

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Demystifying Emergence
A New Exploration Through Scientific Case Studies

Conti Lorenzo, Marta

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Demystifying Emergence: A New Exploration Through Scientific Case Studies

Marta Conti Lorenzo
Department of Philosophy
King's College London

A thesis submitted for the degree of PhD in Philosophy
at King's College London

ABSTRACT

In this thesis I argue that a modified version of Jessica Wilson's subset view, what I call the Holistically Defined Modal view, is the most promising nonreductive physicalist account to accommodate not only mental causation but also other kinds of higher-level causation. In this thesis I do not endeavour to demonstrate nonreductive physicalism's superiority against competitors, such as reductive physicalism or strong emergence. Rather, my goal is to investigate how nonreductive physicalism should be spelled out assessing what I consider the most interesting and successful candidates for the best nonreductive physicalist account. The three candidates are Jessica Wilson's subset view, Sydney Shoemaker's mereological version of the subset view, and Carl Gillett's mutualism.

In Chapter 1, I first provide historical context to the concept of emergence, and I explain its interpretation within this thesis to disambiguate it from alternative definitions or uses. I then introduce Jaegwon Kim's Causal Exclusion Argument, which is one of the most potentially damning and recurrent objections against nonreductive physicalism. In Chapter 2, I present the subset view as well as Wilson's response to Kim's argument. The subset view proposes that for higher-level properties to be causally efficacious in a significant way it is sufficient that they have a distinct set of powers (or causal profile) vis-à-vis their realisers. In this chapter I also defend the subset view against Sara Bernstein's version of Kim's objection.

In Chapter 3 I present Shoemaker's version of the subset view and conclude that Wilson's account is superior to it. In Chapter 4 I introduce Gillett's account, as well as Gillett's Problem of Qualitative Distinctness against the subset view. Gillett claims that science explains the behaviour of special science entities in terms of the joint behaviour of their constituents. From this claim he builds an argument to show that special science properties of composites are realised by the properties of its constituents, which leads him to conclude that properties at different levels do not share any powers —i.e., they are qualitatively distinct. That claim is a direct threat to the subset view, which requires power sharing, and according to Gillett it makes it an untenable view.

To assess Gillett’s objection, I focus on whether realisation is a many-to-one relation – what I refer to as realisation_{M-O} – or a one-to-one relation – realisation_{O-O} . Gillett argues that it must be many-to-one, given that this is the only way to preserve qualitative distinctness between realiser and realisee. Wilson’s way to deal with the apparent many-to-one nature of realisation of special science properties is to provide an ontologically lightweight structural property that bridges the many properties of constituents and the higher-level realised property of the composite. She refers to this kind of properties as “middleman properties”. In the last section of Chapter 4 I argue that Wilson has not provided a successful response to Gillett’s threat, and that the existence of middleman properties is not properly justified. The chapter, therefore, is inconclusive as to whether realisation has a one-to-one or many-to-one structure. Chapter 5 serves as a linking chapter, emphasizing the importance of case studies in elucidating the nature of realisation and discussing their treatment in Wilson’s and Gillett’s works. Chapters 6 and 7 present two scientific case studies and assess whether any of the two kinds of realisation, realisation_{M-O} or realisation_{O-O} , can account for them.

In Chapter 6, I explain how statistical mechanics derives and defines entropy, which is a property of thermodynamic systems from the properties of the particles composing the systems. In Chapter 7, I explain how one of the most influential scientific theories of consciousness, the Global Neuronal Workspace Theory, explains how conscious states arise from brain states.

This thesis concludes that the realisation relation that accounts for both the realisation of entropy and conscious properties is realisation_{O-O} , whereas realisation_{M-O} can, at best, only partially account for them. Hence, realisation_{O-O} is the most general account of realisation. This conclusion is based on one main finding: in both case studies there are scientific grounds to accept the genuine existence of middleman properties. However, the case studies also unveil that Wilson’s definition of “middleman property” –and thus her view– needs to be substantially modified. In Chapters 6 and 7 I present and discuss the modified version of the subset view: the Holistically Defined Modal view –the HDM view. According to this view, a realiser is a holistically defined modal property. I define both the modal aspect and the holistic aspect of the view in a broad sense, that also encompasses cases where a narrow definition applies. The holistic aspect of the view entails that the realiser is defined in terms of properties and states of the system, where those properties and states of the system might or might not be further defined in terms of compositional principles and Boolean operations on its constituents. The modal aspect of the view entails that the realiser has a modal aspect, namely, it considers

possible states the system could be in. In a broad sense, there are cases where the only possible states of the system considered are actual states, which nonetheless are possible states of the system. The Entropy Case Study is an example of a realiser that is a holistically defined modal property in a narrow sense for both aspects, whereas the Neuronal Global Workspace presents an example of a realiser that is holistically defined in a narrow sense and is modal in a broad sense.

ACKNOWLEDGEMENTS

I would like to begin by expressing my deepest gratitude to my supervisor, James Stazicker. Your guidance, expertise, and unwavering support have been instrumental in shaping this thesis. Your insightful feedback and encouragement have been invaluable. I am truly fortunate to have had you as my supervisor throughout this journey. I would also like to extend my sincere thanks to my second supervisor, David Papineau. Your valuable insights and feedback have been crucial in refining my research and enhancing the quality of this thesis. I am grateful to Alex Franklin for supervising the Entropy case study. I truly appreciate your willingness to share your knowledge and give me feedback draft after draft. Your dedication and commitment to my academic development are greatly appreciated. I would like to acknowledge the Philosophy Department at KCL for providing a stimulating and inclusive intellectual environment. Thanks to the London Arts and Humanities Partnership for funding this research, this thesis would not exist without your studentship.

I am indebted to my friends Noelia Iranzo and Nuria Muñoz, with whom I have shared the experience of doing most of our PhD during a pandemic. Noelia, our endless conversations, and your belief in my abilities have been a constant source of motivation. Your intellectual companionship has been invaluable. Nuria, we discovered the topic of emergence together, and our discussions have challenged my thinking and kept me going when the morale was low. Thank you.

I would like to thank my parents and grandparents for their unconditional support throughout my academic journey. Thank you for always believing in me and instilling in me a sense of determination and resilience. You have always been, and always will be, a source of strength and inspiration. Finally, I want to say thank you to my partner, Marina. Your love, understanding, patience, and care have been my rock throughout this journey. I am forever grateful.

TABLE OF CONTENTS

CHAPTER 1. NONREDUCTIVE PHYSICALISM, EMERGENTISM, AND THE CAUSAL EXCLUSION ARGUMENT

1.0 Introduction	8
1.1 British Emergentism	12
1.2 The Causal Exclusion Argument	21
1.2.1 The Causal Closure of the (Micro)Physical	22
1.2.2 The Principle of Non-Overdetermination	25
1.3 Conclusion	28

CHAPTER 2. AVOIDING THE CAUSAL EXCLUSION ARGUMENT: THE SUBSET VIEW

2.0 Introduction	29
2.1 Reformulation of the Causal Exclusion Argument	29
2.2 The Subset View	32
2.3 Scrutinizing Sara Bernstein's Overdetermination Objection against the Subset View	40
2.3.1 Two Solutions to Bernstein's Overdetermination Objection	44

CHAPTER 3. ASSESSING SYDNEY SHOEMAKER'S NONREDUCTIVE PHYSICALIST ACCOUNT

3.0 Introduction	48
3.1 Advantages of Wilson's Schema over Shoemaker's Account	48
3.1.1 Wilson's Neutrality Theses	50
3.1.2 The Advantages of Neutrality	51
3.2 Against Shoemaker's Microphysical Realization (I)	58
3.3 Against Shoemaker's Microphysical Realization (II)	66

CHAPTER 4. CARL GILLETT'S ACCOUNT OF REALISATION AS A COMPOSITIONAL RELATION

4.0 Introduction	69
4.1 Gillett's account	69
4.2 Gillett's Flat Objection against the Subset View	83
4.3 Wilson's Response to Gillett's Objection (and Why It Does Not Work)	85
4.3.1 The Middleman Strategy	87

4.3.2. Middleman Properties: The Controversy	89
4.3.3. Wilson's Positive Argument for the Existence of Middleman Properties (and Where It Goes Wrong)	93
CHAPTER 5. TAKING STOCK AND MOVING FORWARD	103
CHAPTER 6. CASE STUDY I: ENTROPY	
6.0 Introduction	111
6.1 From Large Sets of Particles to Thermodynamic Systems: The Case of Entropy	111
6.1.1 Gibbsian Entropy	114
6.1.2 Boltzmannian Entropy	118
6.2 Thermodynamic Properties Are Causally Efficacious	119
6.3 Testing Wilson's Subset View and Gillett's Mutualism Against the Statistical Mechanics Case Study	120
6.3.1 An Empirically Informed Account of the Realisation Base of Entropy	122
6.3.2 Testing Gillett's Mutualism	124
6.3.2.1 Can Gillett's View Account for the Realisation of Gibbsian Entropy?	124
6.3.2.2 Can Gillett's View Account for the Realisation of Boltzmannian Entropy?	129
6.3.3 Testing Wilson's Subset View	130
6.4 Conclusion	134
CHAPTER 7. CASE STUDY II: THE GLOBAL NEURONAL WORKSPACE	
7.0 Introduction	136
7.1 Bechtel's (2017) View about Mechanisms	137
7.1.1 The Circadian Rhythms Case Study	141
7.2 The Global Neuronal Workspace Case Study	148
7.2.1 The Oscillatory Activity of Neuronal Networks Is a Network Phenomenon	153
7.2.2 The Global Ignition State	159
7.2.3 The Attentional Blink	161
7.3 The GNW Case Study and Realisation: O-O or M-O?	164
7.4 Conclusion	169
CONCLUSION	170
REFERENCES	173

CHAPTER 1. NONREDUCTIVE PHYSICALISM, EMERGENTISM, AND THE CAUSAL EXCLUSION ARGUMENT

1.0 Introduction

The aim of this thesis is to assess how nonreductive physicalism should be spelled out to accommodate not only mental causation but also other kinds of higher-level causation. I believe that in doing so, nonreductive physicalism will be reinforced as a live and tenable position. It is not the goal of this thesis to assess nonreductive physicalism against alternative competitor views. There are many influential nonreductive physicalist accounts, but in this thesis, I focus only on three of them: Jessica Wilson's subset view, Sydney Shoemaker's mereological version of the subset view, and Carl Gillett's mutualism. I believe these views are compelling contenders for the most successful nonreductive physicalist account. Through several arguments, responses to objections, and cases studies, in this thesis I conclude that a modification of Jessica Wilson's subset view is the most promising nonreductive physicalist account.

