal causation. Though horizontal causation is an acceptable way of discussing causation among clearly physical properties, horizontal causation among mental properties is problematic for physicalists.

Kim writes:

[...] if *Supervenience* is assumed, mental-to-mental causation entails mental-to-physical causation [...]. Given *Supervenience*, it is not possible to have causation in the mental realm without causation that crosses into the physical realm. [...] If we were to choose M over P as P*'s cause, *Closure* would kick in again, leading us to posit a physical cause of P*, call it P₁ [...], and this would again call for the application of *Exclusion*, forcing us to choose between M and P₁ [...]. Unless P is chosen and M excluded, we would be off to an unending repetition of the same choice situation; M must be excluded and P retained.¹⁵⁸

¹⁵⁷ Vision 2011, 14. The bracketed text has been added for clarity.

¹⁵⁸ Kim 2005, 40 & 43.

In brief, the Supervenience Argument endorses a view of causation that traces all the mind's causal power back to underlying physical interactions. Under this framework, all instances of mental events are the result of vertical determination.

In Kim's view, the appearance of cause and effect amongst non-reductive mental properties must be attributed to supervenience. The mental causal event supervenes underlying physical components, and it is their interactions that are causally efficacious, allowing macro properties to have the appearance of causal efficacy. Though Kim's supervenience argument undermines the efficacy of non-reductive physicalism, Kim ultimately takes a hardline reductionist approach to causation. He writes:

[...] the principal reason why there are no strict laws at higher levels is because of causal interference from lower levels, ultimately events at the basic physical level. [...] The interfering physical event interacts with the physical system/structure that represents (or realizes) the higher-level system. Interference is a causal process that takes place at the basic physical level. [...] This suggests the following picture: all causal processes, and explanations that track them, are fundamentally physical, in the sense that they are governed and explained by strict physical laws. In this sense, all causation and explanation are at bottom physical causation and explanation. Causation and explanation via nonstrict special-science laws can be *heuristically informative* and *practically useful*, but we must recognize them as derivative from, and *dependent* on, fundamental physical causation and explanation. There always is a causal-explanatory story at the basic physical level, and any purported special-science story must be appropriately related to the physical story if we are to achieve a *single, coherent, and nonredundant* overall picture of the world.¹⁵⁹

We can generally categorize Kim's hardline position on causation as a reductive approach. In the quoted text, Kim uses "physical" to refer to the fundamental physical level, while all other levels of reality are related to that fundamental physical level through supervenience. While heuristically informative and practically useful, the

¹⁵⁹ Kim 2005a, 337-8.

appearance of causation at macro levels has this appearance because everything macro supervenes on the fundamental physical level.

Standing in opposition to Kim's reductionism and supervenience argument is emergentism. As previously discussed, emergentism posits that natural complexity engenders new properties to arise at higher levels of reality. These novel emergent properties are neither predictable nor reducible to the parts from which they emerged. The strong emergentist argues that emergent phenomena are metaphysically substantive and could play a causal role beyond the mere supervenience role that Kim carves out for higher-level phenomena. If Chapter 3 allows the reader to accept the possibility of emergent properties and phenomena with their own real, distinctive, and novel characteristics, then the next step is to dive into their causal role. I propose that the novelty and distinctiveness of emergent properties reveal that they play a necessary causal role in complex systems. Ultimately, I argue that emergent phenomena can affect the systems from which they emerged, which I believe is the most informative view of how a complex system can maintain stability.

The emergentist will maintain that understanding the mind cannot be accomplished solely through identifying neurons firing and cataloging neural pathways; it is also necessary to recognize higher-level abilities, such as beliefs, desires, and feelings. Arguably, these higher-level mental properties could shape the neuronal activity from which they emerged. This is a clear departure from absolute reductionism, but it is a perspective that takes a broader view of the layered conception of our complex reality. In this chapter, I introduce my theory of Dynamically Stable Causal Holism, a view that takes a broader view of complex systems and embraces the idea of feedback loops. As a

general overview, Dynamically Stable Causal Holism (henceforth DSC Holism) endorses the mutual influence between levels, emphasizing that while lower-level components facilitate the emergence of higher-level properties, the latter, once formed, can influence its substrate. This downward causal influence is necessary to explain how a complex, multilayered system can achieve higher-order ends and maintain system stability.

Embracing downward causation and emergentism brings a fresh and dynamic perspective to the philosophy of mind. Recognizing the two-way interaction between levels enriches our understanding, moving us beyond the constraints of supervenience or reductive physicalism. The subsequent sections will delve into this perspective's challenges and critiques, primarily from the vantage of the CEA.

Causal Exclusion Argument

Jaegwon Kim's Causal Exclusion Argument (henceforth CEA) has been and continues to be the principal objection to downward causation. To address this argument, we must first understand its structure and underlying assumptions.

Kim sets the stage for the CEA by laying out four positions that physicalist property dualists maintain. The four positions:

- I. Mental/physical property dualism: "the view that mental properties are irreducible to physical properties."¹⁶⁰
- II. Mind-body supervenience: "Mental properties strongly supervene on physical/biological properties. That is, if any system *s* instantiates a mental property *M* at *t*, there necessarily exists a physical property *P* such that *s* instantiates *P* at *t*, and necessarily anything instantiating *P* at any time instantiates *M* at that time."¹⁶¹

¹⁶⁰ Kim 2005b, 22.

¹⁶¹ Ibid, 33.

III. Causal closure of the physical: “If a physical event has a cause at t , then it has a physical cause at t .”¹⁶²

IV. Principle of Causal Exclusion: “If an event e has a sufficient cause c at t , no event at t distinct from c can be a cause of e (unless this is a genuine case of causal overdetermination).”¹⁶³

From these four positions, the CEA is formed. The crux of the CEA is that if every physical event has a physical cause, and if irreducible mental properties supervene physical properties, how can mental properties have causal force without it being a case of causal overdetermination? Kim’s take is that the combination of mind-body supervenience and the causal closure of the physical relegates mental properties to a lesser causal role, such as causal relevance. Moreover, the Principle of Causal Exclusion advances the idea that if the physical properties on which mental properties supervene represent a sufficient cause for physical events, then supervening mental properties appear causally extraneous. Thus, epiphenomenalism looms large on mental properties unless it can be demonstrated that they are causally relevant through reduction or supervenience.¹⁶⁴

To recreate the above narrative, the following four-premise argument represents the CEA:

1. the mental is not identical to the physical,
2. mental properties cause both physical and mental effects,
3. the causal closure of the physical—the physicalist tenet that every physical effect has a sufficient physical cause, and

¹⁶² Ibid, 15.

¹⁶³ Ibid, 17.

¹⁶⁴ See Kim 1998. Also, for a collected resource of Kim’s work on the mind’s causal problems, see Kim 2002.

4. causal overdetermination is rare and cannot justify large-scale systematic mental causation.¹⁶⁵

The CEA centers on how mental properties can have causal powers when physical properties are causally sufficient (c.f., premise 3). In essence, the argument is that physical causes can fully account for physical effects, leaving no causally efficacious role for mental properties. Given causal closure and the supervenience link between mental and physical properties, mental properties do not bring any causal force to the equation and, therefore, can be *excluded* from the causal explanation. Ultimately, the CEA is meant to refute the practicality and necessity of downward causation. However, the CEA does not eliminate the possibility of downward causation.

Given the above explanation, it seems that the emergentist looking to vindicate downward causation has an uphill climb. The emergentist cannot reject premise 1 or 2. To deny premise 1 is to reject emergentism and embrace absolute reductionism. Also, denying premise 2 summarily rejects mental causal efficacy. The CEA is powerful because each premise is reasonable, and they combine to detract from the possibility of mental causal efficacy.

Kim's CEA challenges the plausibility of downward causation, raising pivotal concerns based on supervenience, causal closure, and causal overdetermination. However, the CEA leaves an opening for downward mental causal efficacy if any of the following can be shown. Either (i) emergent properties and phenomena are realized by but are not in a supervenience relationship with their underlying structure, (ii) emergent

¹⁶⁵ See Vision 2011, 95-7 for a clear overview of the causal exclusion argument and the various options one might select in response. Also, in Chapter 1, I review the consequences of rejecting each premise.

downward mental causation does not violate the causal closure of the physical, or (iii) wide-scale causal overdetermination is not as problematic as the CEA suggests.

Supervenience

Supervenience, a central thesis in the philosophy of mind, is especially crucial to non-reductive theories of mind. In its minimal form, the supervenience thesis maintains that there cannot be a change in mental properties without an accompanying change in physical properties (i.e., there cannot be A-differences without B-differences).¹⁶⁶ It is crucial to revisit supervenience for two reasons. First, strong emergence, the view I have endorsed in this work, is compatible with supervenience in its minimal form. Second, it is worth exploring supervenience alongside the idea of nonlinearity, a relational theory often utilized to interpret emergence in disciplines outside philosophy. Understanding if there is a relationship between nonlinearity and supervenience will help us better understand the entailments of emergence.

Strong Emergence

The common physicalist position that every mental cause can be substituted for a supervening physical cause is not necessarily true. While supervenience describes a metaphysical dependence relationship between the mental and physical, that dependence does not also entail a causal mechanism. In other words, even if we maintain supervenience, we cannot presume that every mental cause is simply an underlying physical cause. For instance, it is conceivable that some mental causes may operate in concert with underlying physical causes but not in a manner that replaces them with the

¹⁶⁶ McLaughlin and Bennet 2018, 3.

causal work of the physical states they supervene. For instance, if emergent mental properties play a contributory causal role that the supervening physical cause alone could not generate, then the reality of mental causal efficacy would be undisputed.

One of the advantages of supervenience is that it supports the interpretation that mental properties are part of the physical world. As such, if mental properties supervene physical states, then the metaphysical thesis and physical completeness can be upheld. It follows that there is a basis for considering emergent phenomena and mental properties to be physically compatible. In contrast, when Kim refers to the “physical,” he seems to refer to the micro-physical phenomena studied by physics and reducible macro-physical phenomena.

In Chapters 2 and 3, I addressed the shortcomings of reductive mental-physical relationships. Despite Kim’s position, there is insufficient reason to find all causation occurring at the fundamental level. At best, the supervenience physicalist can opt for minimal supervenience between levels, including the mental and physical, to ensure physical completeness and determination. In contrast, my position is that there is a distinction between mental and physical properties, but the distinction does not exclude mental phenomena from the physical realm. The primary distinction is that mental phenomena cannot be reduced to other physical phenomena due to nonlinearity and indeterminacy. Regardless of their irreducibility, mental phenomena are intertwined with the physical world. The demonstrable and consistent connection between mental experiences and the physical world makes disentangling the two impossible.

Consider, for instance, a description of an experience that Aldous Huxley chronicles in *The Doors of Perception*. He writes:

[w]hat happens to the majority of the few who have taken mescaline under supervision can be summarized as follows. [...] (3) Though the intellect remains unimpaired and though perception is enormously improved, the will suffers a profound change for the worse. [...] (4) These better things may be experienced (as I experienced them) “out there,” or “in here,” or in both worlds, the inner and the outer, simultaneously or successively. That they *are* better seems to be self-evident to all mescaline takers who come to the drug with a sound liver and an untroubled mind. [...]. In the final stage of egolessness there is an “obscure knowledge” that All is in all—that All is actually each.¹⁶⁷

Through his repeated use of mescaline, Huxley claims to have had profound experiences leading to a better understanding of the relationship between his mind and the wider physical world. Huxley distinguishes his mental experiences from what we might describe as the hallmark examples of physical reality (e.g., tables, chairs, neurons, atoms, etc.), but he never presents them as entirely separate or even disjointed.¹⁶⁸ Interestingly, Huxley is adamant that his inner self (i.e., his mind) and external reality are intertwined in a metaphysically significant manner.

Regardless of how far one wants to take Huxley’s psychedelic journaling, there is value in the evidence of physical substances (e.g., mescaline) having profound effects on mental phenomena and experiences. Whether the causal power is one-way (i.e., bottom-up) or multi-directional and allows for downward causation, there is a causal connection between the physical and mental. So long as one is a mental realist, the evidence suggests that mental and physical phenomena are causally interconnected.

¹⁶⁷ Huxley (1954) 2009, 25-6. I address the concept of the will later in this chapter.

¹⁶⁸ *Ibid*, *passim*.

Even the epiphenomenalists' claim that mental properties may be an after-effect of physical processes cannot undermine this causal linkage. Even if mental properties are not causally efficacious, their emergence from physical states indicates a metaphysical dependency that binds them to the physical domain. This dependency underscores the mental and physical interconnectedness regardless of the causal landscape.