Jessica Wilson's subset view, or weak emergentism, argues that higher-level properties can emerge from lower-level properties in a way that renders them causally efficacious without requiring any new powers, thus guaranteeing physicalism. A challenge that any nonreductive physicalist view must respond to, and that I discuss in detail later in this chapter, is Kim's Causal Exclusion Argument (CEA). It questions how mental causation (or, in general, any kind of higher-level causation) can have any effect in a world that is already causally closed at the physical level. As I explain in Chapter 2, Wilson's subset view is designed to avoid Kim's CEA. However, some authors consider that the subset view's strategy to avoid Kim's CEA render higher-level properties causally redundant or inert. One of these authors, is Sara Bernstein, who introduces what I call the Overdetermination Objection. This objection motivates the following questions: is weak emergentism a coherent view? How can higher-level properties be causally efficacious without new powers? Would not this view still face Kim's CEA? I argue that Bernstein's Overdetermination Objection can be overcome by appealing to the fact that the objection is actually a version of Kim's Problem.

Another objection that the subset view must respond to is Gillett's Problem of Qualitative

Distinctness, or what I call for short, the Flat Objection. Gillett's objection concludes that the subset view cannot account for the making-up relations posited in the sciences. I argue against this objection by presenting two case studies. In the first one, I explain how statistical mechanics derives and defines entropy, which is a property of thermodynamic systems — liquids, crystals, paramagnets, gases, etc.—, from the properties of the particles composing the systems. This case study is particularly illuminating because we can derive a higher-level property from its lower-level realizer without heavy mathematical artillery. In the second case study, I explain how one of the most influential scientific theories of consciousness, the Global Neuronal Workspace Theory, accounts for how conscious states arise from brain states. I then use the case studies to show that Wilson's subset view is able to account for the making-up relations as well as Gillett's view, if not better.

The Global Neuronal Workspace Theory is especially interesting because the debate over higher-level causation was originally motivated by the discussion over the status of the mental in relation to the physical. Although in its origin the debate was restricted to the metaphysics of mind, nowadays the debate is much broader. I believe that this is mostly because currently there are theories of higher-level causation, such as Wilson's, that do not require higher-level features to be substantially different or independent from the physical realm. Therefore, those theories can also include special science features as plausible candidates for higher-level causation, instead of focusing only on mental features.

This chapter is an introduction to the topic of this thesis, nonreductive physicalism, its origin and context, as well as some of its controversies. First, I briefly introduce the mind-body problem and review the most influential responses to it. I then go on to explain British emergentism, which is considered the origin of many contemporary nonreductive physicalist theories. Finally, I present Jaegwon Kim's Causal Exclusion Argument against British emergentism and nonreductive physicalism in general. Kim's argument is so fundamental and damaging to nonreductive physicalism that any reputable nonreductive physicalist account must show in what ways it can avoid its conclusion.

The mind-body problem is, and has been for centuries, a central topic in philosophy, and in particular in metaphysics and philosophy of mind. The problem can be phrased as a question: how is the mind, if there is such thing, related to the physical world? There are at least three general ways to tackle the question: dualism, idealism, and physicalism. Dualism holds that

there are two kinds of entities or, in Cartesian terms, two substances, mental and physical. However, how they are related remains mysterious or at best problematic. Idealism claims that there is only one realm: the mental. Both dualism and idealism are minority views in contemporary analytic philosophy. Currently, physicalism is the view with most advocates and is regarded as the most promising one. Physicalism claims that there is only one substance or realm: the physical. However, this is only the grounding commitment shared by countless physicalist theories and trends, some of them radically incompatible with each other. Just to show how varied those are, I roughly introduce below (in random order) those that I think are the most common:

1. Reductive physicalism: Everything is in principle reducible to its physical basis.
2. Identity theories: Physical features and mental features are either type identical or token identical. The former entails reductionism, while the latter does not.
3. Physicalist Emergentism/Nonreductive physicalism: Mental features depend on physical features but at the same time mental features are nonreducible to physical features and are causally efficacious.
4. Epiphenomenalism: Mental features depend on physical features but at the same time mental features are nonreducible to physical features. However, they do not play any causal role.
5. Eliminativism: There are not mental features, everything is physical.

In this thesis I will focus on what I call “physicalist emergentism”, which is a term that I use to refer to emergentist accounts that are nonreductive physicalist, such as Wilson’s weak emergentism. “Emergentism” is a broad term that comprises an array of many different theories. In particular, in the literature the term “weak emergentism” is often used to refer to epistemic emergentism. This is because Bedau (1997) first introduced the notion of weak emergence to refer to an epistemic kind of emergence, and in an influential paper Chalmers (2006) also uses the term in this sense. However, Wilson is introducing a new ontological characterisation of weak emergence. Clearly, there is not a coherent classification of those views. So the use of the term “physicalist emergentism” is justified given the lack of a comprehensive way to refer to the intersection between the set of emergentist theories and the set of nonreductive physicalist accounts.

Physicalist emergentism endorses the following claims: mental properties not only exist,

against 5 above; but they are irreducible, against 1 and type identity; and causally efficacious, against 4. In particular, much of the discussion in the literature stems from the tension between the causal efficacy of mental properties and physicalist commitments. This tension is exploited by Jaegwon Kim in his Causal Exclusion Argument, which I introduce in the next paragraphs. Many accept that to be real is to have causal powers, so in order to argue that mental properties exist one must argue that they are causally efficacious (this is exactly what epiphenomenalists deny). However, it is also commonly accepted that physicalism entails the causal closure of the physical world. So there is no room for higher-level novel causes. This is the general problem of higher-level causation, and in particular the problem of mental causation, which is one crucial aspect of the mind-body debate. As we can see in the above list, 4 and 5 simply deny any causal role for mental properties. Only reductive theories (including identity theories) and physicalist emergentism account for higher-level causation, although in radically different ways.

Jaegwon Kim's Causal Exclusion Argument and how to overcome it is a central topic in contemporary philosophy of mind and metaphysics, raising questions about the possibility of mental causation, the relationship between mental and physical properties, and the viability of nonreductive physicalism. The argument's targets, as presented by Kim himself, are nonreductive physicalism (NRP) and emergentism, and its conclusion is that these views are unstable and ultimately collapse into either dualism or eliminativism. However, these two categories encompass a heterogeneous collection of views, where proponents of each view may not share the same commitments. Kim's first effort, then, is to provide a unified characterisation that is able to represent the majority of positions falling under the nonreductive physicalist and emergentist categories. In doing so, Kim claims that emergentism and NRP can be considered one and the same view.

(...) nonreductive physicalists are committed to downward causation just as much as the emergentists and for the same reasons, and their fate, too, is intimately tied to the tenability of downward causation. It will become clear, if I am right, that nonreductive physicalism is a form of Emergentism, and that both positions stand or fall with downward causation. I happen to believe that downward causation indeed proves to be their downfall. (Kim, 1992, p.121)

The fact that mentality has emerged, on the emergentist view, must make a genuinely new causal difference in the world. (...) Mentality must have causal powers not had by any physical-

biological properties, not even by those from which it has emerged. I submit that this is precisely the commitment of the nonreductive physicalists (...) nonreductive physicalism in its currently influential form is, in all essential detail, indistinguishable from the emergentist doctrine about mentality. (Kim 1992: 135)

The argument, then, is intended to pose a problem for views that fail under this unified characterisation. Whether Kim's objection successfully shows that NRP and emergentism are untenable views is of course not a resolved issue, but what Kim has undoubtedly achieved is for his objection to be the main driver of most contemporary nonreductive physicalist developments, as it is widely accepted that any successful nonreductive physicalist account must be able to counter Kim's argument.

The Emergentist view that Kim considers NRP to be indistinguishable from is usually referred to as British Emergentism (BE). Although Kim is very thorough in his introduction of BE, I believe there is a crucial aspect of the view that he dismisses in order for his argument to work. As I will show, Kim's CEA only works if BE is taken to endorse one particular reading of the Causal Closure of the Physical. In showing that, I will also undermine his unification of emergentism and NRP. This is an important step before going into a discussion of NRP, as the topic of emergence is famously known for involving conflicting terms and definitions. Therefore, my discussion on British Emergentism and its separation from contemporary nonreductive physicalist accounts is designed to untangle some of these complex concepts and definitions. This will set the groundwork for the subsequent chapters, as it will allow for a better understanding of nonreductive physicalism. The structure of the chapter will begin with an introduction to BE, followed by a detailed explanation of Kim's argument. Finally, I will investigate whether BE is as susceptible to Kim's objection as he proposes.

1.1 British Emergentism

Mill is considered to be the father of British Emergentism, which flourished in the first quarter of the twentieth century with the works of Samuel Alexander, Lloyd Morgan and, most importantly, C. D. Broad, who developed Mill's view into a comprehensive theory. British Emergentism was a middle path between physicalist reductionism, or what they referred to as Mechanism, and Cartesian dualism (or neo-vitalism). On the one hand, emergentists were

physicalists (or in their words, materialists) because they claimed that everything is entirely constituted by elementary physical entities. At that time the existence of some subatomic particles (protons and electrons) was already known, so emergentists claimed that every object was a composite of such fundamental particles. On the other hand, emergentists are nonreductionists because they claim that when fundamental physical entities interact in complex ways, they bring about novel features. The following quote clearly illustrates the British Emergentists' thought:

“(...) every mental process is also physiological but not all physiological ones are mental. Moreover, just as mental processes belong only to a part of the vital structure, so in life we are dealing with a body which performs processes and exhibits features purely material. (...) It is thus a certain constellation or complex or collocation of physico-chemical processes which behaves vitally, and the presence of such constellations which makes the structure to which they belong an organism. To call it organism is but to mark the fact that its behaviour, its response to stimulation, is, owing to the constellation, of a character different from those which physics and chemistry are ordinarily concerned with, and in this sense something new with an appropriate quality, that of life. (...) Accordingly I am prepared in this sense to believe that they may be right who maintain that biology must be treated as a special science, dealing with its own particular subject of organic life which is distinguished by its own delicate capacity of self-regulation.” (Alexander, 1920, p. 61-63)

This fragment is very representative because it involves the BE's three main commitments. Before introducing those commitments, though, I need to bridge the gap between BE's terminology and the terminology used in contemporary philosophy. Taking this text as an example of the terms used by BE, we can see the use of terms like “behaviour”, “quality”, “capacity”. It is said that the behaviour exhibited by the “constellation” of physico-chemical entities is new in relation to the behaviour of its constituents separately. This is explained by the organism having a new capacity, that of self-regulation, and this new capacity in turn encompasses a new quality, that of life. When introducing BE's commitments, Kim translates those terms into contemporary philosophy terms, and I will follow his terminology here. The terms “quality” and “capacity” are substituted by “properties” and “powers”, respectively. This is just meant as a terminological change, so that there is a continuity between discussion of BE's views and of NRP's views. Someone might argue that in doing so Kim is predisposing BE and NRP's unification, and in following him in that, I might be doing the same. However, the aspects on which the views differ are not directly related to any of those two terms, so I do

not think there is anything hinging on this, and on the contrary, this should be understood as a mere translation into contemporary philosophy jargon. Moreover, late emergentists like Roger Sperry, who is also inspired by British Emergentists, do talk in terms of “properties”.

Now, moving on to the three main commitments of BE present in the text. We can find the first one at the beginning of the text: “every mental process is also physiological”. Given that basic entities and properties are physical, there is a sense in which every mental process is also physiological. This is a commitment to physical monism:

- 1) Physicalist ontological monism: The basic entities and properties are fundamental nonemergent physical entities and their fundamental nonemergent physical properties.