I will address this in more detail in the following pages, but first, I will briefly state my position. Emergent phenomena, which include those things we classify as mental, metaphysically depend on the underlying physical structure from which they emerge. Emergent mental phenomena are products of physical reality, but they are distinguishable from hallmark physical things because they are irreducible, unpredictable, and have novel properties. There is nothing about those distinctions that require emergent phenomena to be extra-physical. Emergence is a metaphysically brute relationship occurring between layers of reality. Moreover, nonlinearity and indeterminism support the idea that emergence is metaphysically brute while not undermining the claim that emergent phenomena are physically compatible.

Supervenience physicalists suggest that mental states exist in a metaphysically dependent relationship with physical states. When mental properties or states supervene physical states, mental properties are contingent upon physical states (i.e., mental properties metaphysically depend on physical states). The supervenience thesis is appealing because it empowers at least a minimal or basic form of physicalism that acknowledges the ontological priority of the undergirding physical layers. However, the supervenience thesis also empowers the CEA to posit that all mental causation is just

instances of supervening (i.e., underlying) physical causation. Ultimately, the supervenience connection nullifies the causal efficacy of mental phenomena.

Depending on its formulation, emergence can coexist with or challenge the supervenience thesis. Weak emergentism agrees that higher-level properties supervene on lower-level ones. The distinction is that while these emergent properties might be unpredictable or unexpected based on our current understanding, in principle, they are derivable from the properties and relations of the base level. So, while emergent properties are unpredictable and novel relative to the base level, they are still dependent upon that underlying base level.

Given the complexity of emergence, it may be epistemologically presumptuous to bind emergent phenomena in a supervenience relationship without compelling reasons. In other words, emergentism suggests that reality is complex, and supervenience may only be a provisional descriptor. A descriptor that cannot capture all aspects of the mental-physical relationship.

If we can suspend the idea that all mental causes necessarily supervene physical causes, we pave the way for a richer causal landscape. Moreover, it becomes conceivable that emergent mental states possess causal powers. In this scenario, emergent mental states may have causal powers that work alongside or in tandem with physical causes while not being reducible to them. This viewpoint does not diminish physical causation; rather, it expands our view of the causal landscape, upholding traditional microphysical causes while acknowledging the efficacy of mental causes.

Nonlinearity

Supervenience has occupied a central position in physicalist debates, emphasizing the relationship between higher-level states (e.g., mental properties) and their supervening lower-level states (e.g., brain states). In this subsection, I compare supervenience's rigid dependency with nonlinearity's more dynamic and unpredictable nature.

To understand nonlinearity, consider the sprawling effects caused by something as minor as a traffic jam. For example, an immobilized car on a major highway can cause an impassable jam that affects thousands of people. The repercussions of the traffic jam might span from individuals missing important meetings to others rerouting and experiencing unforeseen events. This example underscores the principle of nonlinearity, where small changes at one level can lead to significant and unpredictable outcomes at another. The unpredictability and possible outcomes align with emergence while challenging supervenience's straightforward dependency assertion.

In discussions about emergent phenomena, the principle of nonlinearity becomes crucial.¹⁶⁹ Strongly emergent properties are not merely unexpected but irreducible to their underlying structure and conditions. These emergent traits could have causal powers beyond supervening underlying components. Stephen Wolfram's concept of computational irreducibility provides insight here. He writes:

[it is] a crucial—and utterly unexpected—feature of the computational universe: that even among the very simplest programs, it's easy to get immensely complex behavior. [...] The first step is to think of every process—whether it's happening with black and white squares, or in physics, or inside our brains—as a computation that somehow transforms

¹⁶⁹ See Wilson 2013.

input to output. [...]. [A]ll these things are computationally equivalent. [...] For one thing, it implies what I call computational irreducibility. If something like rule 30 is doing a computation just as sophisticated as our brains or our mathematics, then there's no way we can "outrun" it: to figure out what it will do, we have to do an irreducible amount of computation, effectively tracing each of its steps. [...] But what computational irreducibility implies is that out in the computational universe that often won't work, and instead the only way forward is just to explicitly run a computation to simulate the behavior of the system.¹⁷⁰

Wolfram suggests that the complexity of some systems prevents us from predicting outcomes by understanding initial conditions. While Wolfram primarily applies this idea to computational systems, there is no apparent reason to exclude biological and cognitive systems from these same considerations. The complexity of biological and cognitive systems highlights a potential gap in supervenience's explanatory power.

Furthermore, I interpret Wolfram's last sentence on computational irreducibility to mean that the complexity of reality suggests that we need, at times, to put aside reductive methods and analyze complex systems as wholes in and of themselves. Simply put, the approach of classical mathematics and reductionism will not work when dealing with some complex systems, such as those that are emergent.

In addition to Wolfram, others have noted the nuanced relationship between the microscopic and macroscopic. For instance, Sollberger writes:

[m]icroscopical exactness does not equal macroscopical exactness: we apply double standards. Even if the causal potential of an individual aggregate supervenes on the compositional details of its structure, it remains quite possible for mereologically different aggregates to exhibit the same prominent macroscopic behaviour. Exactness for special-science kinds does not require the macroscopic pattern of characteristic causes and effects of its members to be exactly identical - all that is needed is similarity in terms of causal homogeneity. That is the reason why many microscopically distinct individuals display equal macroscopic

¹⁷⁰ Wolfram 2017, 199-200. See also Wolfram 2022, *passim*.

behavioural features that are subsumed under a well-defined special-science kind. In slogan form: structurally dissimilar systems can behave in exactly the same way.¹⁷¹

Sollberger explains that microscopic components can differ considerably among systems yet still lead to indistinguishably similar macroscopic behaviors. Sollberger's point feeds into what I have suggested are limitations to a strict supervenience-based understanding of our complex reality where there are straightforward one-to-one relationships between levels.

At its core, supervenience posits a tight dependency relationship between two or more sets of things (e.g., properties, states, levels of reality, etc.). Supervenience tells us that a higher-level state change requires an underlying state change to engender the higher-level change. However, if there is indeterminacy baked into reality, this would offer a metaphysically sound explanation for unpredictability in nonlinear systems. In complex systems, small foundational changes could lead to disproportionate outcomes at higher levels. Conversely, it may also be the case that lower-level changes (or differences) might not always result in significant higher-level shifts (or differences). Put simply, the complexity of nature suggests that the relationship between one level and another may not be as straightforwardly bound together as supervenience requires.

At this point, delineating two more robust versions of supervenience will further clarify the relationship between nonlinearity and supervenience. Kim defines strong supervenience as:

A is said to strongly supervene on B just in case: [...] Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and *necessarily* if any y has G, it has F.¹⁷²

¹⁷¹ Sollberger 2005, 341.

Kim's definition clarifies that strong supervenience requires a direct and necessary relationship between properties or states across different levels. An alternate but similarly strong formulation of supervenience is global supervenience. Global supervenience maintains that two microscopically identical worlds will also be macroscopically identical.¹⁷³ Though these two strong formulations of supervenience do not entail one another, both set up a determinative relationship, where the higher level not only necessarily depends on the lower level but is also determined by the lower level.

Given the inherent unpredictability in nonlinear systems, proponents of strong formulations of supervenience are presented with a significant challenge. Since there is demonstrable evidence of nonlinearity in complex systems and our best science endorses at least some indeterminacy in nature, how can proponents of strong formulations of supervenience explain that specific foundational changes correspond to any given outcome in a complex system with precision (e.g., the human nervous system)? In other words, these strong formulations of supervenience make major claims about the relationship between an outcome and its supervenience base. Yet complex systems, including nature in general, are not universally determinative.

The mere affirmation that mental states are metaphysically anchored in physical ones is insufficient for capturing the extent of the claims made by strong and global supervenience. The rigidity of strong and global supervenience should also require some evidence that microscopic conditions entail specific macroscopic conditions (i.e., extraordinary claims require extraordinary evidence).

¹⁷² Kim 1987, 316.

¹⁷³ Paull and Sider write, "[a] *globally supervenes* on B iff any two worlds with the same distribution of B-properties have the same distribution of A-properties as well." Paull and Sider 1992, 834.

In other words, the apparent nonlinearity of complex systems and indeterminacy in nature shifts the burden of proof to the proponents of strong supervenience. They must elucidate how specific foundational changes give rise to emergent outcomes within the framework of supervenience physicalism. This demands more than a cursory affirmation that mental states are determined by physical ones. The challenge is in explaining the precise dynamics of this relationship, considering evident nonlinearity and indeterminacy.

Supervenience is a central concept of physicalism, but nonlinearity in our complex reality demands scrutinizing its assertions. Supervenience must account for nonlinear systems' dynamic and unpredictable nature, but the rigidity of supervenience limits how far it can stretch.

Stronger formulations of supervenience are undermined by nonlinearity; however, let us briefly look at the weakest version. The weakest version of supervenience is the slogan form, namely "there cannot be A-differences without B-differences"). However, even this version is challenged by nonlinear emergence. For example, imagine two parallel realities with identical B-differences, yet there are distinctions in the A-differences in these two realities. So long as indeterminacy is a feature of both realities, there is the possibility that B-differences are identical between realities while A-differences are not. Notably, A-differences are still realized by B-differences in both realities, but those A-differences need not be identical to one another. An emergent explanation for this is that the emergent A-state metaphysically depends on the B-state. However, the process of emergence is metaphysically brute. Ultimately, indeterminacy and nonlinearity allow for identical B-states and differing A-states.

Stepping away from our modal example. So long as indeterminacy is a feature of reality, then supervenience need not be the case for emergent phenomena. Put simply, indeterminacy in nature entails the possibility of nonlinearity, which means that identical B-differences may not result in identical A-differences. Notably, this nonlinear emergence does not undermine the metaphysical priority of the physical state (i.e., the B-state). Ultimately, this should be the more important thing for physicalists to cling onto since it allows nonlinear emergent to be physical.

Causal Closure

Causal closure, a foundational concept of physicalism, posits in its most basic form that physical effects must have physical causes. Closure has profound implications, especially in metaphysics and the philosophy of the mind. Chalmers notes both the importance and one of the implications of closure when he writes:

[t]he best evidence of contemporary science tells us that the physical world is more or less causally closed: for every physical event, there is a physical sufficient cause. If so, there is no room for a mental ‘ghost in the machine’ to do any extra causal work.¹⁷⁴

Chalmers’s quote affirms that the leading scientific evidence suggests the causal closure of the physical domain. Chalmers makes it clear that one of the consequences of closure is that it impacts mental causal efficacy, at least in so far as it is described as a “ghost in the machine.” Emergent mental properties and their causal powers are not akin to a “ghost in the machine,” however, it is important to entertain the sentiment of Chalmers’s quote and press against my position that mental properties are part of the physical. Thus,

¹⁷⁴ Chalmers 1996, 125.

I grant that, on the face of it, causal closure calls into doubt the causal role of mental phenomena.

Causal closure can be formulated in a variety of ways, and its wording can have a variable impact on causal propositions. For instance, even Kim has interpreted causal closure in several ways. In a recent interpretation, he defined it as “[i]f a physical event has a cause at *t*, it has a physical cause at *t*.”¹⁷⁵ Kim clearly states that all physical events must have physical causes, but the definition leaves room for the possibility of non-physical causes. These non-physical causes could impact other non-physical phenomena or even physical phenomena so long as it works in conjunction with a physical cause.

Later in the same work, Kim advances a stronger version of causal closure. He writes, “[a]ny cause of a physical event is itself a physical event—that is, no nonphysical event can be a cause of a physical event.”¹⁷⁶ Kim’s stronger definition directly challenges mental causal efficacy by precluding the possibility of non-physical events causing physical events. Simply put, no physical event can trace its cause outside the physical domain.

Popper and Eccles also emphasize that physicalists must take a strong position on the causal closure of the physical. Stronger versions of closure are typically associated with physical completeness, the idea that “every physical effect has a sufficient physical cause.”¹⁷⁷ To have a sufficient physical cause would entail that all other

¹⁷⁵ Kim 2005, 22. Kim has also described closure as “all physical states have pure physical causes.” Kim 1993, 280. Also, he has stated, “[...], no causal chain will ever cross the boundary between the physical and the nonphysical.” Kim 1998, 40.

¹⁷⁶ Kim 2005, 50.

¹⁷⁷ Robb, Heil, and Gibb 2023, 10.

potential causes would be unnecessary additions. As expected, Popper's and Eccles's definition of closure severely limits the possibility of mental causal efficacy; they write:

[...] physical processes can be explained and understood, and must be explained and understood, entirely in terms of physical theories. I call this the physicalist principle of the closedness of the [physical world]. It is of decisive importance, and I take it as the characteristic principle of physicalism or materialism.¹⁷⁸

Popper's and Eccles's formulation of causal closure states that physical effects must be *entirely understood* in terms of physical theories. If it is true that no physical event is caused—partially or fully—from outside the physical domain, then mental events that are understood as separate from the physical domain *cannot be causally efficacious*. As such, Popper's and Eccles's formulation ostensibly eliminates mental causal efficacy unless mental causes can be properly integrated into the physical domain and physical theories.