Second, the text argues that the constellation of basic entities can show a distinct behaviour when its parts are related and propertied in a certain way. It is because the constellation shows this new behaviour, that we attribute to it a new quality, that is, a new property. This is what Kim (1992, p. 123) refers to as the doctrine of Property Emergence, and it is directly related with the emergence of new powers:

- 2) Property Emergence: (a) Genuine new properties emerge when basic entities aggregate in certain ways so that they form a complex structure. (b) Those emergent properties bestow the complex structures they characterise with new powers.

The second part of the doctrine of Property Emergence, (b), is not that clearly present in the quote above. However, it is a fundamental commitment of BE. This is shown in the following passage of the same book by Alexander:

[epiphenomenalism] 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, as undoubtedly would in time, be abolished. (Alexander, 1920, vol. 2, p. 8)

This is known as Alexander’s dictum or the Eleatic Principle, that can be translated in terms of causal powers, as Kim (1992, p.134) does: to be real is to have causal powers. Although I discuss the definition of causal powers in detail in Chapter 2, I need to briefly mention here what is meant by “causal powers” in this context. Throughout this thesis I take Wilson’s neutral

definition of causal powers as the causal contribution a property can make to the bringing about of an effect (Wilson, 2015, p. 354). Alexander's dictum is the reason why BE needs to accept that higher-level properties have distinct causal powers, because if they did not then they would need to be eliminated.

Finally, the third doctrine is about the relation between emergent properties and the base entities they arise from.

- 3) The irreducibility of emergents: Emergent properties cannot be reduced to the base properties they arise from.

In most of the British Emergentists' works, irreducible emergent properties determine the existence of new levels of reality. The world is taken to be layered into levels according to a hierarchy of complexity and abstractness. Given that emergent properties bestow new powers upon the composites that have them, levels can be used to distinguish and hierarchise the different sciences and their laws. I will say more about special science laws below, in here I focus on trans-ordinal laws—which for my purposes I consider to be akin to Nagel's bridge laws, as they bridge the gap between events at different levels (see Kim, 1992, p. 125). Heteropathic or trans-ordinal laws are a key element of British Emergentism. These laws are contingent and a posteriori—precisely for this reason, a Laplacian calculator could not deduce the emergent features arising from a complex configuration—and describe how emergent properties manifest when lower-level components interact in specific configurations. However, for BE the fact that these laws exist does not entail in any way that emergent properties can be reduced to, or explained solely by, the properties and behaviours of their constituent parts.

The higher quality emerges from the lower level of existence and has its roots therein, but it emerges therefrom, and it does not belong to that lower level, but constitutes its possessor a new order of existent with its special laws of behaviour. (...) A [Laplacian] calculator given the state of the universe (...) could not on our interpretation predict what qualities would be evoked by the complexes he predicts in Space-Time, unless he lived to observe them. (Alexander, 1920, p. 73)

In Nagel's (1961) account of reduction, the connection between higher-level theories and lower-level ones established by bridge laws is sufficient for reduction. Bridge laws establish

relationships between the entities, properties, or processes described by two different theories, allowing the higher-level theory to be derived from, hence reduced to, the lower-level theory. However, BE's reject this notion of reduction. For them, reduction requires the ability to *explain how* emergents arise from base level collections of entities. The fact that if appropriate trans-ordinal laws are added to the statements of a lower-level theory we can then derive the statements of the higher-level theory does not provide the sort of explanation BE require for reduction. Reduction requires an explanation for trans-ordinal laws. For BE, the lack of such an explanation highlights the unique character of emergent phenomena, which cannot be fully predicted or explained by lower-level theories alone. Even if those bridge laws exist, which BE accepted they do, reduction as understood by BE is not achieved. This leads them to the claim that emergents must be accepted with "natural piety": that life "is a new quality" that emerges from physico-chemical processes and is governed by special laws must "be accepted with the 'natural piety' of the investigator. It admits no explanation" (Alexander, 1920: vol.2, 46-47). "Consider and bow the head.' That, in some sense, should be our loyal attitude towards all emergents." (Lloyd Morgan, 1923, p. 4). As I will explain in Chapter 4, there are many similarities between Gillett's Emergentist view and BE. But there is one key difference: Gillett's view attempts to explain, with varying degrees of success to be assessed, what British Emergentists considered inexplicable—and importantly the fact that such an explanation exists does not entail reduction.

Having emphasised the importance of irreducibility of emergents for BE, it is important to remind ourselves of doctrine 1, Physicalist ontological monism. British Emergentist's views remain committed to naturalism. In the following quote Alexander explicitly denies entelechies (i.e. non-physical entities) as the right way to describe the mind. Moreover, we can clearly read that, although entities at different levels of material complexity might behave differently, they are nonetheless chunks of matter.

Instead of straightway postulating an entelechy to act as a guide, it would seem to me more reasonable to note that a given stage of material complexity is characterised by such and such special features, and that these are part and parcel of the nature of the principle or plan of the new order of complex. It is quite true that no merely material complex will regenerate itself or reproduce itself or grow up into a small perfect specimen from half the stuff of a full-sized one. But the fact is that the new complex is no longer purely material, though it is also material. By accepting this we at any rate confine ourselves to noting the facts, observing loyally the

differences of these existents from existents of a lower order; and do not invent entities for which there seems to be no other justification than that something is done in life which is not done in matter. (Alexander, 1920, p. 65)

Having explained the fundamental tenets of British Emergentism, I will now explore their arguments in greater detail. As mentioned, the concept of emergence has its roots in John Stuart Mill's work. In particular, in the *System of Logic* (1843), where he introduces the distinction between homopathic and heteropathic laws and effects. To comprehend this distinction and its significance, it is essential to explain the principle of Composition of Causes, introduced by Mill himself. This principle is analogous to the principle of Composition of Forces in mechanics and states that "the joint effect of several causes is identical with the sum of their separate effects" (Mill, 1859, p. 428). Homopathic effects adhere to this principle, while heteropathic effects arise from violations of the Composition of Causes principle. According to Mill, chemical processes often provide examples where the joint effect of several causes does not equal the sum of their separate effects, such as the combustion of methane. Methane combustion is an exothermic chemical reaction between methane (CH_4) and oxygen (O_2) to produce carbon dioxide (CO_2) and water (H_2O). This reaction plays a crucial role in Mill's argument, as it seems to be a process that is not aggregative where there is a presence of heteropathic effects: where there was methane (CH_4) and oxygen (O_2), there is now carbon dioxide (CO_2) and water (H_2O); the chemical knowledge at the time could not show how the effects could be obtained by merely aggregating the causes. In this case, given that the collection is able to cause effects that were unexplainable at the time by the properties of the individual elements, it was concluded that the aggregation of methane (CH_4) and oxygen (O_2) must have emergent properties that endow it with the powers to cause those heteropathic effects. The existence of violations of the Composition of Causes principle is what justified doctrine 2, Property Emergence.

Let me move on to discuss how British Emergentism argues for the existence of special science laws. The existence of special sciences and laws relies on the violation of the principle of Composition of Causes. At the same time, this principle can be in place again when some of these heteropathic effects combine. In such cases, they produce homopathic effects at their own level, and new homopathic laws appear that can be derived from the effects of separate agents. The existence of special sciences thus relies on both the violation and the compliance with the principle of Composition of Causes because it allows us to derive, from an initial set of basic

special science laws, all the laws of a given special science.

Although Mill does not yet use emergentist terminology, Lloyd Morgan, in his *Emergent Evolution* (1923, p. 2), suggests that heteropathic effects and heteropathic laws can be considered fewer appealing synonyms for the now common terms “emergents” and “emergent laws,” respectively. G. H. Lewes (1875) was the first to introduce the term “emergence,” referring to the distinction between “emergents” and “resultants.” This distinction echoes that between heteropathic effects and homopathic effects of a certain configuration. Resultants are the result of the sum of the separate effects of the agents constituting the configuration, while emergents are effects that are more or at least different than the sum of the separate actions of their parts.

In a very famous paper that helped to bring emergentism back to the fore of philosophical discussion, Brian P. McLaughlin (1992) reviews the weaknesses and strengths of British Emergentism, as well as the historical context in which the view was conceived. An important argument of the paper, and the one that is most relevant to my purposes, defends the compatibility of BE with our best contemporary physical theories: Classical Mechanics and Quantum Mechanics. In the following lines I delve onto this argument.

First, a terminological note: in order to assess whether Classical Mechanics and Quantum Mechanics are compatible with the BE views, we must decide what those views imply for those physical theories. I have explained that BE claim that emergent properties bestow new powers to the collections that instantiate them. In Classical Mechanical terms, this implies that:

(...) Types of structures that compose certain special science kinds can affect the acceleration of a particle in ways unanticipated by the laws concerning forces exerted by pairs of particles, general laws of motion, and the spatial or spatio-temporal arrangements of particles. In a framework of forces, the view implies that there are what we may call “configurational forces”: fundamental forces that can be exerted only by certain types of configurations of particles. (McLaughlin, 1992, p. 21).

Heteropathic forces, initially proposed by Alexander Bain (1870) and referred to as configurational forces by McLaughlin, emerge when specific arrangements of particles produce heteropathic effects. In other words, these configurations have properties that grant

them forces not available to their individual constituents. These forces are fundamental, and McLaughlin contends that they are not only compatible with Classical Mechanics (McLaughlin, 1992, p. 36) but also with Quantum Mechanics (McLaughlin, 1992, p. 23).

Regarding Classical Mechanics, one simply needs to incorporate heteropathic forces into the collection of forces that could potentially influence a body's motion. In Classical Mechanics, forces combine through vector addition, resulting in a total vector sum of the forces acting upon a body. If such a fundamental configurational force were to exist, it would be included during vector addition, becoming part of the overall vector sum. In the Hamiltonian formulation of Classical Mechanics, where energies are used instead of forces to characterise the time evolution of a system, the introduction of a new force does not present a problem either. The energy associated with this force could as well be added to the Hamiltonian characterising a given state. In this scenario, the Hamiltonian would account not only for kinetic and potential energy of the individual particles, but also for energies specific to particular particle configurations.

Regarding Quantum Mechanics, the argument is very similar, but in this case I need to introduce Schrödinger's equation, as the Hamiltonian alone is not sufficient to characterise the time evolution of a system. Schrödinger's equation is a fundamental equation that describes the time evolution of a quantum system's wavefunction. The wavefunction is a mathematical representation of the quantum state of a system. The equation, formulated by Austrian physicist Erwin Schrödinger in 1926, describes the behaviour of a quantum system based on its present state. It is usually presented using the Hamiltonian formulation, as it is much simpler to work with energies than forces, although a formulation using forces is also possible. The general form of Schrödinger's equation is the time-dependent Schrödinger equation, where $\hat{H}$ is the Hamiltonian operator, $\Psi(x, t)$ represents the wavefunction of the system, $\hbar$ is the Planck's constant, and t is time:

$$\hat{H}\Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}$$

The Hamiltonian represents the total energy of a system and typically includes various forms of energy, such as kinetic energy, potential energy, and interaction energies between particles (e.g., electromagnetic interactions). McLaughlin's argument is that if a particular configuration of particles gives rise to a unique form of energy that affects the system's behaviour, this energy

could be incorporated into the Hamiltonian. The modified Hamiltonian would then be used in Schrödinger's equation to account for the influence of this energy on the system's wavefunction, allowing for the prediction of the system's time evolution and properties.