Given the landscape of physicalist closure, the critical challenge is determining how mental and physical properties are distinguishable yet both physical. I have previously characterized mental properties as qualitative, private, and irreducible. In contrast, hallmark physical properties are quantitative, public, and reducible (unless fundamental). Though these features distinguish hallmark physical properties from emergent mental properties, they do not necessarily indicate that mental properties belong to a separate domain from physical properties. The distinctions simply reveal that we understand them differently, and we can point to our relationship to these properties as the reason for that difference.

Crane remarks that emergentism grapples with twin dilemmas; he writes: “[...], emergentism is generally thought to encounter two problems: first, it is committed to

¹⁷⁸ Popper and Eccles 1985, 51.

mental properties having their own causal powers and second, it is committed to the inexplicability of the mind-body relation.” Crane argues that “nonreductive physicalists must share these commitments. This is problematic for nonreductive physicalists because their physicalism ought to commit them to the causal closure of the physical world, and to the physical explicability (in principle) of the mind-body relation.”¹⁷⁹ So, if emergentism aligns with nonreductive physicalism, it must also align with causal closure.

As noted in the previous chapter, emergentism posits that certain systems exhibit properties or behaviors irreducible to their underlying properties and behaviors. It also utilizes a multilayered framework to reality. The layered conception allows for mental properties to be distinguishable from hallmark physical properties without separating them from the physical domain. Ultimately, I find that this allows emergent mental properties to not only be affected by physical circumstances but also to affect physical circumstances.

All physicalist mental realists will maintain that physical circumstances affect mental properties (e.g., a psychedelic drug altering perceptions). Given this widely held physicalist position, we can safely conclude that the physical and mental are linked. With half the connection already established (i.e., the physical affecting the mental), we only need to determine how the mental affects the physical to complete the interface. In other words, if physical completeness can encompass emergent mental properties, then causal closure, regardless of its stringency, may not be at odds with downward mental causation.

¹⁷⁹ Crane 2001, 210.

Closure In Light of Conservation

The physicalist doctrine of causal closure hinges primarily on the conservation of energy in closed systems (henceforth, Conservation). Since Conservation is an idealized principle, any divergence from its official articulation would lead to considerable uncertainty. I have no intention of undermining Conservation; however, physicalists seem to inappropriately utilize the principle in their formulation of closure.

Feynman conveys the scope of Conservation when he writes, “[t]here is no theorem that says that the interesting things in the world are conserved – only the total of everything.”¹⁸⁰ Here, Feynman states that Conservation applies to the whole of physical reality, and the principle is not meant to apply to any physical system that is less than the totality. Consider, for example, the pacemaker. The pacemaker is a surgically implanted device that interfaces with the heart and nervous system in a feedback loop. Its proper functioning requires detecting the heart’s actual rhythm, processing that information, and sending adjusted electrical impulses to the heart to ensure rhythmic beating. Energy exchanges in and out of the pacemaker, the heart, and the nervous system, indicating that none of these systems are closed.¹⁸¹ This example demonstrates the risks of localizing the concept of causal closure, a principle intended to apply only at the scale of the entire universe, as Feynman suggests.

¹⁸⁰ Feynman [1965] 1990, 83.

¹⁸¹ The following description of a pacemaker was written by cardiologists; they write: “[p]acemakers consist of a pulse generator or can, which contains the battery and electronics, and leads, which travel from the can to contact the myocardium, to deliver a depolarizing pulse and to sense intrinsic cardiac activity [...]. [...] *Sensing* is the process whereby a pacemaker determines the timing of the cardiac depolarization of the chamber that a lead is in. To effectively sense, the pacemaker must optimally sense near-field depolarization signals, reject near-field repolarization signals (T-waves), and reject far-field signals (signals generated by tissues that the lead electrode is not in contact with) as well as non-physiological signals (electromagnetic interference generated by cellphones, and so on.” Mulpuru et al 2017, 191-2.

Since Conservation only applies to the physical totality, this implies that closure cannot be localized as well (i.e., applied to small systems). In other words, applying closure to a localized system, such as the human nervous system, would be an improper application of Conservation and would not be an extrapolation supported by physics. This means physicalists cannot arbitrarily draw a border around the human nervous system (or the body itself) to claim the system is causally closed.

In addition to this fundamental limitation set by Conservation, there are also practical limitations. For instance, we cannot measure the total energy of some small-scale systems, such as the human nervous system. Undoubtedly, the nervous system contains chemical and electrical energy, but our most advanced tools (e.g., fMRI and PET scans) cannot accurately measure that energy. These technologies are revolutionary, but they function primarily by measuring blood flow, which is not representative of total energy.¹⁸² So, not only is the system not closed, but we cannot even determine how efficiently it conserves energy.

To dive even deeper, consider the quantum mechanical phenomena of tunneling. Tunneling is a physical phenomenon where particles (e.g., electrons) move between energy barriers without an apparent cause.¹⁸³ Polkinghorne writes:

¹⁸² Chen et al. writes, “[f]unctional maps of cerebral blood flow (CBF) can be used to monitor hemodynamic changes in the healthy brain as well as alterations associated with cerebrovascular disease. Positron emission tomography (PET) is capable of providing in vivo quantitative measures of CBF and has evolved to be considered the gold standard for studying cerebral hemodynamics. [...] [Arterial Spin Labeling] is based on the detection of magnetically labeled arterial blood water spins and has therefore been used with more success in functional MRI (fMRI) studies.” Chen et al 2008, 1.

¹⁸³ Rather than happening due to an apparent cause, tunneling is understood by way of probability. Polkinghorne writes, “[r]eplacing such picturesque story-telling by precise calculations leads to the conclusion that a particle whose kinetic energy is not too far below the height of the barrier will have a certain probability of getting across and a certain probability of bouncing back.” Polkinghorne 2002, 59

[...] energy is, broadly speaking, a conserved quantity in quantum theory—just as it is in classical theory—this is only so up to the point of the relevant uncertainty. [...], there is the possibility of “borrowing” some extra energy, provided it is paid back with appropriate promptness. [...] The earliest example of a process of this kind to be recognized related to the possibility of *tunneling* through a potential barrier.¹⁸⁴

Tunneling underscores that energy may be impossible to track and small-scale systems are not causally closed.

Given these critiques and the official articulation of Conservation, we must approach causal closure through the lens of the entire (physical) universe. When we utilize a holistic understanding of closure, we embrace a nuanced perspective of how physicalist concepts ought to operate within our complex layered reality.

Montero, in her critical analysis of Conservation and its implications for physicalism, is instructive to my broader points. Montero asks why mental properties are thought to violate causal closure in the first place; she writes:

[w]hat is the argument that is supposed to show that nonphysical minds cannot causally affect physical bodies without violating the conservation laws or, assuming that there are causal relations between the mental and the physical, that physicalism must be true in general?¹⁸⁵

Montero goes on to note:

[f]or if energy is conserved among the physical components of the universe, and if the mental causally affects the physical, and if causation requires the transfer of energy, it would seem that unless the mental only acts as a kind of energy conduit—siphoning energy from one physical thing every time it causally affects another physical thing—the mental must be physical.¹⁸⁶

Montero further refines her argument when she presents the Valid Argument from the

Conservation of Energy, or VACE. She describes VACE as:

¹⁸⁴ Polkinghorne 2002, 58. Italics added for emphasis.

¹⁸⁵ Montero 2006, 385.

¹⁸⁶ Montero 2006, 385-6.

1. Energy is conserved in any closed
2. The universe is a closed system.
3. There are causal relations between the mental and the physical.
4. Causation involves the transference of energy.
5. Anything with energy is physical.

Thus: The mental is physical.¹⁸⁷

I concur with Montero's holistically motivated perspective on causal closure and the physical status of mental properties.

In the next section, I aim to explore how emergent properties fit within this broader, holistic perspective. A holistic perspective that may enable a more comprehensive understanding of downward causation in the context of causal closure.

Holism and Emergence

In this subsection, I address how emergentism and holism can illuminate our understanding of downward causation. This discussion involves delineating holism from reductionism and exploring their interplay.

Reductive and holistic approaches can be viewed as opposing methodologies, but I see them as different lenses that can jointly enhance our understanding of complex systems. The reductive method analyzes a system by breaking it down and demonstrating how its parts contribute to the broader system. Holism, on the other hand, emphasizes that there is inherent value in viewing each system as it functions and interacts as a whole. The holist perspective recognizes that the system can exhibit properties that cannot be deduced from its parts alone. Rather than viewing these contrasting

¹⁸⁷ Montero 2006, 394.

perspectives as mutually exclusive, I find them complementary, each offering insights that the other overlooks.

Emergence exemplifies the interplay between reductive and holist perspectives. A reductive analysis of emergent phenomena scrutinizes components, provides insight into how those components interact, and could identify necessary conditions. For example, reducing hurricanes allows us to recognize necessary components (e.g., barometric pressure, wind speed, ocean temperature, etc.). Additionally, reductions provide insight into the necessary conditions for higher-level phenomena.

However, emergent phenomena, while founded on their constituent parts, exhibit behaviors and attributes that are not deducible from their components. Thus, a comprehensive understanding of emergent phenomena requires more than reductive analysis, it also requires an overarching holist perspective. A holistic look at higher-level and emergent systems reveals a comprehensive whole with functional attributes. I posit that emergent phenomena become part of an interdependent system, capable of affecting behavior and playing a role in maintaining the system. However, this reciprocal causation raises crucial questions about the nature of causation and how it relates to our understanding of the physical world.

The reciprocal interaction between emergent mental states and their foundational physical systems can be elucidated through feedback loops. These loops, prevalent in various complex systems, describe the continuous adaptation that complex systems undergo in response to internal and external changes.

Crucial to this view is recognizing that mental states are not in a separate realm from the physical. Instead, they represent a higher-order level of physical reality, necessitated by and dependent on the complex interactions of the base level from which they emerged. While mental states possess novel properties that distinguish them from their bases, they are not external or unrelated entities. They simply represent another level in a multilayered reality.

Consider the experience of pain as an illustrative case. While an external stimulus may initiate a sensation of pain (e.g., a stubbed toe), the interpretations of this sensation, including its intensity, associated mental distress, and physical reaction, can all be modulated by our mental states. For instance, focusing on the pain can enhance its sensation, activating neural pathways related to distress. Conversely, practices like meditation or positive reinforcement as a motivation to endure the pain can potentially mitigate the sensation. These instances exemplify the complex interplay of mental processes and physical sensations.

The interplay between neural processes and the emergence of mental properties and states as well as the apparent power of the mind to act in intentional ways to affect mental properties and actions suggests the potential for downward causation. At this point, it is worth reflecting on the emergent story. Crowther writes:

[i]n the synchronic conception of emergence, Band E [i.e., emergent phenomena] represent different levels of description: B [i.e., base] is said to describe the system at the lower level and E at the higher level. [...]. In the diachronic conception of emergence, E and B describe the system at the same level. These theories, or models, are supposed to apply to the system at different times, or otherwise under different conditions. The idea

is that the system has undergone some change: typically, B describes it before, and E after.¹⁸⁸

The distinction that Crowther draws between synchronic and diachronic emergence further refines our understanding. Under diachronic emergence, emergent phenomena do not simply coexist with their base in each moment (i.e., synchronically) but serve as a successor to the base state.

Diachronic emergence allows for the emergence of metaphysically significant phenomena and sets up the possibility of a dynamic feedback loop between the base level and the emergent phenomena. This dynamic feedback loop can help bridge the gap between emergent mental properties and their neural substrates. The idea is that there is a necessary two-way interaction between the mind and brain for the benefit of the whole system. This challenges the simplistic causal models traditionally posed by physicalists and instead emphasizes the complex relationship between an emergent mind and its neural substrate.

To summarize this section before turning to examine this dynamic feedback loop idea more carefully, viewing causal closure through a holistic lens fosters a broader understanding of Conservation, emergence, and downward causation. Holism suggests that we should recognize wholes as something beyond the amalgamation of their parts. Approaching reality as a series of interconnected and interwoven complex systems and causal pathways challenges traditional physicalist views without invalidating physicalism.

¹⁸⁸ Crowther 2021, 7286. Bracketed text added for clarity.