Given these theoretical considerations, it turns out that the validity of British Emergentism mostly depends on empirical evidence, rather than on any contradictions or theoretical shortcomings in the modifications it suggests should be made to our most advanced scientific theories. This perspective already points to what McLaughlin considers was one of the main reasons for the decline of British Emergentism: with the progress of quantum mechanics, researchers did not find empirical evidence to support the existence of heteropathic forces. In fact, evidence indicated the opposite. Chemical bonds, once believed to be one of the most promising domains for demonstrating emergentism, were explained by quantum mechanics without the need to invoke heteropathic effects. The explanation of chemical bonds relies on the principles of quantum mechanics, such as wave functions, electron orbitals, and molecular geometry. These concepts help elucidate the nature of the bonds between atoms, as well as the properties arising from complex molecular structures, solely using quantum mechanics and non-configurational forces.

After providing a summary of the main tenets of British Emergentism, I would like to draw attention to two key ideas, both of which are related to the earlier discussion of the compatibility of emergentism with classical mechanics and quantum mechanics. The first idea concerns the reasons behind the decline of British Emergentism. It is important to note that the undermining of this philosophical perspective was not due to any intrinsic tension or contradiction within the view itself. Rather, it was the result of empirical evidence: quantum mechanics successfully explained many of the phenomena that emergentists had previously cited as striking examples of emergence, such as chemical bonds (McLaughlin, 1992, p. 49). It is crucial to emphasise, however, that this development does not necessarily undermine the overall coherence and validity of British Emergentism as a philosophical framework. That being said, it is perhaps one of the reasons why Kim's argument was so influential when it was initially introduced. The objection demonstrated that there is, in fact, a tension within British Emergentism. By bringing attention to the potential issues in the view, Kim's argument provided an opportunity for further examination, analysis, and development of the emergentist framework. This exploration allowed for a better understanding of the underlying concepts and their implications, ultimately contributing to new contemporary emergentist approaches, such

as Wilson's and Gillett's which are at the centre of this thesis.

The second idea relates to the stance of British Emergentists on physicalism. They regarded themselves as physicalists, which means that, despite their claim that emergent properties and configurational forces exist, they also defended Physicalist ontological monism. As I have shown, their definition of physicalism allows for the existence of fundamental configurational forces, or higher-order causes, which are not reducible to lower-level physical processes. According to British Emergentists, the presence of such forces is a brute empirical fact that should be accepted. However, one of the key contemporary principles of physicalism is the Causal Closure of the Physical (also known as the Completeness of the Physical World) — roughly, the idea that every physical effect has a sufficient physical cause. How could this principle accommodate such forces and features? According to Kim, acceptance of this principle is at the heart of the tension within BE and NRP views. With this context in mind, I can now proceed to introduce Kim's Causal Exclusion Argument.

1.2 The Causal Exclusion Argument

The Causal Exclusion Argument (CEA) can be found in several of Kim's papers, such as Kim (1992, 1993a, 1993b, 1999, 2006), and its primary objective is to demonstrate that emergentism and nonreductive physicalism embrace claims that, when combined, result in a contradiction. It is important to note that Kim has proposed various formulations of the problem and made some adjustments over time, so I will present the version that I believe is most suitable for my purposes. For the sake of simplicity, I will introduce the problem in terms of mental properties, however it can be generalised to any higher-level emergent property.¹ I will start by introducing the five claims Kim's considers are definitory of emergentism and nonreductive physicalism. The first two are claims that according to Kim no physicalist would dispute, while the

¹ For instance, we can find Kim's account of Special Sciences in his 2005's paper *Laws, Causation, and Explanation in the Special Sciences*. In p. 331 Kim argues for the following principle: "Massive failure of causal closure for all special sciences fragility of special-science causation. For any putative cause from S1 to S2, involving special science kinds S1 and S2, the lower-level condition C such that if C were to occur along this would disrupt the causal relation and prevent S2. The idea is that any special-science causal process is liable for disruption from below - that is, by occurrences, or non-occurrences of events at the lower-levels. Computer programs malfunction account of hardware breakdowns; normal biological causal can be disrupted by interfering conditions at the physicochemical level. And so on. I believe this generalized version of massive failure is not at all implausible."

subsequent three represent the core tenets of emergentism and nonreductive physicalism.

- (1) *Physical Causal Closure*. For every physical effect there is a sufficient physical cause
- (2) the *Non-overdetermination Principle*. Physical effects are not systematically overdetermined by two or more sufficient causes (excluding double-rock-throw cases and causal chains)
- (3) *Dependence*. Mental properties depend on their base properties.
- (4) *Efficacy*. Mental properties are causally autonomous from their base properties.
- (5) *Distinctness*. Mental properties are distinct from their base properties.

Before discussing how these claims together might give rise to a contradiction, some considerations about them must be in place. Most philosophers would argue that (1) and (2) are both claims that any physicalist should endorse.

1.2.1 The Causal Closure of the (Micro)Physical

Starting with the first claim, Kim himself argues for its necessity in the following quote:

(...) to give up this principle [Physical Causal Closure] is to acknowledge that there can in principle be no complete physical theory of physical phenomena, that theoretical physics, insofar as it aspires to be a complete theory, must cease to be pure physics and invoke irreducibly nonphysical causal powers - vital principles, entelechies, psychic energies, élan vital, or whatnot. If that is what you are willing to embrace, why call yourself a "physicalist"? (Kim, 1993a, p. 356)

This seems a compelling argument: denying Physical Causal Closure implies accepting the existence of “irreducibly nonphysical causal powers - vital principles, entelechies, psychic energies, élan vital, or whatnot”. However, there is a certain ambiguity in how the causal closure is defined. To start with, we can read in the above passage that all the examples of “irreducibly nonphysical causal powers” are some kind or other of mental or vital entelechies. This means that Kim is either considering 1) special science’s features are non-problematically reducible to fundamental physics or 2) taking special science’s causal powers as physical and only problematising mental causation. However, the former approach seems question begging to me: Kim’s argument precisely aims to show that emergent properties (not only of the mental

kind, but also of the special sciences kind) must be reducible, so in assuming that they are he is implicitly using as a premise the claim that he wants to prove. Although one might argue that special science properties are not of the kind usually considered to engender problematic emergence, and for this reason Kim left them out of the discussion, in the current debate that cannot be assumed. On the other hand, the latter approach, where he would be considering special science's causal powers as physical, entails that the physical causal closure does include broadly physical properties, namely special science properties. This approach entails redefining the notion of "physical" to encompass a wider range of properties that are grounded in or supervene upon the basic physical entities. By allowing for the existence of "broadly physical" properties, the principle of physical causal closure could be preserved even if emergent properties exist. However, the formulation in Kim's framework would be problematic, as it opposes to the possibility of a complete theory of fundamental physics, which is what Kim considers is at stake in this debate. When talking of a unified theory of physics philosophers and physicists alike are referring to a unified theory of *fundamental* physics, which does not include special science features.

I believe that the crux of this discussion revolves around the challenge of defining what qualifies as physical. As I explained above, a key distinction for British Emergentists is that between homopathic effects, or resultants, and heteropathic effects, or emergents. Homopathic effects are those produced when powerful features follow the Principle of the Composition of Causes, while heteropathic effects are those produced when powerful features violate such Principle. According to British Emergentists, the latter entailed the existence of configurational forces, however in their view this is not against physicalism. The view that McLaughlin (1992) attributes to British Emergentists is that those configurational forces are new forces that higher-level features contribute. Although this reading of British Emergentism is charitable because it allows McLaughlin to provide an idealised and coherent account of British Emergentism, Gillett (2006, p. 262) claims that in doing so McLaughlin ends up attributing them an anti-physicalist commitment: that of the existence of non-physical forces. However, McLaughlin explicitly defines configurational forces as fundamental physical forces. Moreover, I have explained before that McLaughlin convincingly argues that configurational forces are compatible with classical mechanics and quantum mechanics.

If British Emergentism's view of the laws of the special sciences were correct, then some fundamental laws of physics would concern structural properties specific to very specific kinds

of things, for example, various chemical kinds, biological kinds, and so on. And there would be moreover downward causation from the psychological, to the biological, to the chemical, to the subchemical (...) for at least some kinds from each special science will be composed of types of aggregates of particles that exert configurational forces” (McLaughlin, 1992, p. 22)

Many would agree, with Kim, that physicalism is mainly defined through the Completeness of Physics, or Physical Causal Closure. So the problem of what counts as physical can be reframed as the problem of how the Completeness of Physics should be defined. Moreover, as I shall explain, Physical Causal Closure is a crucial claim in Kim’s Exclusion Argument. He defines the principle as follows: For every physical effect there is a sufficient physical cause. However, this way of defining Physical Causal Closure is vague, as we can provide two more fine-grained definitions: either the microphysical domain is causally closed, or the broadly physical domain is. Honouring Mill’s distinction I call those principles Homopathic Causal Closure and Heteropathic Causal Closure.

- (1) Homopathic Causal Closure: For every microphysical effect there is a sufficient microphysical cause.
- (2) Heteropathic Causal Closure: For every broadly physical effect there is a sufficient broadly physical cause.

Microphysical features correspond to fundamental physics features, and broadly physical features correspond to fundamental physics features plus any other feature in a certain appropriate relation to those fundamental features. For instance, Yates (2016) argues that physicalists do not take Physical Causal Closure to entail the completeness of the microphysical, rather he argues that the preeminent way of understanding Physical Causal Closure is the second reading, what I call Heteropathic Causal Closure. This line of reasoning can be found also in Crook and Gillett (2001).

I have briefly explored a potential disambiguation of the Causal Closure principle, but it is important to note that this discussion is far from exhaustive. There is an extensive body of literature dedicated solely to this topic, such as Papineau’s (2008) “Must a Physicalist be a Microphysicalist?” and Ney’s (2016) “Microphysical Causation and the Case for Physicalism.”

1.2.2 The Principle of Non-Overdetermination

In relation to the principle of non-overdetermination, it is essential to note that it is designed to apply exclusively to systematic overdetermination between independent events. This distinction ensures that cases of occasional overdetermination, such as double-rock throw scenarios, are not considered problematic. In a double-rock throw case, two entities coincidentally cause the same effect (for example, two rock throwers breaking a window) even though only one throw would have been sufficient to break the window. Although these situations represent instances of overdetermination, as there are two distinct causes sufficient to produce the effect, they are not regarded as problematic since they are not systematic occurrences. The principle of non-overdetermination also encompasses an ambiguity that has been exploited by nonreductive physicalists and will play a significant role in later discussions. This ambiguity arises from the fact that overdetermination can be considered at the level of causes, which involves the properties themselves, or at the level of causal powers, which does not involve the properties themselves but rather their possession of certain powers. Kim's argument is framed in terms of the former interpretation. To provide a more detailed and nuanced understanding of the principle of non-overdetermination, it is essential to consider both the distinction between systematic and occasional overdetermination and the ambiguity between considering overdetermination at the level of causes or causal powers. By carefully examining these aspects, we can better comprehend the implications of the principle and how it relates to the broader debate surrounding emergentism and nonreductive physicalism. This discussion will be continued in Chapter 2, where I will assess Wilson's proposed solution to the Exclusion Argument.