Downward Mental Causation

Given the proposed interplay between mental states and their underlying physical states, understanding how emergent wholes influence their parts (i.e., the underlying physical states) is the next essential step. There are many views about causation, ranging from energy transference, counterfactual dependence, primitivism, dispositionalism, and so on.¹⁸⁹ Some of these views are better suited to accommodate downward causation, such as interventionism. Woodward describes intervention as follows:

[...] it is heuristically useful to think of an intervention as an idealized experimental manipulation carried out on some variable *X* for the purpose of ascertaining whether changes in *X* are causally related to changes in some other variable *Y*. However, although the notion of an idealized experiment suggests an activity carried out by human beings, we will see shortly that any process, whether or not it involves human activities, will qualify as an intervention as long as it has the right causal characteristics. The idea we want to capture is roughly this: an intervention on some variable *X* with respect to some second variable *Y* is a causal process that changes the value of *X* in an appropriately exogenous way, so that if a change in the value of *Y* occurs, it occurs only in virtue of the change in the value of *X* and not through some other causal route.¹⁹⁰

At its core, interventionism is about the effects of changing a variable to see how it will affect the effects. Woodward, the founder of interventionism, suggests that causation encompasses any process with the right causal attributes. Such processes could include human-level actions.¹⁹¹

I do not subscribe to one causal theory and do not wish to tie the success of downward causation to a particular theory of causation. While each theory presents compelling points that make them plausible, I doubt that a single causal theory could

¹⁸⁹ See Mumford and Lill Anjum 2013 for an overview of various theories of causation and their shortcomings.

¹⁹⁰ Woodward 2003, 94.

¹⁹¹ In fact, Woodward has published extensively defending downward causation by utilizing his interventionist account. See, for example, Woodward 2003, Woodward 2021a, and Woodward 2021b.

account for colliding billiard balls, gravitational attraction, entangled particles, the observer effect in quantum mechanics, good economic job reports, emergent mental phenomena, and downward causation. For these reasons, I find myself pushed towards a stance of causal pluralism. Nevertheless, I strongly object to causal fundamentalism—the notion that all causation stems solely from the interactions of fundamental particles.¹⁹²

Such a view renders all types of macro causation incapable of influencing their foundational parts. Ultimately, causal fundamentalism is where the CEA leads.¹⁹³ For instance, in his study of emergence, Kim writes:

[t]he idea of downward causation has struck some thinkers as incoherent, and it is difficult to deny that there is an air of paradox about it: After all, higher-level properties arise out of lower-level conditions, and without the presence of the latter in suitable configurations, the former could not even be there. So how could these higher-level properties causally influence and alter the conditions from which they arise? Is it coherent to suppose that the presence of X is entirely responsible for the occurrence of Y (so Y's very existence is totally dependent on X) and yet Y somehow manages to exercise a causal influence on X?¹⁹⁴

Kim questions the coherence of emergent downward causation, suggesting it is paradoxical to believe that a higher-level property, which owes its existence to a foundational state, can, in turn, influence the substrate from which it emerged.

Challenges like Kim's are pivotal to our understanding, yet causal fundamentalism is not exempt from scrutiny. One major critique lies in its assumption that all macro phenomena can and should be reduced to quantum states—an assumption

¹⁹² Friederich and Mukherjee describe fundamentalism as "causation is in fact a pure low-level affair, i.e. it is confined to the most fundamental level of elementary physics that exists." Friederich and Mukherjee 2021, 297.

¹⁹³ For instance, Friederich and Mukherjee write, "[causal fundamentalism] is, notably, the conclusion of the causal exclusion argument, championed by Jaegwon Kim." Friederich and Mukherjee 2021, 297. Bracketed text added for clarity.

¹⁹⁴ Kim 1999, 25.

that Nancy Cartwright argues goes well beyond the evidence. For instance, Cartwright notes:

[t]he easiest way to ensure that no contradictions arise is to become a quantum imperialist and assume there are no properties of interest besides those studied by quantum mechanics. In that case classical descriptions [...] must be reducible to (or supervene on) those of quantum mechanics. But this kind of wholesale imperialism and reductionism is far beyond anything the evidence warrants. We must face up to the hard job of practical science and continue to figure out what predictions quantum mechanics can make about classical behaviour.¹⁹⁵

Furthermore, the complexity of reality and the existence of emergent phenomena are serious stumbling blocks for causal fundamentalism.¹⁹⁶ The special sciences (e.g., biology and psychology) rely on multilayered explanations. In Chapter 3, I addressed the elusiveness of a unified theory and our inability to reduce the theories and concepts of the special sciences to fundamental physics. The explanatory gaps between higher and lower levels are reasons to reject causal fundamentalism, a limiting view that requires a high degree of certainty to which it is not entitled.

By dismissing causal fundamentalism, we pave the way for downward causation. Montero's VACE argument, for instance, provides a theoretical foundation for the interplay between the mental and physical. The VACE argument parallels the dynamics of causal feedback loops seen in biological, computational, and ecological systems. In feedback loops, systems can adapt to variable changes, including internal and external changes, to maintain stability. Applying the function of feedback loops to emergent

¹⁹⁵ Cartwright 1999, 233.

¹⁹⁶ Cartwright provides several arguments against causal fundamentalism. Cartwright 1999, *passim*. For instance, she writes, "[d]espite some formal analogies, however, quantum and classical quantities fall under different mathematical characterisations, and there is no general rule that enables us to represent any classical quantity with a quantum one and vice versa. We do make attempts to formulate a general rule that enables us to represent classical quantities with quantum operators and vice versa, but they have all proven problematic. Cartwright 1999, 217.

mental states would mean that they could affect the underlying physical states (e.g., brain states) from which they emerge. This would mean that mental states are not epiphenomenal byproducts; instead, they would have an essential and active role in guiding system responses by influencing physical states.

By applying the concept of dynamic feedback loops to emergent states, we enrich our understanding of the mental-physical relationship. Emergent mental states are not merely byproducts of complex physical processes; they are essential to processing and guiding system responses, allowing the system to navigate complex reality. Hence, it is not individual neurons, cells, organs, and so on that manage to coalesce in shared goals; rather, higher-order objectives (e.g., survival, prosperity, fame, and so on) can be tied to the emergent state. These higher-order objectives serve a role in orchestrating the underlying substrate and ensuring the continuation of emergent states (i.e., subsequent mental states).

Now, we must determine exactly how downward mental causation works. I concur that physicalism must maintain causal closure, but I find it appropriate to reframe its parameters. A holistic perspective could reconcile causal closure with the challenges of the CEA. This reframing would maintain the integrity of physicalism while recognizing the multilayered complexity of reality and the diversity of explanations required to navigate it.

To explore the possibility of downward mental causation, I think it worthwhile to review several mental-physical causal diagrams to illustrate the objective and the issues. These diagrams elucidate the need for causal closure while helping us to avoid the troubled causal model of interactionist dualism.

Figure 1 illustrates mental states being realized by physical states and direct mental-to-mental causation.



Figure 1: Mental to Mental Direct Causation

If M₁ causes M₂, then P₁ and P₂ must necessarily be present to realize those mental states. The problem is that Figure 1 cannot explain how P₂ comes about.¹⁹⁷ Since I have endorsed emergence as the metaphysical basis of mental states and properties, mental states cannot precede their underlying physical states.

Alternatively, Figure 2 addresses the causal vacuum in Figure 1 by suggesting M₁ causes P₂, which can then realize M₂.

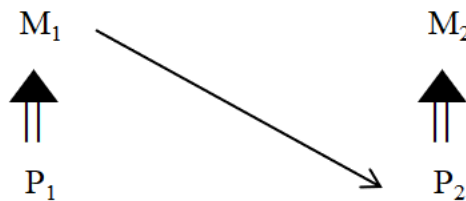


Figure 2: Direct Mental to Physical Causation

Figure 2 is problematic as well for two reasons. First, as illustrated, there is still no P-level causation occurring, which will be objectionable to most physicalists who expect to see causation at that level. For example, the physicalist expects to observe sequences of neurons firing, which serves as an intra-level causal explanation. Second, figure 2

¹⁹⁷ Kim (2005, 15) describes causal closure of the physical as “[i]f a physical event has a cause at *t*, then it has a physical cause at *t*.”

requires an explanation of how M_1 affects P_2 . Without a satisfactory explanation, we have a problem of interaction like the one Princess Elisabeth identified in Descartes's dualism.

Figure 3 attempts to save downward mental causation while recognizing the physicalist desire for P-level causation. In this diagram, both M_1 and P_1 affect P_2 .

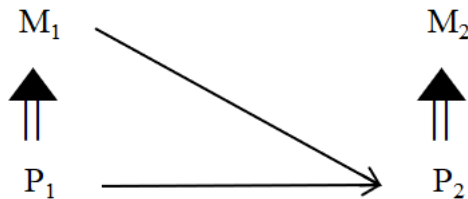


Figure 3: Mental and Physical Causal Overdetermination

There are challenges to claiming there are simultaneous mental and physical causes occurring. Both Kim and Papineau argue that an event can only have one sufficient cause, which, if right, would make the causal influence of the mental on the physical superfluous or redundant.¹⁹⁸ In genuine cases of causal overdetermination, each cause must have a distinct causal role where two independent causal chains intersect at a common effect. However, the concept of a single sufficient cause itself can be problematic, as the very nature of causation remains a matter of debate.

Of these causal diagrams, Figure 3 is the most promising for downward mental causation; however, I think the diagram is overly simplified. In synthesizing these discussions, I propose a framework called “Dynamically Stable Causal Holism.” In the following section, I elaborate on this concept and aim to pave the way for a more coherent understanding of the relationship between the mental and physical. I will also

¹⁹⁸ Kim (2005) describes exclusion as “[n]o single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination” (42). Papineau (1998) writes, “[e]very physical effect has a sufficient physical cause” (375).

propose a fourth and fifth causal diagram illustrating a more accurate depictions of what I am proposing.

Dynamically Stable Causal Holism

In this section, I introduce Dynamically Stable Causal Holism (henceforth “DSC Holism”). DSC Holism serves as a metaphysical framework and bridge between reductive and holistic perspectives. It suggests that while complex systems are underpinned by components, higher-level and emergent phenomena serve a coordinating role and influence the underlying level which in turn influences subsequent emergent states. In other words, in some complex systems, higher levels are essential for empowering the system to achieve higher-order objectives (e.g., beliefs, desires, goals, etc.). This interplay creates a feedback loop wherein higher-level causes play an essential role in sustaining and stabilizing the system.

Consider, for example, the behavior of a flock of birds. A reductive analysis reveals individual birds adhering to simple behavioral rules. For instance, studies using computer simulations have indicated that these rules can be as simple as (i) collision avoidance, (ii) velocity matching, and (iii) flock centering.¹⁹⁹ Abiding by these rules will allow a flock to maintain its structure. However, the reductive perspective overlooks the collective behavior and abilities of the flock. The flock also exhibits collective patterns and behaviors that mutually influence both individual birds and the flock as a whole, creating a *dynamically stable* feedback loop. DSC Holism suggests that while individual

¹⁹⁹ Reynolds 1987, *passim*. Reynolds describes the three rules as “avoid collisions with nearby flockmates, [...] attempt to match velocity with nearby flockmates, [and] attempt to stay close to nearby flockmates.” Reynolds 1987, 28.

behaviors are essential, the flock's higher-level (potentially emergent) behaviors and abilities offer additional explanatory power.

When viewed through the reductive lens, birds follow simple rules. When viewed holistically, a flock exhibits a higher-order purpose that guides each bird's individual behavior. For instance, a murmuration, which is the flight pattern of the flock, creates the appearance of being a larger organism that some predators will instinctually avoid. The flock's collective advantage arises only when pattern flocking.²⁰⁰ Here, the advantage does not stem from any bird's behavior or the reductive rules, but from the collective properties of the murmuration. In other words, reduction misses the higher-level properties of collective intimidation and confusion.

Some critics might argue that systems like bird flocks are mereological constructs without unified agency. Yet, DSC Holism maintains that complex systems, be they ecological, biological, or mereological, have higher-level (and sometimes emergent) properties that cannot be ignored. By focusing on the reductive perspective, we may miss the forest for the trees. Just as understanding an individual tree will not illuminate the ecosystem's intricacies, reductionism alone will overlook the overarching patterns and dynamics that emerge from collective interactions.

Bird flocks serve as a real-world demonstration of nature's complexity. A cursory glance reveals a collection of birds following simple behavioral rules, but a holistic examination reveals a harmony acting in accord with higher-order objectives. Flocking

²⁰⁰ Flocking birds are also more efficient foragers. (Gentry) Richardson et al. hypothesize, "[i]ncreased space use may result from an increased need to expand access to foraging habitat, which in turn increases the requirement for broad-scale information flow. Information flow will depend critically on the social relationships between flock members, which in some respects depend on the type and strength of associations among core individuals." (Gentry) Richardson et al, 2022, 2-3.

birds are, of course, only one such example. The human body is another prime candidate. A purely reductive approach might emphasize the function and role of cells in determining macro characteristics; however, DSC Holism underscores that the larger system itself plays an essential role in guiding cellular behavior.

DSC Holism suggests a broader, more integrated approach to understanding causation. It advocates for acknowledging and valuing the influences and effects at all system layers, leading to a richer and more holistic understanding of the world around us. DSC Holism tells us that complex systems may adapt, evolve, or transform to maintain system stability. Consider, for example, Reid's explanation of natural selection; he writes:

[w]hat if, without prejudging alternatives, natural selection were better understood as the hypostasis (or imposer) of dynamic stability? That is how selectionists actually see it, when they open their eyes. The conventional journey has already taken us close to that vista since normalizing selection and "stable evolutionary" strategies are the most common conditions. The exception, directional selection, may only be the consequence of autonomous amplifications and epigenetic constraints. The reverse of the question is more important: what might be biologically possible in the absence of the agents of natural selection?²⁰¹

In this quotation, Reid suggests that natural selection might be more than an evolutionary mechanism; it also drives "dynamic stability." The concept of hypostasis suggests that natural selection is not just about how traits emerge but why populations stabilize.

To summarize, DSC Holism serves as a framework for understanding complex systems. It is closely aligned with emergentism and emphasizes that both reductive and holistic perspectives are necessary to comprehend a complex system. By looking beyond individual components, we also recognize higher-order patterns, behaviors, properties,

²⁰¹ Reid 2007, 407-8.

and purposes. It is through the interplay of lower-order and higher-order (i.e., parts and wholes) that we recognize the interplay occurring within systems and enrich our understanding of a complex interconnected reality.

DSC Holism, Emergentism, and Downward Causation

The intersection of DSC Holism, emergentism, and downward causation warrants a nuanced examination. To this end, this subsection delves into a general description of how DSC Holism might allow for downward causation by reframing our understanding of the causal interplay occurring within complex systems.

DSC Holism contends that higher-level phenomena, including emergent mental states, are not passive outcomes of their underlying components. Instead, they actively guide underlying activities, steering the system toward stability. By “stability,” I mean the balance or equilibrium a complex system pursues to maintain itself, including the ability to achieve intended or desired outcomes. I classify the attainment of outcomes as "stability" because a system’s ability to achieve desired outcomes signifies the system’s effective coordination, which requires harmonious interplay between the system’s higher and lower levels.

DSC Holism maintains that the interplay between higher and lower systems is not additive or subtractive but a dynamic redistribution of a system’s existing energy. Thus, under DSC Holism, understanding Conservation and causal closure holistically becomes pivotal. When Conservation is perceived not as discrete components conserving energy but as the whole system utilizing energy equilibrium to achieve stability, the idea of the higher level influencing its lower underlying parts still aligns with causal closure.

One of the prominent criticisms of emergence, and especially the possibility of emergent downward causation, is Kim's CEA. Kim argues that coexisting mental and physical causal powers would lead to causal overdetermination. Basically, if both higher-level mental properties and lower-level (i.e., underlying) physical processes have causal influence, then there is causal redundancy. Moreover, given the evidence Kim believes we have for physical causation, we should prefer reductive physical causation and exclude mental causation. Instead, if we opt for redundancy, we have widescale causal overdetermination, which must be addressed.

Kim would likely find DSC Holism to rely on causal overdeterminism to function. I will explore the overdeterminism critique in the final subsection, but for the time being, I will emphasize that both higher levels and lower levels of causation are integral to the system's stability. Under DSC Holism, causality is multilayered and multidirectional. These multilayered causal influences must coexist to explain the system's stability as defined. In a complex system, like a human, the highest layer is the mind. It is in the mind that intentions and objectives are formed, which play an essential role in coordinating behaviors so that the system can maintain stability.

To help illustrate what I am describing, consider the following analogy. The underlying structure of a computer is its hardware. The hardware is essential to the computer, allowing the computer to run in a basic state without malfunctioning. However, computers also need software to achieve their potential, including an operating system and installed programs, to function as a useful interactive interface. The higher-layer software allows the computer to process information in specific ways and provide desired outputs. In effect, the software changes the hardware's behavior, altering the

processes and electrical currents of the hardware's basic state. In other words, it is both the software and hardware working together that allows the system to complete its higher-order tasks. While this analogy does provide a real-world example of a complex system that requires both higher level and lower levels to function to its full abilities, the analogy does break down. For one, a system, like a human, is far more complex than hardware and software analogy. The human brain and mind represent far greater sophistication than the computer, and the types of inputs humans are subjected to and the possible outputs far surpass those of our cutting-edge technologies.

Humans, influenced by a wide variety of desires and their societal structures, must adapt their strategies to achieve stability. DSC Holism recognizes this complexity and posits that our base physical conditions cannot be solely responsible for the wide variety of behaviors and outcomes humans achieve. Instead, those underlying physical properties are steered by the emergent mind, which is itself capable of navigating through a myriad of external inputs (e.g., societal expectations) as well as internal challenges (e.g., anxiety). This perspective reinforces the idea that causality in complex systems is multilayered and dynamic, involving a reciprocal relationship between higher and lower levels. It is not just that higher levels emerge from lower levels. They also provide feedback that influences lower levels. Ultimately, the higher and lower levels create a feedback loop of mutual causality.

DSC Holism underscores the nuanced interdependence within complex systems. Under this view, higher and lower levels operate as a feedback loop, where higher-level properties can reciprocally influence their underlying components. The mutual causality inherent in some systems signifies that higher levels are not passive phenomena; instead,

they are dynamic participants striving for stability through influencing the lower level. In other words, DSC Holism tells us that this causal interplay is necessary because it *dynamically* drives the system and its constituent parts toward *stability*. Through this understanding, DSC Holism serves as a framework, reframing our perspective on emergentism and downward causation in the face of physicalist critiques.

Mental Casual Efficacy

In this section, I overview how DSC Holism allows for downward causation. The framework of DSC Holism promises a comprehensive approach, hinging on emergentism and holism. As a brief recap, emergentism suggests that some complex physical processes give rise to unpredictable novel properties and phenomena. Holism suggests that isolating the parts of a whole will not provide a comprehensive understanding of all macro phenomena; instead, the whole itself must also be analyzed in at least some cases. Both concepts are pivotal to DSC Holism.

The central challenge is to explain the viability of downward causation in a physically viable manner, and the CEA is the main driver for doubting the physical viability of mental causal efficacy. As a refresher, Kim's CEA is a four-premise argument:

1. the mental is not identical to the physical,
2. mental properties cause both physical and mental effects,
3. the causal closure of the physical—the physicalist tenet that every physical effect has a sufficient physical cause and

4. causal overdetermination is rare and, therefore, we cannot justify large-scale systematic mental causation.²⁰²

We can slightly revise premise 1 based on our discussion so far. It is not that the mental is non-physical; rather, a more accurate premise is 1* the mental is not identical to reductive physical states. According to the revised argument, every caused event, such as neuron x firing, must have a sufficient physical cause. Kim finds only reductive physical causes to be sufficient causes. Moreover, he finds that these reductive causes can explain all macro events. Considering this, any effect involving both a mental and a physical cause can be viewed as causally overdetermined.

Kim's CEA demands that DSC Holism respond to three questions. They are (1) Efficacy, or how does a higher-order cause alter its parts; (2) Closure, or where does the higher-order cause originate; and (3) Overdeterminism, or is widespread overdeterminism necessary to explain downward causation?²⁰³ I will proceed in order, and begin by addressing Kim's overarching conclusions about downward causation and why we should doubt them.

Efficacy

Kim advocates for a reductive approach to causation, suggesting that the only way to uphold mental causation is by reducing mental processes to physical ones. He writes:

[s]ome philosophers are still gamely holding on, trying to somehow fashion an account of mental causation within the nonreductive scheme, but I believe that if we have learned anything from the three decades of debate, it is the simple point that unless we bring the supposed mental causes fully into the physical world, there is no hope of vindicating their

²⁰² See Vision 2011, 95-7 for an overview of the causal exclusion argument.

²⁰³ Vision prefers "massive causal overdetermination," whereas I prefer "widescale causal overdetermination." See Vision 2011, 95 *op cit*.

status as causes, and that the reality of mental causation requires reduction of mentality to physical processes, or of minds to brains.²⁰⁴ This quotation notes Kim's skepticism of non-reductive schemes while affirming his commitment to embedding mental causes within the underlying physical structure. Kim sees this embedding, typically understood as a form of identity, as the only way to attribute a causal role to mental properties. For Kim, the essence of mental causation demands reducing mental processes to physical processes.

Many physicalists sympathize with Kim's position, but his bias is apparent. He states explicitly that three decades of non-reductive attempts have yielded no gains. However, the same can be said about reductionist attempts to diminish mental properties. For instance, in Chapter 2, I addressed attempts to reduce mental properties to physical processes and explained how they come up short.

Kim's position on reductive physicalism is by no means flawless. In fact, Kim acknowledges the shortcomings of reductive physicalism, while finding it to be as close to truth as he thinks we can get. He writes:

[t]he position is, as we might say, a slightly defective physicalism—physicalism *manqué* but not by much. I believe that this is as much physicalism as we can have, and that there is no credible alternative to physicalism as a general worldview. Physicalism is not the whole truth, but it is the truth near enough, and near enough should be good enough.²⁰⁵

Here, Kim attempts to solidify reductive physicalism as the best approach to mental realism. However, a more critical reading indicates that the reductive physicalist project also fails to achieve its ends; it only manages to get *near enough*. Kim may find that

²⁰⁴ Kim 2005b, 156.

²⁰⁵ Kim 2005b, 174.

satisfactory, but I find it biased and am skeptical of the supposed success of physicalism *manqué*.

Vision rebuts the idea of causal drainage, which Kim uses to support his reductive approach.²⁰⁶ Vision argues that causal drainage—the idea that causation at higher levels (e.g., mental) is entirely accounted for by lower levels (e.g., atomic interactions)—is a poor reason to dismiss non-reductive approaches to mental causation on the basis that it is an oversimplification of the causal landscape. Vision contends that even if higher-level properties derive their causal powers from lower-level properties, derivation does not negate the reality and efficacy of higher-level causation. I also find the reductive approach to be myopic and believe it can only reveal an incomplete picture. Instead, I find that DSC Holism does not deny the metaphysical necessity of parts to wholes while also emphasizing that reduction alone misrepresents reality. Ultimately, DSC Holism is better understood as a holist perspective on the reductive paradigm. Through a wider view of complex reality, the rich interplay of parts and wholes provides a more complete perspective. For example, Maudlin writes:

[t]he physical state of a complex whole cannot always be reduced to those of its parts, or to those of its parts together with their spatiotemporal relations, even when the parts inhabit distinct regions of space. [...]. The result of the most intensive scientific investigations in history is a theory that contains an ineliminable holism.²⁰⁷

To put it plainly, DSC Holism puts forth the best explanation for the distinction between the mental and physical while leaving the two interconnected and causally intertwined.

²⁰⁶ Vision 2011, 116-21.

²⁰⁷ Maudlin 1998, 56.

One factor motivating DSC Holism is that the ghost of reductionism still haunts non-reductive physicalism. Though it is called non-reductive, the physicalist expectation is that the reductive framework is rigid and only makes exceptions for the existence of the non-reductive phenomena. In other words, reductionism may not be able to explain the phenomena, so they can be classified as non-reductive, but that phenomena must still play by reductive rules. One way to ensure non-reductive phenomena play by the rules is by insisting on a supervenience relationship between non-reductive phenomena and lower-level phenomena. So long as there is a firm expectation of reductionism, non-reductive phenomena having causal influence will be incompatible. DSC Holism, on the other hand, tells us the reductive framework is too myopic, leading to inconsistency.

Being largely inspired by emergentism, DSC Holism suggests that reality is akin to nested complex systems, where each nested system is part of a multilevel feedback loop reaching from the fundamental level to the whole of reality. Such a perspective becomes illuminating when we consider the complexity intrinsic to various systems and their objectives.

Various complex entities, like the human body, which is constructed out of DNA molecules replicating, cells interacting, the brain firing, and the mind emerging from the brain, are all intertwined layers of intricate complex systems. Each system operates in a manner best described at their respective levels (see Chapters 2 and 3). While DSC Holism acknowledges that there are foundational components necessitating each higher system, each layer overlaid plays a pivotal role in coordinating the broader behavior of the whole system. For instance, to suggest that complex behaviors, like choosing a career or creating art, arise solely from rudimentary cellular actions seems overly simplistic.

DSC Holism maintains that the emergent mind plays a necessary role in allowing the system to navigate reality. For example, the mind allows a person to live through and understand their inherent talents and dispositions while navigating the contextual landscape of the world and acting on the mind's intentions. From the DSC Holist perspective, the downward causation that explains the relationship between creating art and cellular activity is not vastly different than how the epithelial cells in a ballet dancer's feet do not decide to be on the left side of the stage. The location of those cells is determined by the choreography, which was created in the choreographer's mind, which was then intentionally followed by the dancer.