Before introducing the argument, I need to provide one last comment. As the reader surely has noticed, talk of levels is predominant in nonreductive physicalist and emergentist literature, and is also pervasive in this thesis. However, how levels are individuated, or even whether levels exist, is a discussion I will not pursue here. In this thesis, unless stated otherwise, I follow Wilson (2021, p. 35) in considering levels to be individuated in relation to natural laws, and thus to sciences. So every feature associated to science *S* will belong to a certain level *L*, and all the entities in level *L* will be composed by entities at level *L*-1. On the other hand, all properties at level *L* will be realized by properties at level *L*-1. Wilson also includes at any given level, say *L*, complex entities made up of entities at that level, as long as their features and relations still belong at level *L* (more on this point on Chapters 3 and 4). In the next chapter

I provide a thorough explanation of Wilson's argument for the individuation of levels in relation to special science laws, which is based on the elimination of degrees of freedom.

Finally, let us move on to the argument. According to (4), mental properties are causally autonomous. This means, according to Kim, that they have new powers to produce effects either at their own level, or in levels below. The former is known as same-level causation, and the latter as downward causation. Kim claims that same-level causation (except in the fundamental level, if there is one) requires downward causation. This is due to (3), *Dependence*. Given that mental properties depend on their base properties, by causing a change at their own level, mental properties also cause a change in the lower-level, so in a same-level case of causation there is also downward causation (Kim, 1999, p. 23). For the sake of conciseness, I shall explain in detail the argument only regarding downward causation.

Suppose that M is a mental property at level L that causes property P* at the lower level L-1. According to Physical Causal Closure (1), P* must have a physical cause. Let's assume that M's physical base is P, and P is a sufficient cause for P* (it is important to note that if P* were caused by another unrelated physical property, the argument would remain unaffected). Since P and M both cause P* and, given that they are (4) causally autonomous and (5) distinct, P* is overdetermined. Moreover, as (3) posits that M depends on P, this scenario is not a double-rock-throw case, and consequently, Non-overdetermination is denied in every instance of downward causation. In other words, every case of downward causation is overdetermined. In order to avoid the violation of *Non-overdetermination* one could try to reject one of the causes. However, rejecting the physical cause P would contradict *Physical Causal Closure*, which is Physicalism's key principle. As both nonreductive physicalism and emergentism are physicalist theories, it is the higher-level causes that must be rejected. This is why Kim refers to his argument as the causal exclusion argument since its conclusion is that lower-level causes exclude higher-level causes (Kim, 2006, p. 558).

Kim has employed this argument to contend that nonreductive physicalism and emergentism are unstable positions because they either collapse into dualism or eliminativism: "nonreductive materialism is not a stable position. There are pressures of various sorts that push it either in the direction of outright eliminativism or in the direction of an explicit form of dualism" (Kim, 1993b, p. 47). This suggests that Kim only considers two solutions to his argument: either one denies (1) *PCC* and becomes a dualist, or one denies (4) *Efficacy* and

becomes an eliminativist (or epiphenomenalist).

I briefly argue now that there is a potentially straightforward response to CEA available to British Emergentists. Remember that for the objection to fail we need either to deny PCC and argue that this does not entail anti-physicalism, or accept PCC and then argue that in some way Kim's conclusion is still avoided. Clearly, as I have shown British Emergentists deny PCC: they accept the existence of configurational fundamental forces, therefore higher-level properties have new powers not had by their lower-level constituents, and they can produce new effects at the lower-level. Hence, if the truth of physicalism depended on accepting PCC, Kim's Exclusion Argument would succeed against British Emergentism. Kim was certain that any physicalist emergentist account would claim that higher-level features are causally efficacious, so they would need to deny PCC. However, why should physicalism a formulation of PCC as Homopathic Causal Closure? The British Emergentist would argue that Heteropathic Causal Closure is a more accurate representation of physicalism than Homopathic Causal Closure. By rejecting the Homopathic formulation of the CCP, British Emergentism can maintain that higher-level features are causally efficacious without violating the principles of physicalism. This move would allow the view to avoid the consequences of Kim's Exclusion Argument without resorting to dualism or eliminativism. But of course, one can still press the point, as Kim would do, and argue that Physical Causal Closure must be understood as Homopathic Causal Closure.

1.3 Conclusion

In this chapter I have presented British Emergentism in order to provide historical and philosophical context to the chapters ahead. I have explained Kim's CEA, which has such a strength that it has forced any nonreductive physicalist account to provide a solution to it. It is very difficult to find a NRP view that at some point does not discuss how it avoids or overcomes Kim's CEA. In the literature one can find two predominant ways to overcome the objection:

- Some strategies argue that the Physical Causal Closure can be denied while avoiding the threat of dualism. This is the case of Yablo's (1992) difference-making approach, and List & Menzies counterfactual approach. In both views, in cases of higher-level causation, only higher-level causes are taken to be relevant to the production of the effect, as it is the higher-level cause the one that makes a difference to the bringing

about of the effect (or in Yablo's terms, it is proportional to the effect). Although they do not deny that lower-level properties are sufficient to cause the effect, in those instances they are not the actual cause. This takes Menzies to reformulate the CCP: "Causal closure of the physical reformulated: For every physical property G, there is a physical property F that is a difference-making cause of G." And then argue that the CEA is avoided because this CCP is false (a nuanced discussion of these views can be found in Menzies' chapter in *Being reduced: new essays on reduction, explanation, and causation*, 2008).

- Some other strategies revolve around questioning whether the entailed overdetermination is genuinely problematic. In that case, it is accepted that properties at both levels are causally efficacious. For instance, by offering an account of the dependence relation that, while securing causal autonomy, only entails a benign form of overdetermination. This is the case of the subset view.

In this thesis I focus on the subset view, which provides the second kind of solutions to Kim's CEA, although as it will be apparent in subsequent chapters, both solutions share some commitments. A central reason why I focus on the subset view is because it aims to preserve the PCC as formulated by Kim, and yet it can overcome the CEA without denying that lower-level properties are causally efficacious. Kim's formulation of the PCC is arguably the one accepted by reductive physicalists, so trying to argue for an alternative PCC doubles the burden of the NRPist, who not only needs to convince reductive physicalists of the existence of powerful higher-level properties, but also of a reformulation of the PCC.

CHAPTER 2. AVOIDING THE CAUSAL EXCLUSION ARGUMENT: THE SUBSET VIEW

2.0 Introduction

In this chapter I introduce and assess Jessica Wilson's Nonreductive Physicalist view of higher-level properties. Wilson has introduced her account in several papers (Wilson, 1999, 2011, 2015) and in her book *Metaphysical Emergence* (2021). Throughout her work she has presented two ways to solve Kim's problem. On the one hand, she introduces the Strong Emergentist solution, which is anti-physicalist because it entails that higher-level features have new powers. On the other hand, she introduces the Weak Emergentist solution, which is a nonreductive physicalist solution that requires higher-level features to have fewer powers than their lower-level physical realizers. Both solutions are, in her own words, "schemas". This is because these views only impose general conditions on causal powers, but aim to remain neutral on all other concerns (such as which dependence relation must hold between realiser and realisee, what powers are, what theory of causation is best, etc.). In Wilson's own words, "these schemas encode the *only* options for characterizing the metaphysical emergence of higher-level, broadly scientific entities" (Wilson, 2015, p. 363). This is a strong claim, as it means that any theory of higher-level features that successfully deals with Kim's problem must implement one of these schemas and meet its conditions. Nevertheless, advocating for Wilson's schemas does not necessitate the acceptance of such a claim. In this thesis, I will neither defend nor endorse the claim that these schemas are the only two ways to characterise metaphysical emergence.

2.1 Reformulation of the Causal Exclusion Argument

Before explaining Wilson's solutions, I will introduce a new formulation of the Causal Exclusion Argument. Wilson's solutions involve a powers-based approach: rather than talking in terms of causes she presents her arguments in terms of causal powers. However, Kim's exclusion argument, as presented by Kim and as introduced in Chapter 1, is not fully stated in terms of causal powers. As I will show in 2.1.1, this introduces an ambiguity to the argument that can only be overcome by reformulating it in terms of causal powers. But before discussing any of that, what are causal powers? In Wilson's words:

Talk of ‘powers’ (...) is simply shorthand for talk of what causal contributions possession of a given property makes (or can make, relative to the same laws of nature) to an entity's bringing about an effect, when in certain circumstances (where the circumstances alone are not up to the task of bringing about that effect). (Wilson, 2015, p. 354 & 2021, p. 45)

This is Wilson’s neutral understanding of powers. To put it in an example, the power the property of being sharp has to cut bread when its bearer is a knife just is the causal contribution the property of being sharp makes to the knife producing that effect (when this property is co-instantiated with other enabling properties such as the property of being made of a hard material such as steel). In the literature there is discussion of whether properties themselves have powers, or only their bearers (Baysan, 2016, p. 386), but in Wilson’s neutral understanding of powers this does not make a difference to her account. One can understand properties to bestow powers on their bearers, to have powers, or to have their powers derivatively on the powers of their bearers. I cannot provide a detailed discussion of this issue in this thesis, so following Wilson and most of the literature, I talk of properties *having* powers. According to Wilson, and as I explore in more detail in Chapter 3, the subset view is devised such that one should be able to accommodate their preferred theory of the nature of powers as well as their preferred relation between powers and properties.

Wilson’s metaphysically neutral stance on the nature of powers requires that her schemas work for any view on causal powers. In order to demonstrate this, she contends that her model effectively works for what are arguably the two most divergent and extreme perspectives on powers: the view that powers are contingent regularities and the view that powers are genuine and real metaphysical categories. In this thesis I remain neutral on whether Wilson can deal with all and every theory of powers and causation, as this is a strong thesis that would need thorough discussion. However, I think how she deals with the contingentist categoricallist Humean in this fragment is persuasive:

Suppose a contingentist categoricallist Humean wants to take a physicalist approach to the problem of higher-level causation, and so aims [this will become clearer below] to identify every token power of a token higher-level feature with a token power of its lower-level base feature. As previously discussed, such a Humean understands powers in terms of actual or potential instances of a (contingent) regularity. Where the aim is to avoid overdetermination,

the Humean may suppose, to start, that the (relevant instances of the) regularities overlap, both with respect to the (single) effect, and with respect to the (single) circumstances in which the two token features occur. If the Humean aims to be a reductive physicalist, they may suppose that such overlap motivates identifying the token features at issue, and hence the associated powers. If the Humean aims to be a nonreductive physicalist, they can reject this identification of features, on difference-making or other grounds (...). Such a Humean will suppose that attention to broader patterns of regularities can provide a basis for identifying token powers of token features, even when the token features are not themselves identical. Whether reductive or nonreductive, the contingentist categoricalist Humean can make sense of the claim that some, all, or none of the token powers of token features are identical. (...) this case is like the case of New York: if we can make it (out) here, we can make it (out) anywhere. (Wilson, 2021, p. 45, fn. 15)

Thanks to metaphysical neutrality, Wilson's schema can implement a Humean view on causal powers, which understands powers just as actual or potential contingent regularities. In this framework, the numerical identity between token powers necessary for her view to work would occur between instances of regularities at different levels. As I said, I will not thoroughly discuss this line of reasoning, as I believe Wilson's view is compelling enough even if it were unable to work for each and every view on powers. In Chapter 3 I revisit Wilson's neutrality theses.