DSC Holism weaves the mental and physical together in a manner that neither reductive nor non-reductive approaches have managed. Instead of focusing on reductionism or non-reductionism, DSC Holism emphasizes the complexity of reality and the importance of understanding the layered interactions therein. DSC Holism is a fresh perspective, a nuanced look allowing us to appreciate the intricacies of the mind-body relationship.

Closure

Traditional understandings of causation, such as the views espoused by Kim, emphasize the importance of bottom-up processes. In these causal frameworks, underlying processes wholly shape overlaid phenomena. The aim of this section is to explain how DSC Holism understands downward causation while also addressing reductive critiques about closure. Since the CEA addresses mental properties explicitly and because the overarching focus of this work has been mental causal efficacy, I will

primarily utilize the human mind to explain how DSC Holism explains downward causation.

To begin, let me reassert my stance on emergence, which I previously endorsed as a key metaphysical model for comprehending minds and mental properties. Emergence posits that the human mind is a product of the complex processes of lower-order systems, culminating in a metaphysically distinct layer that overlays those underlying processes. In an average human brain, this overlaying layer will consist of emergent properties like qualia, beliefs, intentions, and more. These emergent properties, while resistant to reductive interpretations, stand on their own according to mental realists.

The emergent mind's causal "energy" is fundamentally distinct from the traditional reductionist form of energy (i.e., energy transference). In other words, the emergent mind's energy is not equivalent to the kinetic energy exchanged between billiard balls.²⁰⁸ The energy required for a mind to emerge stems from biological processes (e.g., consuming food and burning calories), but the downward causal power of the emergent mind cannot be found in cellular energy stores and electrochemical reactions. If that were the case, a simple identity relationship would suffice. Undoubtedly, the reductionist will be curious about how an emergent mind causally influences its foundational components (e.g., individual neurons and neural pathways).

Chalmers suggests that while consciousness might supervene on the physical, it may necessitate its own fundamental laws distinct from physical laws. He writes:

[if] consciousness [...] is responsible for wave-function collapse, the emergent quality of consciousness is not epiphenomenal but plays a

²⁰⁸ Such a position would tread too closely to causal interactionism and simply would reopen the critique presented to Descartes by Princess Elisabeth. Elisabeth of Bohemia and Descartes 2007, 61.

crucial causal role. [...] I do not know whether there is any strong downward causation, but it seems to me that if there is any strong downward causation, quantum mechanics is the most likely locus for it. If both strongly emergent qualities and strong downward causation exist, it is natural to look at the possibility of a close connection between them, [...].²⁰⁹

Chalmers suggests that if emergent mental states play a role in wave-function collapse, these states cannot be redundant; instead, they would be causally significant. I am unsure whether it is necessary for emergent mental states to bring about quantum events directly or if it is possible to affect the brain directly; however, I believe the exact mechanics of how higher-order emergent mental processes influence their underlying constituent parts cannot be fully understood. Since emergence is a brute process (see Chapter 3), we cannot expect to fully understand or demonstrate with any degree of certainty the causal relationship between higher-order emergent phenomena and their underlying parts. In other words, emergence is a metaphysically brute relationship, and this relationship does not only apply in one direction. If it is brute in the upward direction, then it can also be brute in the downward direction as well.

I can speculate, just as Chalmers does, that the emergent mind interacts with its underpinnings through causal feedback loops in the form of regulatory mechanisms. To illustrate, consider the effects of repeatedly walking across grass. Over time, the grass will wear away, and a well-trodden path will form. I propose that something parallel occurs within the mental-physical relationship. Similarly, imagine neural pathways forming and strengthening due to repeated behaviors, just as repeated walks over grass create a path. Human behaviors, shaped by a myriad of complex inputs (e.g., societal

²⁰⁹ Chalmers 2008, 250.

expectations, parental guidance, inherent dispositions, etc.), aid in solidifying those pathways. These structural constraints become our default actions, and the idea aligns well with how habits and learned behavior can become like second nature.

There is a relationship between the emergent mind and the underlying structure from which it emerged. The underlying structure comes with structural constraints that play a role in determining our actions without the need for mental intervention; however, the emergent mind retains the potential to process information in a way not entirely constrained by those structural constraints. Since the emergent mind exists on a higher level, it cannot simply be reduced to those structural constraints. To elaborate, the emergent mind need not be deterministically bound to individual structural constraints. Those structural constraints will urge one to act in a given way, but the emergent mind, realized by many underlying components and informed by other higher-level considerations, can act against default actions. In other words, the non-reductive, emergent relationship, supported by the layered conception, affords the emergent mind some freedom to act otherwise.

Expanding on this concept, imagine an anxious student attending a lecture. They listen to the content intently, process it, and, after much consternation, decide to raise their hand to ask a question. This seemingly simple action may be an internal struggle for the student reinforced by their neural structural constraints (e.g., anxiety, low confidence, etc.). Through the lens of DSC Holism, the decision to raise their hand is not an example of energy transfer, but an informational feedback loop between the emergent mind and the neural foundations. This feedback loop represents the constant two-way communication between the emergent mind and its neural foundations. After the brain

processes external stimuli, it rises to conscious awareness. Subsequently, the emergent mind synthesizes this information with memories, emotions, passions, curiosities, and so on. DSC Holism maintains that this synthesized information is then fed back to the neural network, modulating which neural pathways fire.²¹⁰

Interestingly, DSC Holism aligns with Ego Depletion, a controversial theory in psychology. This theory proposes that willpower is a finite resource, and continuously resisting ingrained behaviors or habits can deplete this abstract reservoir.²¹¹ Consider, for instance, the habitual drinking of an alcoholic. The neural pathways associated with drinking are continuously reinforced, making resisting the urge extremely difficult. However, DSC Holism suggests that the emergent mind can resist structural constraints and influence behavioral changes.

To be clear, the struggle to overcome an addiction is no small feat. Ego Depletion suggests that resisting alcohol consumption requires actively engaging one's willpower. Possibly for that reason, the success of abstaining can often be tied to several techniques that aim to help slow the depletion of willpower (e.g., avoiding bars and other people drinking alcohol). This commonplace occurrence and struggle for many speaks to the challenge the emergent mind has in reshaping structural constraints (i.e., well-trodden neural pathways that lead to embedded behaviors). It suggests that while the emergent mind can override structural constraints, doing so requires focus and conscious action.

²¹⁰ Here, one can anticipate the reductionist asking, "how?" The DSC Holist may simply have to say, we cannot know for certain due to the nature of emergence. Someone like Chalmers might speculate that it can be tied to the wave-function collapse. He writes, "[if] consciousness [...] is responsible for wave-function collapse, the emergent quality of consciousness is not epiphenomenal but plays a crucial causal role." Chalmers 2008, 250.

²¹¹ Baumeister et al. 1998.

DSC Holism's causal diagram resembles the following two scenarios:

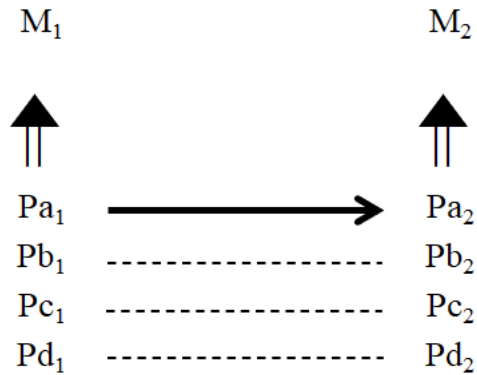


Figure 4: Emergent Mind without Intervention

In Figure 4, M_1 represents an emergent mind, realized by all its underlying processes. M_1 has the ability to look introspectively upon at least some of those processes and has a sense of what they are capable of making happen. For example, Pa_1 may be a compulsion to do x , while Pb_2 is a desire to change that behavior and not do x . The emergent mind is conscious of at least some of these underlying processes while remaining unconscious to others. All the underlying processes ($Pa_1, Pb_1, Pc_1, \dots Pn_1$) carry on in various ways, some potentially carrying all the way through to action and effect. In this example, let us assume that Pa_1 represents a default action reinforced by past behaviors. Though Pa_2 comes to be, $Pa_2, Pb_2, Pc_2, \dots Pn_2$ all play a role in realizing M_2 , empowering M_2 to be aware of the action and at least some of the alternative possibilities that could have been.

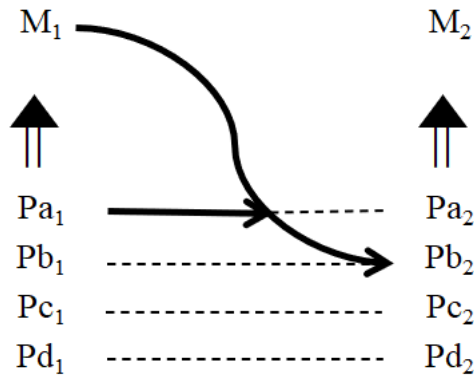


Figure 5: Emergent Mind with Intervention

Figure 5 represents a similar scenario to Figure 4, except that the emergent mind, M_1 , has intervened by redirecting the default cause Pa_1 toward making Pb_2 the outcome. This is possible because M_1 was informed by $Pa_1 \dots Pn_1$ and capable of foreseeing higher-level considerations like the consequences of $Pa_2 \dots Pn_2$ as well as other same and higher-level considerations, such as societal expectations, onlooking bystanders, and any number of reasons that might affect one's desire to act otherwise.

Thus, DSC Holism posits a nuanced relationship between the emergent mind and its underlying infrastructure. It tells us that the mind is unlike libertarian free will or Cartesian substance dualism because the mind is not the master of the body. Rather, the emergent mind, informed by all its underlying architecture and higher-level considerations, exists in a symbiotic relationship within the whole complex structure. It is through mutual *causal* influence that the system (as a whole) can *dynamically* achieve *stability*.

One important thing to note, since the above diagrams will be somewhat misleading, is that the emergent mind is not simply sparking particular neurons into action. Rather, there is a dynamic back-and-forth or interplay between the various

components of this complex system. Moreover, as noted, whether the emergent mind can directly influence wave-function collapse as Chalmers speculates, or if it is something else entirely, emergence suggests we cannot know because the relationship is metaphysically brute.

By recognizing that wholes have causal roles of their own, DSC Holism compels us to recognize the collective dynamics of a system. In certain complex systems, notably conscious humans, these dynamics have the potential to reshape or even counteract the inherent behaviors of individual components. The emergent mind can affect at least some human actions through causal feedback and dynamic stability.

Causal Overdetermination

This brings us to the third question, namely does DSC Holism entail causal overdetermination. The CEA's causal closure premise maintains that every physical effect must have a physical cause. If there are distinct mental causes, then there must also be physical causes to explain physical state changes. So, if it is proposed that M_1 can cause P_2 , it means that P_1 is necessary to bring about P_2 . Kim's position is that only the physical cause (P_1) is necessary to bring about the physical effect (P_2), making M_1 an overdetermined cause. Arguably, redundant causes threaten the notions of causal closure and physical completeness, serving as a type of metaphysical baggage accumulating with each mental cause.

Notably, overdetermination is not an inherently problematic philosophical issue. There are cases where physical effects arise from concurrent causes. An often-used

example is of a firing squad where two bullets enter a heart simultaneously.²¹² The CEA's real focus regarding overdetermination is to problematize widespread overdetermination, which would happen if Kim's notion of downward mental causation were the case (see Figure 3). This formulation of downward causation would be widespread because, according to causal closure, every mental cause would also need a corresponding physical cause so that every mental cause would be packaged with a distinct physical cause. This leads to the accumulation of a type of metaphysical baggage. Thus, the argument against widespread causal overdetermination is that because lower-level physical causes sufficiently explain all effects, the introduction of mental causes is superfluous.

However, DSC Holism provides an alternative perspective. First, it rejects causal fundamentalism. Second, by approaching causation from a holistic perspective, DSC Holism reconceptualizes the causal relevance of mental properties. Where proponents of the CEA see one sufficient cause and one overdetermined cause, DSC Holists see the necessary interplay between the mental and physical in a complex multilayered system. When we understand causation holistically, mental causes are not seen as redundant; rather, they are a necessary causal contributor to any outcome that does not follow a default path. Thus, DSC Holism aims to dilute the force of the CEA. It does this by pointing to the reductive emphasis of the CEA and finding the purely reductive approach to be insufficient for explaining multilayered interlevel causation.

To elucidate this point, consider a game of pool. From a reductive perspective, a frame of pool boils down to basic physics. The kinetic energy stored in a struck ball is

²¹² See, for example, Kroedel 2019, 167-8.

transferred to another ball, and this transference of energy occurs until the 8-ball's kinetic energy terminates in one of the table's pockets. However, the reductive perspective omits significant details of the game, such as the player's strategy and nervous demeanor, as well as any other higher-level elements that influence the game.

DSC Holism emphasizes that, in complex systems, certain causal narratives cannot be captured by reductive explanations alone. Since emergent phenomena are unpredictable, irreducible, and novel, their presence would introduce new causal dimensions that a purely reductive approach would overlook. While DSC Holists do not refute the validity of the basic physical explanation of the pool game, they would find that it is just one layer of a more intricate causal story and is therefore limited.

DSC Holism's causal position aligns neatly with emergentism. While emergentism holds that complexity gives rise to higher levels with novel properties and behaviors, DSC Holism suggests that emergent phenomena can impact lower levels. According to DSC Holism, emergent phenomena's ability to influence the foundational components from which they emerged does not require introducing new energy into the system. Instead, it involves a nuanced redirection and coordination of existing processes. These idea aligns with Reid's idea of hypostasis, wherein a system will intrinsically aim for stability by adapting as needed.²¹³

DSC Holism does not ignore the overdetermination criticism; instead, it reframes the issue. Rather than view causation through a myopic reductive lens, DSC Holism encourages zooming out and embracing a richer causal environment. This perspective perceives what might be labeled overdeterminism in a purely reductive worldview as an

²¹³ Reid 2007, 407-8. See Chapter 4's Dynamically Stable Causal Holism section.

inherent feature of sophisticated multilayered systems. DSC Holism, in effect, undercuts the power of the overdeterminism objection. Importantly, DSC Holism does not posit new energy sources but rather claims that existing energies are redirected through the interlevel back and forth. By embracing the multilayered complexity of emergent phenomena, DSC Holism recognizes the possibility that some things could be otherwise. In humans, the power to make those things be otherwise is made possible by the downward causal powers of the emergent mind.

Thus, DSC Holism challenges the overdetermination concern by offering a different perspective of causation. DSC Holism promotes a richer understanding of causation, one that sees overdetermination not as a problematic anomaly but as an integral feature of complex systems. Higher-level phenomena and properties do not insert extra energy into the system, they redirect existing energies at lower level through causal feedback loops. It follows that in cases where there is no downward causal power, then lower-level events would follow a purely reductive causal perspective and the macroworld would unfold per physical expectations.²¹⁴ In cases where there are emergent phenomena, both lower and higher levels are crucial to the causal story. As I see it, this sidesteps concerns about overdetermination.

In summary, DSC Holism provides a more comprehensive look at causation beyond the misleadingly simplified reductive perspective. Instead, it underscores the need for a nuanced, multilayered understanding of causal relations. Through its ability to accommodate multiple theories of causation while respecting emergentism, Conservation,

²¹⁴ In essence, rendering the idea of philosophers' zombies nonsensical. It is nonsensical not because they are physically impossible, but because of the idea that philosophers' zombies would behave identically to humans possessing emergent minds. See Aulisio 2022.

closure, and the intricacies of complex systems, DSC Holism offers a more robust and versatile framework for understanding our complex reality.

Emergence Revisited

The nonlinear emergence that I present in this work captures the idea of emergent phenomena as the culmination of an evolving series of processes. In the case of the conscious mind, it is not the static outcome of a specific neural state; instead, it is very many neural processes, some of which can be more well-defined than others, coalescing in a dynamic, ongoing process. The emergent phenomena, in this case, a particular mental state, is influenced by many factors, including behavioral predispositions, environmental considerations, and contexts. It is further influenced by things like weakened emotional states or mind-altering drugs, as well as internal regulatory mechanisms. The emergent mind adapts, evolves, and reorganizes itself in response to stimuli and new information.

This nonlinear theory of emergence rejects the supervenience model because we cannot neatly tie the emergent state to a static underlying state. The supervenience emergentist claims a slice of time will show that M_1 supervenes on P_1 , but the nonlinear emergentist can reject the simplicity of this view and instead note that M_1 is the result of many preceding P states, and subsequent P states will be further underwritten by M_1 and subsequent M states. Making the idea of any given emergent state supervening on a physical state too narrow to capture the fluidity of emergence. In other words, the emergent mind is not just passively determined by its underlying components, it also actively affects the underlying substrate from which it emerged.

My view of emergence recognizes the dynamic interplay between the parts in an emergent system, including the emergent mind and its downward causal ability to affect the underlying substrate from which it emerges. Under this view, the emergent mind is dynamic and continuously influencing and being influenced by its underlying substrate. This picture of emergence highlights that each new experience or thought does not exist in isolation but is layered upon existing conditions (e.g., the underlying structure). Thus, present emergent instances of the mind are influenced by the underlying “hardwiring” in place. The dynamic, evolving nature of the emergent mind and the semi-stability of the underlying substrate underscores a holistic understanding of the emergent mind, where both change and continuity play integral roles.

To elaborate further, consider the following. The emergent mind is dynamic and adaptable, but there is continuity because of the underlying substrate. The underlying substrate’s hardwired aspects could exist on a spectrum from "permanent" (i.e., unalterable by way of downward mental causation) and “deeply etched” (i.e., behaviors or traits that are difficult for us to influence) to “easily defeasible” (i.e., able to be altered by downward mental causation or by external pressures).

For instance, consider the role that memory could play in this theory. Memory serves as a foundational element that anchors the emergent mind. It ensures that despite the ongoing adaptation of the emergent mind, the hardwired aspects serve as the thread of continuity. This continuity is crucial for identity and behavioral consistency over time. The emergent mind and self, therefore, are not just a product of immediate stimuli or current cognitive processes but also a culmination of past experiences, learned behavior that has been "written" into the neural pathways, and the processing of information.

As another example, consider the ability of strong emotions to hardwire us in certain ways. Emotional experiences, whether joyful or sorrowful, can carry with us and influence how future stimuli are processed, affecting the emergent mind. In cases like these, the initial strong emotion underwrites and hardwires the brain affecting future emergent states. The individual affected by this hardwiring may need to make a concentrated effort to overcome what has been deeply etched within them, such as when they face their fears or address the underlying emotions so that they can alter their future emergent mental states. One can picture cases where this can be successful, as well as cases where this may not be possible and even cases where it is possible but extremely challenging, requiring multiple sustained efforts to instill positive change.

As a lighter example, it is easy to imagine a habit, such as fingernail biting entering people's regular behavior. The level of difficulty one has at overcoming the habit depends on many factors, but it is easy to picture an individual who, through concentrating on the urge, can defeat the habit in most situations and eventually rid themselves of the urge. It is also easy to picture another nailbiter who can only defeat the urge in public or professional settings but continues to succumb to the urge in other settings.

Conclusion

As I explained in Chapters 1 and 2, we require a viable mental-physical relationship once we commit to mental realism. I find emergentism, which accepts that the mind is a real but non-reducible higher-level phenomenon, to be the best explanation for mental realism. The emergent mind allows us to perceive reality in the ways we do,

formulate beliefs, process a vast amount of internal and external information, and synthesize those details into decisions and actions.

Supervenience emergence is limited in its ability to address nonlinear phenomena and indeterminism in nature. Moving beyond the constraints of supervenience, my theory of emergence better accommodates a wide range of phenomena, where emergent properties often arise from nonlinear interactions that are not strictly predictable from their constituent parts. This approach, therefore, offers a more inclusive and applicable framework for understanding emergence in a broader range of contexts.

Nonlinear emergence suggests the importance of an interdisciplinary understanding of the emergent mind. Understanding the emergent mind is not in the domain of any one discipline; instead, several disciplines will offer glimpses into a better understanding of the mind and ourselves. However, because emergence is metaphysically brute, the culmination of those glimpses will not reveal a complete picture. However, those disciplinary perspectives jointly offer a clearer picture of reality, underscoring the importance of a holistic approach to understanding emergent phenomena.

While emergentism provides the metaphysical foundation, DSC Holism delivers the theoretical framework for understanding the relationships between the layers of hierarchical complex systems. DSC Holism and emergentism offer a compelling lens through which we can understand mental causal efficacy. Emergentism suggests that the mind represents a distinctive higher-order layer realized by complex neural processes.

DSC Holism challenges the idea that causation is a one-way, bottom-up process by underscoring the multilayered bidirectional nature of complex systems. It suggests that higher-order systems, like the mind, can influence the same components from which it

was realized. DSC Holism also posits that some complex systems have inherent objectives that serve a role in orchestrating their operations and guiding their constituent parts toward achieving certain ends.

When complex systems introduce emergent phenomena, they also introduce novel properties and behaviors that further shape a system's trajectory. The mind's novel properties explain our ability to formulate diverse objectives, solve complex problems, and navigate complex reality in ways that allow system stability. Emergent phenomena like the human mind do not exist in isolation. To achieve system stability, emergent phenomena must actively interact with their foundational components. The human mind's ability to process, interpret, and synthesize information plays a role in guiding actions and behaviors. These mental determinations affect the brain and, in turn, influence subsequent events. In other words, the emergent mind and underlying substrate are in a dynamic feedback loop—where emergent properties and their foundational components exist in a symbiotic relationship of dynamic stability.

By integrating DSC Holism and emergentism, we arrive at a reframed view of downward mental causation. Rather than viewing the mind as a byproduct of brain processes (i.e., an epiphenomenon), this framework elevates its role, making it capable of influencing, just as it is influenced by, underlying physiological processes. Accepting this intricate causal narrative encourages us to recognize higher-level systems as essential threads in the vast causal fabric of our reality.

Thus, we have causally efficacious downward mental causation.

BIBLIOGRAPHY

- Alexander, Samuel. 1920. *Space, Time, and Deity*. London: Macmillan.
- . 1927. *Space, Time and Deity*. Vol. II. London: Macmillan and Co., Limited.
- Aristotle. (1984) 1995. *The Complete Works of Aristotle*. Edited by Jonathan Barnes. 6th ed. Vol. Two. Princeton: Princeton University Press.
- Aulisio, George J. 2022. “Epiphenomenal Minds and Philosopher’s Zombies: Where Do Mental Properties Originate?” *National Taiwan University Philosophical Review* 64 (Special Issue on Self and Other): 267–300.
- Batterman, R. W. 2000. “Multiple Realizability and Universality.” *The British Journal for the Philosophy of Science* 51, no.1 (March): 115–45. <https://doi.org/10.1093/bjps/51.1.115>.
- Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. “Ego Depletion: Is the Active Self a Limited Resource?” *Journal of Personality and Social Psychology* 74 (5): 1252–65. <https://doi.org/10.1037/0022-3514.74.5.1252>.
- Bedau, Mark A. 2010. “Weak Emergence and Context-Sensitive Reduction.” In *Emergence in Science and Philosophy*, edited by Antonella Corradini and Timothy O’Connor, 46–63. New York: Routledge.
- Bennett, Karen. 2011. “Construction Area (No Hard Hat Required).” *Philosophical Studies* 154 (February): 79–104. <https://doi.org/10.1007/s11098-011-9703-8>.
- . 2017. *Making Things Up*. New York: Oxford University Press.
- Bickle, John. 2020. “Multiple Realizability.” In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020, 1–58. <https://leibniz.stanford.edu/friends/preview/multiple-realizability/>.
- Cangialosi, John P. 2022. “National Hurricane Center Forecast Verification Report.” *National Hurricane Center*. Miami: National Oceanic and Atmospheric Administration. https://www.nhc.noaa.gov/verification/pdfs/Verification_2021.pdf.
- Cargile, James. 2003. “On ‘Alexander’s’ Dictum.” *Topoi* 22 (2): 143–49. <https://doi.org/10.1023/a:1024926205716>.
- Cartwright, Nancy. 1999. *The Dappled World*. Cambridge: Cambridge University Press.