Let us finally provide a new version of the Causal Exclusion Argument, where some of the premises of the exclusion argument are modified to refer to Wilson's neutral notion of causal power, rather than referring to causes. Given that there is a strong link between causes and causal powers, this might seem a minor change, but as I mentioned it will be crucial when assessing Wilson's solution to Kim's argument in section 2.1.1. Another point I have altered in this new version of CEA is related to the fact, mentioned in the Introduction, that most contemporary nonreductive physicalist philosophers, including Wilson herself, do not restrict their theories to mental properties. Their theories' scope also considers what I have been referring to as higher-level properties or special science properties, which are broadly scientific properties belonging to the special sciences. For this reason, I will also modify Kim's argument to generalise over higher-level properties, which include mental properties.

- (1)* *Physical Causal Closure*. Every physical effect is the manifestation of a causal power bestowed by a physical property.

- (2)* *The Non-overdetermination Principle*. An effect is not systematically produced by two distinct and sufficient causal powers.
- (3)* *Dependence*. Higher-level properties depend on lower-level physical properties.
- (4)* *Efficacy*. Higher-level properties have different causal powers than their physical realizers.
- (5)* *Distinctness*. Higher-level properties are distinct from their lower-level physical realizers.

The Causal Exclusion argument is restated as follows. Consider again that M is a mental property in the L level that bestows the causal power to cause property P* in the lower-level L-1. Given *Physical Causal Closure* (1)*, P* is produced by a causal power bestowed by a physical property. Say that M's physical base is P and P bestows a causal power sufficient for P*. Both P and M bestow causal powers sufficient for P* and, by (4)*, which aims to capture causal autonomy, those powers are different. Therefore, there is a violation of *Non-overdetermination* in every case of downward causation. Again, the causal power bestowed by the physical property excludes the causal power bestowed by the mental property. As I will show below, Wilson's nonreductive physicalist solution consists in providing a dependence relation that guarantees (3)* and that at the same time endows the higher-level property with causal autonomy. Her key move is to question what it takes to have *different* causal powers. According to Kim "different" causal powers actually means *new* causal powers: "mentality must contribute genuinely new causal powers to the world—that is, it must have causal powers not had by any physical-biological properties, not even by those from which it has emerged" (Kim, 1992, p. 135). This, I think, is the straightforward reading of (4)*, but as we will see having different powers can also be defined as having a different set of causal powers. Framing Kim's argument in terms of powers has unveiled an ambiguity in premise (4)*, which I will dispel in Section 2.3.1.

2.2 The Subset View

In this section I introduce Wilson's solutions to the argument. I begin by briefly introducing the anti-physicalist schema: the schema for strong emergence. This solution requires the denial of Physical Causal Closure: the condition essential to the strong emergence schema is that higher-level properties, such as M above, must have a new causal power not had by their lower-level bases, such as P above. If M has a new power, whenever this power manifests there will

be effects at the lower level. The reason for this is (3)* *Dependence*, which enforces that there cannot be a change in a higher-level property without a change in its lower-level realiser. Say that this lower-level effect is P*, and assuming that there is no feature (or power for that matter) at that lower level, such as P, that would be able to cause it. Therefore, overdetermination is avoided, although at the expense of Physical Causal Closure: there is only one power producing P*, namely the one had by M and not by P. Therefore, the lower level is not causally closed. This schema is no longer within a physicalist framework because it entails that there is higher-level downward causation and therefore, if the schema were true, then there would be no such thing as a complete theory of physics. It is clear that this schema is not suitable for nonreductive physicalists.

Wilson's nonreductive physicalist solution is to bite the bullet: accept that there is overdetermination but argue that it is not of a problematic kind. As no claims are denied, this is a Nonreductive physicalist account, which Wilson calls "Weak emergence". Now onto the solution. As I have said, the core of Wilson's solution is the proposal of a new dependence relation between higher-level properties and physical properties. This kind of solution is not new, as many nonreductive physicalists have proposed dependence relations before in order to avoid Kim's conclusion. However, Wilson's work is pioneering because it offers a general account—or schema—of such a relation by only providing two necessary and sufficient conditions that must hold between the causal powers of the realiser and realisee. As she argues, the general relation she proposes is the one underlying all the other dependence relations previously proposed, such as the determinable/determinate relation proposed by Yablo (1992), or Shoemaker's (e.g. 2001, 2013) mereological approach.

Wilson's dependence relation is a nonreductive realization relation and is defined by the powers-based subset strategy. According to this strategy, a property is the realizer of another property, which is its realizee, if they fulfil two conditions regarding causal powers. The first condition should not be controversial from a physicalist stance, as it is also present in reductive physicalists accounts. In fact, Kim himself accepts a principle similar to this condition, the Causal Inheritance Principle, that I will discuss later. The condition has been recently (Wilson, 2015, 2021) stated to include higher-level properties generally, although its first formulations (Wilson, 1999, 2011) only referred to mental properties. I will focus on the latter formulation.

- Condition on Causal Powers (CCP): “Every token power of [a given higher-level feature] S, on a given occasion, is identical to a token power of the base feature P upon which it depends, on that occasion.” (Wilson, 2015, p. 357)

Later formulated as:

- “Token Identity of Powers Condition: Every token power of token feature S, on a given occasion, is identical with a token power of the token feature P on which S cotermporally materially depends, on that occasion.” (Wilson, 2021, p. 58)

In her 2021, Wilson formulates the condition in terms of cotermporal material dependence, instead of realisation. Throughout her book, she uses this term to refer to what in previous works was referred to with the term “realisation”. I believe this is in line with her aim to be as metaphysically neutral as possible. By using the term “cotermporal material dependence” her account is able to work for a process ontology. Moreover, by talking in terms of cotermporal material dependence, which primarily applies to relations between particulars, Wilson is able to capture the idea that “the only matter or substance is physical matter or substance, such that any and all macro-entities (objects, systems, or other particulars) inherit their matter from (configurations of) their ultimately physical constituents” (Wilson, 2021, p. 2). However, although cotermporal material dependence operates between entities, Wilson argues that it can be easily translated to operate between properties: “What it is for a feature S to cotermporally materially depend on a feature P is for the entity bearing S to cotermporally materially depend on the entity bearing P” (Wilson, 2021, p. 11). In this thesis, however, I use the term “realisation”, which is the common term in the literature used to refer to the kind of metaphysical relation between properties I am dealing with.

The second condition imposed is the Subset Condition on Causal Powers. Again, this condition has been restated in recent publications in order to include not only mental features but higher-level features more generally.

- Subset Condition on Causal Powers (SCCP): “Token higher-level feature S has, on a given occasion, a non-empty proper subset of the token powers of the token lower-level feature P on which S synchronically depends, on that occasion.” (Wilson, 2015, p. 357)

Again, later formulated as:

- “Proper Subset of Powers Condition: Token feature S has, on a given occasion, a non-empty proper subset of the token powers of the token feature P on which S coterminally depends, on that occasion.” (Wilson, 2021, p. 59)

Let us explain how Wilson uses SCCP to solve Kim’s causal exclusion problem, in particular how she deals with overdetermination while assuring physical causal closure and non-reduction. To do so I follow Wilson (2011, 2015). In the first place, provided that every token power of a realized higher-level feature is numerically identical with a token power of its physical realizer, these causal powers do not compete with each other. There is only one power, which is had by both the higher-level feature and its lower-level realizer. *Prima facie* this is at odds with premise (4)* but, as will be clear soon, the core of Wilson’s strategy lies precisely in how we should understand the claim that higher-level features have different causal powers. Secondly, Physical Causal Closure is respected in both formulations, (1) and (1)*. Every physical effect has a proper physical cause, and every physical effect is the manifestation of a physical causal power (contrary to the strong emergentist, who claimed that there is a new power not had by the physical base). Finally, given that the higher-level feature, say M, has only a subset of the causal powers had by its physical base, P, M is distinct from P. However, we are left with the question of how premise (4)* is assured. Kim emphasises that to be causally autonomous is to have a *new* causal power. Contrary to this assertion, Wilson argues that having new powers is only one of the two ways – i.e. the strong emergentist way – for higher-level features to be causally autonomous. The other way – i.e. the weak emergentist way – a higher-level feature might be causally efficacious is by having fewer powers than its physical realizer. In my restatement of claim (4), (4)*, then, having different causal powers does not entail having a new causal power; it suffices to have a distinctive set of powers. I will return to this issue after explaining how Wilson solves the causal exclusion problem for downward causation.

Again, I focus on mental features as a particular case of the most general argument. Mental feature M (say a pain) synchronically depends on a physical base feature P_i (say, using a common example, C-fibre firing), and causes another physical feature Q (say a yelp). Given that M satisfies the SCCP in such a way that it has the power to cause Q and that this power is numerically identical to a power had by P_i , when M causes Q, P_i does cause it too. M and P_i

are both causes for Q, and given that they have a distinct set of powers they are distinct features. Hence, there is a violation of Kim's Non-overdetermination Principle but not a violation of my reformulation of it in terms of causal powers. There are two features causing Q, but not two powers causing it. Wilson works with Kim's formulation, so she cannot technically deny that there is overdetermination, rather she argues that there is overdetermination, but of an unproblematic kind. The reason, she claims, is that the effect is caused only by one token power (Wilson, 2011, 2015, 2021). She is not specific enough either about what is a problematic kind of overdetermination or about what kind of overdetermination she is accepting. I think that, as shall be clear in section 2.1, a formulation of the non-overdetermination principle in terms of powers allows a more convincing solution that, moreover, will be able to avoid Bernstein's objections to the foregoing solution.

Let us explain how having a distinct power profile is sufficient to be causally autonomous, which is key for mental causation. Kim claims that a property is causally efficacious only if it manifests a new causal power.