- Cat, Jordi. 2017. "The Unity of Science." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Spring 2023 Edition, 1–71. <https://plato.stanford.edu/entries/scientific-unity>.
- Chakravartty, Anjan. 2017. "Scientific Realism." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2017 Edition. <https://plato.stanford.edu/entries/scientific-realism/>.
- Chalmers, David. 2008. "Strong and Weak Emergence." In *The Re-Emergence of Emergence*, edited by Philip Clayton and Paul Davies. Oxford: Oxford University Press.
- Chen, Jean J., Marguerite Wieckowska, Ernst Meyer, and G. Bruce Pike. 2008. "Cerebral Blood Flow Measurement Using FMRI and PET: A Cross-Validation Study." *International Journal of Biomedical Imaging* 2008 no. 516359 (September): 1–12. <https://doi.org/10.1155/2008/516359>.
- Churchland, Paul M. (1984) 2013. *Matter and Consciousness*. Cambridge: MIT Press.
- Collier, John. 2011. "Holism and Emergence: Dynamical Complexity Defeats Laplace's Demon." *South African Journal of Philosophy* 30 (2): 229–43. <https://doi.org/10.4314/sajpem.v30i2.67786>.
- Crane, Tim. 2001. "The Significance of Emergence." In *Physicalism and Its Discontents*, edited by Carl Gillett and Barry Loewer, 207–24. Cambridge: Cambridge University Press.
- Davidson, Donald. (1980) 2001. *Essays on Actions and Events*. Oxford: Clarendon Press.
- Dunlop, Storm. 2017. *Weather: A Very Short Introduction*. Oxford: Oxford University Press.
- Elisabeth of Bohemia, Princess Palatine, and Descartes, René. 2007. *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*. Edited by Lisa Shapiro. Chicago: The University of Chicago Press.
- English, Lars Q. 2017. *There Is No Theory of Everything: A Physics Perspective on Emergence*. Cham, Switzerland: Springer Verlag.
- Feynman, Richard. (1965) 1990. *The Character of Physical Law*. Cambridge: MIT Press.
- Feynman, Richard P. 1982. "Simulating Physics with Computers." *International Journal of Theoretical Physics* 21 (6/7): 467–88.

- Fine, Kit. 2012. "Guide to Ground." In *Metaphysical Grounding*, edited by Fabrice Correia and Benjamin Schneider, 37–80. New York: Cambridge University Press.
- Fodor, Jerry A. 1974. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28 (2): 97–115. <https://doi.org/10.1007/bf00485230>.
- . 1989. "Making Mind Matter More." *Philosophical Topics* 17 (1): 59–79. <https://doi.org/10.5840/philtopics198917112>.
- Friederich, Simon, and Sach Mukherjee. 2021. "Causation as a High-Level Affair." In *Top-down Causation and Emergence*, edited by Jan Voosholz and Markus Gabriel, 297–304. Cham, Switzerland: Springer Nature.
- (Gentry) Richardson, Katherine E., Daniel P. Roche, Stephen G. Mugel, Nolan D. Lancaster, Kathryn E. Sieving, Todd M. Freeberg, and Jeffrey R. Lucas. 2022. "Social Dynamics of Core Members in Mixed-Species Bird Flocks Change across a Gradient of Foraging Habitat Quality." Edited by Vitor Hugo Rodrigues Paiva. *PLOS ONE* 17 (2): 1–20. <https://doi.org/10.1371/journal.pone.0262385>.
- Hellman, Geoffrey, and Frank Wilson Thompson. 1977. "Physicalist Materialism." *Nous* 11 (4): 309–45. <https://doi.org/10.2307/2214560>.
- Hempel, Carl G. 1969. "Reduction: Ontological and Linguistic Facets." In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, edited by Sidney Morgenbesser, Patrick Suppes, and Morton White, 179–99. New York: St. Martin's Press.
- Horgan, Terence. 1989. "Mental Quausation." *Philosophical Perspectives* 3, 47-76. <https://doi.org/10.2307/2214263>.
- Huxley, Aldous. (1954) 2009. *The Doors of Perception*. New York: Harper Perennial.
- Kidd, D.A., and Mary Wade. 2003. *Collins Latin Dictionary plus Grammar*. Edited by Joyce Littlejohn. New York: HarperCollins Publishers.
- Kim, Jaegwon. 1987. "'Strong' and 'Global' Supervenience Revisited." *Philosophy and Phenomenological Research* 48 (2): 315. <https://doi.org/10.2307/2107631>.
- . 1990. "Supervenience as a Philosophical Concept." *Metaphilosophy* 21, no. 1-2 (January): 1–27. <https://doi.org/10.1111/j.1467-9973.1990.tb00830.x>.
- . 1993a. *Supervenience and Mind: Selected Philosophical Essays*. New York: Cambridge University Press.

- . 1993b. “The Non-Reductivist’s Troubles with Mental Causation.” In *Mental Causation*, edited by John Heil and Alfred Mele, 189–210. New York: Oxford University Press.
- . 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge: MIT Press.
- . 2002a. “Précis of Mind in a Physical World.” *Philosophy and Phenomenological Research* 65, no. 3 (November): 640–43. <https://doi.org/10.1111/j.1933-1592.2002.tb00226.x>.
- . 2002b. “Responses.” *Philosophy and Phenomenological Research* 65, no. 3 (November): 671–80. <https://doi.org/10.1111/j.1933-1592.2002.tb00231.x>.
- . 2005a. “Laws, Causation, and Explanation in the Special Sciences.” *History and Philosophy of the Life Sciences* 27 (3/4): 339–44. <https://www.jstor.org/stable/23333895>.
- . 2005b. *Physicalism, or Something near Enough*. Princeton: Princeton University Press.
- . (1993) 2010. *Essays in the Metaphysics of Mind*. Oxford: Oxford University Press.
- Kroedel, Thomas. 2019. *Mental Causation: A Counterfactual Theory*. Cambridge: Cambridge University Press.
- Kuhn, Thomas S. (1962) 2012. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Laplace, Pierre-Simon. (1825) 1995. *Philosophical Essay on Probabilities*. Translated by Andrew I. Dale. New York: Springer-Verlag.
- Lewis, David K. 1966. “An Argument for the Identity Theory.” *Journal of Philosophy* 63 (1): 17–25. <https://doi.org/10.2307/2024524>.
- Long, Anthony A. 1999. *The Cambridge Companion to Early Greek Philosophy*. New York: Cambridge University Press.
- Lowe, E.J. 2008. “Dualism (Property Dualism, Substance Dualism).” In *Encyclopedia of Neuroscience*, edited by M.D. Binder, N. Hirokawa, and U. Windhorst, 1018–21. Berlin: Springer.
- MacKinnon, Flora I. 1924. “The Meaning of ‘Emergent’ in Lloyd Morgan’s ‘Emergent Evolution.’” *Mind* 33 (131): 311–15.

- Maslin, Mark. 2013. *Climate: A Very Short Introduction*. Oxford: Oxford University Press.
- McGinn, Colin. 1989. "Can We Solve the Mind–Body Problem?" *Mind* 98: 349–66. <https://doi.org/10.1093/mind/xcviii.391.349>.
- McGivern, Patrick, and Alexander Rueger. 2010. "Emergence in Physics." In *Emergence in Science and Philosophy*, edited by Antonella Corradini and Timothy O'Connor, 213–32. New York: Routledge.
- McLaughlin, Brian, and Karen Bennett. 2018. "Supervenience." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021 Edition. <https://plato.stanford.edu/entries/supervenience/>.
- Melnyk, Andrew. 2019. "Physicalism." In *Routledge Encyclopedia of Philosophy*. Taylor & Francis. <https://www.rep.routledge.com/articles/thematic/physicalism/v-1>.
- Montero, Barbara. 2006. "What Does the Conservation of Energy Have to Do with Physicalism?" *Dialectica* 60 (4): 383–96. <https://doi.org/10.1111/j.1746-8361.2006.01073.x>.
- Montero, Barbara Gail. 2013. "Must Physicalism Imply the Supervenience of the Mental on the Physical?" *Journal of Philosophy* 110 (2): 93–110. <https://doi.org/10.5840/jphil2013110240>.
- Moore, Dwayne, and Neil Campbell. 2010. "Functional Reduction and Mental Causation." *Acta Analytica* 25 (4): 435–46. <https://doi.org/10.1007/s12136-010-0107-8>.
- Morgan, C. Lloyd. 1925. "Emergent Evolution." *Mind* 34 (133): 70–74.
- Mulpuru, Siva K., Malini Madhavan, Christopher J. McLeod, Yong-Mei Cha, and Paul A. Friedman. 2017. "Cardiac Pacemakers: Function, Troubleshooting, and Management: Part 1 of a 2-Part Series." *Journal of the American College of Cardiology* 69 (2): 189–210. <http://dx.doi.org/10.1016/j.jacc.2016.10.061>.
- Mumford, Stephen, and Rani Lill Anjum. 2013. *Causation: A Very Short Introduction*. New York: Oxford University Press.
- Neurath, Otto. 1983. *Philosophical Papers, 1913-1946*. Dordrecht: D. Reidel Publishing Company.
- O'Connor, Timothy. 2021. "Emergent Properties." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2021 Edition. <https://plato.stanford.edu/entries/properties-emergent/>.

- O'Shea, Michael. 2005. *The Brain: A Very Short Introduction*.
New York: Oxford University Press.
- Okasha, Samir. 2012. "Emergence, Hierarchy and Top-down Causation in Evolutionary Biology." *Interface Focus* 2: 49–54. <https://doi.org/10.1098/rsfs.2011.0046>.
- Oppenheim, Paul, and Hilary Putnam. 1958. "Unity of Science as a Working Hypothesis." *Minnesota Studies in the Philosophy of Science* 2: 3–36.
- Paull, R. Cranston, and Theodore R. Sider. 1992. "In Defense of Global Supervenience." *Philosophy and Phenomenological Research* 52 (4): 833–54. <https://doi.org/10.2307/2107913>.
- Polkinghorne, John. 2002. *Quantum Theory: A Very Short Introduction*.
New York: Oxford University Press.
- Popper, Karl R., and John C. Eccles. 1977. *The Self and Its Brain*.
New York: Springer International.
- Putnam, Hilary. 1967. "Psychological Predicates." In *Art, Mind, and Religion : Proceedings*, edited by Daniel D. Merrill and William H. Capitan, 37–48.
Pittsburgh: University of Pittsburgh Press.
- Rabin, Gabriel Oak. 2018. "Grounding Orthodoxy and the Layered Conception." In *Reality and Its Structure: Essays in Fundamentality*, edited by Ricki Bliss and Graham Priest. Oxford: Oxford University Press.
- Reid, Robert G.B. 2007. *Biological Emergences: Evolution by Natural Experiment*.
Cambridge: MIT Press.
- Reynolds, Craig W. 1987. "Flocks, Herds and Schools: A Distributed Behavioral Model." *ACM SIGGRAPH Computer Graphics* 21 (4): 25–34. <https://doi.org/10.1145/37402.37406>.
- Schneider, Susan. 2012a. "Why Property Dualists Must Reject Substance Physicalism." *Philosophical Studies* 157 (1): 61–76. <https://doi.org/10.1007/s11098-010-9618-9>.
- . 2012b. "Non-Reductive Physicalism Cannot Appeal to Token Identity." *Philosophy and Phenomenological Research* 85 (3): 719–28. <https://doi.org/10.1111/j.1933-1592.2011.00534.x>.
- Shoemaker, Sydney. 2007. *Physical Realization*.
New York, NY: Oxford University Press.

- Skiles, Alexander. 2014. "Against Grounding Necessitarianism." *Erkenntnis* 80, no. 4 (September): 717–51. <https://doi.org/10.1007/s10670-014-9669-y>.
- Smart, J.J.C. 2007. "The Mind/Brain Identity Theory." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2022 Edition. <https://plato.stanford.edu/entries/mind-identity/>.
- Stoljar, Daniel. 2021. "Physicalism." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021 Edition, 1–75. <https://plato.stanford.edu/archives/sum2021/entries/physicalism/>.
- Toulmin, Stephen. 1970. *Physical Reality*. New York: Harper and Row.
- Vinerbi, Lorenzo, Andrea Bondavalli, and Paolo Lollini. 2010. "Emergence: A New Source of Failures in Complex Systems." In *Proceedings of the Third International Conference on Dependability*. IEEE Computer Society.
- Vintiadis, Elly. 2019. "Property Dualism." In *Introduction to Philosophy: Philosophy of Mind*, edited by Heather Salazar, 29–36. Quebec: Rebus Foundation Publishing. <https://press.rebus.community/intro-to-phil-of-mind/>.
- Vision, Gerald. 2011. *Re-Emergence: Locating Conscious Properties in a Material World*. Cambridge: MIT Press.
- Wilson, Jessica. 2013. "Nonlinearity and Metaphysical Emergence." In *Metaphysics and Science*, edited by Stephen Mumford and Matthew Tugby, 201–29. New York: Oxford University Press.
- Wolfram, Stephen. 2017. "A New Kind of Science: A 15-Year View." *Complex Systems* 26 (3): 197–224. <https://doi.org/10.25088/complexsystems.26.3.197>.
- . 2022. *Twenty Years of a New Kind of Science*. Champaign: Wolfram Media, Inc.
- Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.
- . 2021a. *Causation with a Human Face: Normative Theory and Descriptive Psychology*. New York: Oxford University Press.
- . 2021b. "Downward Causation Defended." In *Top-down Causation and Emergence*, edited by Jan Voosholz and Markus Gabriel, 217–51. Cham, Switzerland: Springer Nature.
- Yablo, Stephen. 1992. "Mental Causation." *The Philosophical Review* 101 (2): 245–80. <https://doi.org/10.2307/2185535>.