[The Causal Inheritance Principle] If M is instantiated on a given occasion by being realized by P, then the causal powers of this instance of M are identical with (perhaps, a subset of) the causal powers of P. (...) The implications of the Causal Inheritance Principle are devastating to Nonreductive Physicalism: if the causal powers of M are identical with those of its realization bases, then M in effect contributes nothing new causally, and M's claim to be a new, irreducible property is put in jeopardy. (Kim, 1993a, p. 355)

Wilson's response to the causal exclusion argument will succeed only if her account of Weak Emergence is able to endow weakly emergent features with reality and autonomy, in particular causal autonomy, and thus avoid Kim's objection. So far, Wilson has proposed two arguments to defend this. I call the first argument "The Proportionality Argument" because it is very much in debt to Yablo's (1992) proportionality constraint, which states that causes should be proportional to their effects. I call the second argument "The Causal Joints Argument" because it is grounded on the claim that natural laws track causal joints in nature.

The Proportionality Argument

This argument aims to show that mental states are the proportional cause of at least some other

mental states, and it could be expanded to show that they are the proportional cause of some physical states. The argument is based on the multiple realizability of mental properties. The subset view can account for multiple realizability by imposing that the causal powers of M are the powers inherited from the intersection of its multiple physical realizers, as shown in Figure 1.

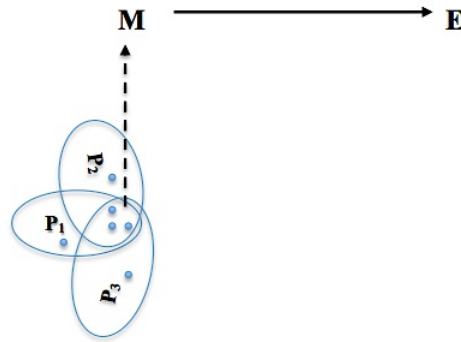


Figure 1. M inherits the causal powers contained in the intersection of its physical realizers P_1 , P_2 and P_3 , which include the power to cause E. Therefore, M's causal profile is distinct from the causal profiles of each of its physical bases (consider that every dot is a causal power).

I shall explain the argument following Wilson (2011). Let M be a feeling, such as being thirsty, realized by base property P_1 and let E be a desire, such as the desire to drink water, caused by M. Consider that, as it is shown in Figure 1, M is multiply realizable by physical base states P_1 , P_2 and P_3 . Wilson's point is that when M is realized, in similar circumstances, by P_2 or P_3 instead of P_1 , M will also cause E. This is because E is a manifestation of the causal powers of M and these powers are shared by P_1 , P_2 and P_3 . The powers had by P_1 and not by P_2 or P_3 make no difference. The fact that M has a distinctive power profile that contains only the causal powers relevant to cause E is sufficient to claim that M's causal role regarding the effect E is distinct from the causal role of its physical realizers. In Stephen Yablo's (1992) terms, which Wilson endorses, we would say that M is the proportional cause of E because it is M that makes a difference to the bringing about of the effect E, whereas its physical realizers do not. That said, Wilson is able to retain the causal sufficiency of P_1 because the power had by M, E is a manifestation of is numerically identical to a power of P_1 . This is very interesting, as Yablo also claimed that the causal sufficiency of the lower-level realiser of M to cause E should be preserved. If not, there would be a violation of the causal closure of the physical (which we are trying to avoid). However, I believe Wilson's view provides a more nuanced way to explain how we get to keep the causal sufficiency of the realiser yet avoid overdetermination. Surely, Yablo persuasively argues that M is the proportional cause to the bringing about of effect E,

and there is a sense in which it is compelling that the kind of overdetermination between P_i and M seems unproblematic: given that in Yablo's view P_i and M stand in a relationship of determinate/determinable, these two properties are not causal competitors. Nevertheless, by introducing her two conditions on causal powers, Wilson's view is able to offer a nuanced explanation that, moreover, does not rely on there being a determinate/determinable relationship between realiser and realisee.

The Causal Joints Argument

This argument was first put forward in Wilson (2010), and it is complementary to The Proportionality Argument in that it does not require higher-level features to be multiply realized, but only singly realized. Wilson (2010) introduces a *degrees of freedom* (DOF)-based account of weak emergence: she argues that the fact that some higher-level entities and states are described by fewer degrees of freedom than their lower-level realizers entails that the former are irreducible and moreover causally autonomous with respect to the latter. Degrees of freedom are the minimum number of parameters (or coordinates) —independent from each other— required to describe the state of an entity. For instance, a quantum state is characterised, among other degrees of freedom, by the one corresponding to its spin. Those degrees of freedom, in terms of coordinates and associated values, are the information required for natural laws to operate at the quantum level. However, at the classical limit, that is, when quantum entities assemble to form complex macroscopic entities, the degrees of freedom corresponding to quantum features, such as colour charge or the already mentioned spins, are not needed to characterise such macroscopic entities. Therefore, they are eliminated: no description of a macroscopic entity would ever need the spin or colour charge as a parameter, and hence no natural law at such higher level operates with spin or colour charge parameters. In short, lower-level entities behave in more unconstrained ways than those at higher levels, and therefore there are more degrees of freedom needed to account for their behaviour. This is what grounds her DOF-based account of weak emergence. The fact that higher-level properties have less degrees of freedom than their realisers parallels the subset condition on causal powers: higher-level features can do less than their realizers. I put aside now discussion of the DOF-based weak emergence, which I will resume in Chapter 5. Let us go back to the Causal Joints argument. Here is a nicely summarised version of her argument:

To start, laws require certain kinds of information in order to operate. For example, in order for the quantum laws to operate on a configuration of fundamental physical particles, the input into the laws has to contain information about the individual quantum spins of those particles. Suppose also that it is a metaphysical consequence of the quantum configuration's existing and evolving as per the quantum laws that a given planet exists and evolves as per the laws of astronomy, where (as a Weak emergentist might well maintain) the characterization of planets and their behaviors is taken to abstract away from any details concerning quantum spin. Were the planet, so characterized, to be given as input into the quantum laws, this input would not contain the information about quantum spin required for the laws to operate. Correspondingly, and more generally, planets and their features would not be appropriately placed at the fundamental physical level, since the physical laws would not be able to operate on such entities or features. (Wilson, 2021, p. 30)

The above provides sufficient reason, Wilson claims, to argue in favour of ontological irreducibility and causal autonomy. The argument is the following. In those and similar cases of degrees of freedom elimination higher-level features constitute a new level, because there is no way to describe those features by lower-level laws —laws need a certain set of parameters to operate and cannot operate with less. For instance, you do not get a planet's behaviour from, say, Schrödinger's equation, but from astronomy laws. In Wilson's words, the fact that M is causally autonomous is supported by the fact that "M produces E as part of a different system of laws than P" (Wilson, 2011, p. 129). How does this relate to having a distinct causal profile? Being part of a different system of laws is directly related to having a distinct set of powers or causal profile because "systems of laws track causal joints in nature" (Wilson, 2011, p. 129), and properties with distinct power profiles are related to distinct causal joints in nature. From these reasonings, we can infer that what is required for causal autonomy is having a distinct causal powers profile. Therefore, the fact that, as in our previous example, P and M cause E in virtue of different laws is sufficient to assert that M is causally efficacious vis-à-vis E. The fact that every power M has is numerically identical to a power had by its realiser, P, is irrelevant.

The range of higher-level features to which one can apply this argument will depend on the extent to which the argument relies on the elimination of degrees of freedom. Although there are some clear examples from the sciences (such as the classical limit already explained, but also statistical mechanics aggregates) that entail the elimination of degrees of freedom and thus can accommodate the argument, this is not the case for many other features Wilson would like to argue are weakly emergent. In particular, mental features. This is acknowledged in her 2010,

as we can read: “Can attention to DOF [degrees of freedom] (...) establish that NRP is a live position, for the cases that most matter [i.e. mental features]? The answer is: I’m not sure. It is not obvious that mental phenomena involve an elimination in DOF; but given our present rudimentary level of theoretical insight, it is also not obvious that they don’t.” (Wilson, 2010, pp. 308 – 309). However, in *Metaphysical Emergence* (2021), as well as in her 2011 paper, she presents a more general version of the argument, which does not hinge on degrees of freedom elimination, but on the mere assertion that there actually are different special science laws governing phenomena at different levels. Wilson then takes the previous argument from here: given that laws are related to causal joints in nature, then higher-level features are related to distinct causal joints than their realizers. However, given that degrees of freedom elimination was the main argument to claim that higher-level features cannot be encompassed by the laws at the level of their realizers, there is a sense in which this more general version of the argument for causal autonomy is weaker.

From Kim’s stance the only way to be causally efficacious is to have a new causal power. This is the reason why Kim claims that Nonreductive Physicalism cannot be sustained and collapses either into Strong Emergentism or reductionist physicalism. However, Wilson’s Nonreductive physicalist account, namely Weak Emergentism, is able not only to deal with the causal exclusion argument but also to avoid the conclusion derived by Kim from the Causal Inheritance Principle. A given higher-level property M has only powers numerically identical with powers of its lower-level constituents and, assuming SCCP, arguably it is still causally efficacious.

2.3 Scrutinizing Sara Bernstein’s Overdetermination Objection against the Subset View

As I have mentioned above, most Nonreductive physicalist theories, and Wilson is no exception, aim to neutralise the exclusion argument by proposing a dependence relation between physical properties and mental properties that is used to deal with overdetermination. Sara Bernstein (2016) argues that Nonreductive physicalists tackle the problem of overdetermination in two different ways, so they can be divided in two groups: authors in the first group, so-called “accepters”, accept overdetermination and then argue that the kind of overdetermination taking place in higher-level causation is not problematic; authors in the second group, so-called “deniers”, deny that the form of higher-level causation they defend

entails overdetermination in the first place. Bernstein's ambitious goal is to show that Nonreductive physicalists of both groups have not properly addressed the problem of overdetermination and that, for this reason, they have not realised that Nonreductive Physicalism is an untenable position. For this reason I call her objection the Overdetermination Objection. Before going into detail about what problems Wilson may face, I discuss whether Wilson is an "accepter" or a "denier". To do so, I review fragments of some of her papers on the Subset View in which she refers to overdetermination.

In Wilson 1999, 2009 she argues in favour of the proper-subset strategy as underlying the determinable/determinate relation proposed by Yablo (1992). She claims that "determinables and determinates are not causal competitors; (...) [hence] overdetermination and exclusion are avoided" (Wilson, 2009, p. 150) and "this conception of the mental allows mental properties to be efficacious, without this efficacy introducing overdetermination" (Wilson, 1999, p. 47). From these two claims one infers that Wilson is a denier. However, in Wilson (2011, 2015, 2021) she is clearly an accepter: "a realization relation satisfying SCCP avoids *problematic* overdetermination" (Wilson, 2011, p. 128, my italics) and "allow that there is overdetermination, but deny that it is of the firing squad variety that would be intuitively problematic as generally characterizing higher-level causation." (Wilson, 2015, p. 353). I think that underlying this change of phrasing there are two different understandings of overdetermination, which unfortunately Wilson does not make explicit. My aim in the following section is to make them explicit, which will be the core of the first solution I propose to block Bernstein's arguments against Nonreductive Physicalism.

Bernstein's arguments against Nonreductive Physicalism are directed to deniers and accepters separately. First of all, let me introduce Bernstein's schema for overdetermination, which she claims is widely endorsed although the understanding of the underlined terms may vary among Nonreductive physicalists:

OD Schema: Causes c1 and c2 overdetermine an effect e if (i) they are distinct, (ii) they occur, and (iii) they are each sufficient to cause e in the way it occurs. (Bernstein, 2016, p. 20).

As this is a schema, Bernstein aims to be neutral about the understanding of the underlined terms. I will discuss in due time how Wilson understands some of them. However, I need to briefly mention here that by referring to "the way e occurs", Bernstein is not saying here that

the event will be caused in the same way when only one of the two causes as when both causes occur. As I discuss in the *Argument against accepters* subsection, she claims that when both causes occur, the event will be caused in a different way. What she is referring to is that the two causes independently are sufficient to cause the event *e* in the way it occurs.

Argument against deniers

In order to deny overdetermination Nonreductive physicalists should deny either (i), (ii) or (iii). However, denying (i) would entail reduction or elimination of one of the causes, denying (ii) would entail causal irrelevance, and finally denying (iii) would entail that either (or both) *c*₁ and *c*₂ do not have the causal efficacy the nonreductive physicalist needs them to have. Hence, Bernstein claims that given that nonreductive physicalists cannot deny overdetermination without putting the core of their theory in jeopardy, no nonreductive physicalist should be a denier. In the following section I shall scrutinise this assertion because I believe that the strength of Wilson's theory is precisely that she can be a denier, as she in fact used to be, without compromising the core of her theory.

Argument against accepters

Let us discuss Bernstein's argument on why accepters such as "late" Wilson face an overlooked and crucial problem regarding overdetermination. According to Bernstein, they have addressed easy problems: ubiquity —i.e., the fact that a nonreductive physicalist view implies that overdetermination is pervasive—, and coincidence —i.e., overdetermination implies the apparently unexplainable repeated coincidence of two causes producing the same event in the same way—. Of course, ubiquitous overdetermination is not desirable, but it can be arguably accepted from a nonreductive physicalist stance if the metaphysical excess is counterbalanced with the explanatory power of the theory. Regarding coincidence, the nonreductive physicalist does not really face a problem because the two causes are not coincidental: there is a dependence relation binding them (Bernstein, 2016, p. 29-30).

I shall explain what Bernstein considers the hard problem in what follows. According to Bernstein in every case of overdetermination one should consider two questions: where the extra causal power "comes from" and where it "goes" (Bernstein, 2016, p. 30). The two-rocks-throw case —i.e., rock A and B are each sufficient to break a window, and they impact the

window at the same instant of time— is “causally satisfying” because these two questions can be easily answered. The extra causation comes from the extra rock, and it goes to the shattering of the window with twice the force it would have been shattered if only one rock had been thrown. So, there is a difference between the case where only one rock shatters the window and the case where two rocks shatter the window. If only A or B impacts the window with force N , the window will shatter with that force, N . If both A and B impact the window at the same instant, the window will shatter with force $2N$. What about mental/physical overdetermination?

Mental/ physical overdetermination has no such difference, since, holding the physical cause fixed, removing the mental cause does not change the way the mentally caused outcome occurs. This difference, I suggest, is the reason overdetermination is problematic for the nonreductive physicalist: it reinforces the central tension of the nonreductive physicalist’s view rather than resolving it. (Bernstein, 2016, p. 30)

According to Bernstein, the central tension of Nonreductive Physicalism is the following. On the one hand, if mental/physical overdetermination is too similar to two rocks, then physicalism is threatened. On the other hand, if mental/physical overdetermination is not enough like two rocks, then the mental cause becomes “derivative and un-causation-like” (Bernstein, 2016, p. 31) and this puts nonreduction of mental features at stake because, Bernstein holds, if higher-level features are not causes then there is no reason for them not to be reduced to their realizers or eliminated. Bernstein’s reasoning sounds familiar. This is because it is none other than Kim’s claim that Nonreductive Physicalism is an unstable view that collapses either into Reductive Physicalism or Strong Emergentism.

Moreover, Bernstein’s way of arguing for the central tension enables her to go beyond the central tension itself and propose a new objection to Wilson’s theory. The weak emergentist is able to answer where the extra causal powers come from, but they cannot say where the extra force goes to. The question about where extra causal powers come from is answered by appeal to the fact that higher-level features share their powers with a set of the powers of their realizers. However, this raises an already known problem: “there is only one set of causal powers that the physical and mental share. But shared causal powers do not vindicate efficacy of the mental *qua* mental, because the mental itself is not the source of the efficacy.” (Bernstein, 2016, p. 31-32). In this quote Bernstein is arguing that the subset strategist cannot answer where the extra cause goes to because there is nothing the mental causes *qua* mental, there is no difference it

makes to the effect. Therefore, mental causation as described by the subset strategy is nothing like two-rocks and, as per the central tension, nonreductionism is at stake.

It is evident from the quote that Bernstein is making the same assumption as Kim, i.e., causal autonomy requires *new* causal powers, with the same goal in mind: undermining nonreductive physicalism. Nonetheless, she does not mention the assumption. In fact, it looks like it is the result of a commitment she makes at the beginning of the discussion, when after considering different kinds of causation she claims: “I will limit the present discussion to views of mental causation that involve “oomph”-style causation” (Bernstein, 2016, p. 23). According to “oomph” theories of causation, causation requires a transfer of quantity from the cause to the effect, and for this reason this kind of causation is also called “production”. It is straightforward to understand why oomph-style causation requires new causal powers: if the mental is to be causally relevant at all, it must produce an identifiable change on the physical, it must contribute with a specific force to the causing of the effect. Certainly, Wilson is not committed to this view on causation, and this is a flaw in Bernstein’s argument that I shall explore in what follows.

2.3.1 Two Solutions to Bernstein’s Overdetermination Objection

In what follows I present two ways to overcome Bernstein’s argument against the subset view. In particular, I shall focus on how it affects Weak Emergentism. The first is to propose an alternative reading of *OD schema* tailored to Weak Emergentism, that would allow a denial of overdetermination without jeopardising nonreduction. The second consists in making explicit Bernstein’s undiscussed assumption underpinning the central tension and the subsequent objection to weak emergentism, which is no other than Kim’s claim that only new causal powers confer causal autonomy. Once this is done, it shall be clear that Wilson can remain an acceptor without compromising nonreductionism.

I think that *OD Schema* is ambiguous enough to admit (at least) two different readings. In fact, Wilson herself is also unspecific about the principle, as I have mentioned above. I think that her change from being a denier to be an acceptor could be due to the fact that she has considered two different views on overdetermination through time, or simply to the fact that she did not consider it necessary to disambiguate it so it really does not matter whether one is a denier or an acceptor. Nonetheless, Bernstein shows that it does matter, so at this point it is key to

disambiguate the two readings, and then show that Wilson's view can actually deal with both of them. Moreover, disambiguating the two readings may also help to understand Wilson's change. The readings are the following:

OD Schema: Causes c_1 and c_2 overdetermine an effect e if (i) they are distinct, (ii) they occur, and (iii) they are each sufficient to cause e in the way it occurs. (Bernstein, 2016, p. 20).

Causes c_1 and c_2 are understood by Bernstein as states or events, but they can also be understood as powers. Accordingly, I rephrase *OD schema* as follows.

OD Schema*: Causal powers c_1 and c_2 overdetermine an effect e if (i) they are distinct, (ii) they occur, and (iii) they are each sufficient to cause e in the way it occurs.

Denier's solution

The *OD Schema** disambiguation of *OD Schema* allows for a Denier's solution that does not jeopardise nonreductive physicalism. As the reader may notice, *OD Schema** resembles the understanding of overdetermination I proposed when rephrasing the Exclusion argument. In this case, the Weak emergentist denies that cases of mental causation satisfy (i) because c_1 and c_2 are numerically identical. Hence, in the light of this reading the Weak emergentist is a denier. Causal powers c_1 and c_2 are not distinct because the powers of mental properties are a subset of the powers had by the physical properties. However, the fact that they are not distinct does not preclude the discussion over the causal efficacy of mental properties. As I have explained above, Bernstein focuses on "oomph" theories of causation. Weak emergentism, however, does not require an extra "oomph" for there to be a case of real causation. I shall discuss this when explaining the acceptor's solution to Bernstein's objection, because it is relevant to both solutions. Finally, the denial of overdetermination avoids the "central tension": there is no dichotomy regarding whether causation is too much or not enough like two rocks; it is nothing like two rocks, and that is just fine.

The Acceptor's Solution

The acceptor's solution does not need to disambiguate *OD Schema*, which is the version actually discussed by Bernstein. According to *OD Schema*, Weak Emergence entails

overdetermination because there are two distinct properties, a mental property and its physical realizer, causing a physical effect. Assuming that coincidence and ubiquity are not a problem, as Bernstein argues, the real problem of accepting overdetermination is that it reinforces the “central tension” of Nonreductive Physicalism views. In particular, Bernstein claims that, per the “central tension”, the subset view is not a nonreductionist account. The reason is that it is nothing like two rocks. Two-rocks is “causally satisfying” because we can identify where the extra cause comes from and where it goes. However, for the Weak emergentist there is no way to tell where the extra cause goes, because in fact there is no extra cause: there is only one shared power causing the effect and, Bernstein claims, “shared causal powers do not vindicate efficacy of the mental *qua* mental, because the mental itself is not the source of the efficacy.” (Bernstein, 2016, p. 32).

What I shall argue is that the Weak emergentist can accept overdetermination without falling prey to the upshot of the “central tension”. First of all, as I have shown above, Bernstein’s “central tension” is nothing but Kim’s argument that Nonreductive Physicalism is an untenable view because it collapses either into Strong Emergentism or Reductive Physicalism. However, Kim’s argument has been answered by Wilson’s subset strategy. Of course, the claim that having distinct power profiles is sufficient for causal efficacy is not straightforwardly true, but Wilson provides two arguments that show in what sense it might be. Therefore, the burden of the proof is not on the Weak emergentist side, as Bernstein intends: she is just rephrasing Kim’s tension, which has already been answered by appeal to distinct and causally autonomous causal profiles. Even so, from the Weak emergentist side we can make a stronger claim: Bernstein is testing Weak Emergence against too restrictive a view of causation. It is worth mentioning that Bernstein claims that she will limit her objections to theories understanding causation as “oomph”, and nonetheless she addresses weak emergentism. Of course, weak emergentism does not provide higher-level properties with the causal autonomy an “oomph” theorist would require.

(...) there are two ways for a mental feature (event, property, state) M to be causally relevant—two ways for a mental feature to be efficacious *qua* the type of mental feature it is—as compared to the physical feature upon which the mental feature depends. One way—emphasized by Kim and others—is for M to be associated with a new power to produce the effect (...) here the distinctive efficacy reflects facts that are broadly *intrinsic* to the effect’s production. Another way—the way at issue in the powers-based strategy for accommodating Nonreductive

Physicalism—is for M to be associated (...) with a collection of powers that are relevantly proportional to the effect in the ways indicated by difference-making considerations and special-science laws. Here, the distinctive efficacy reflects, in part, facts (concerning which power profiles are associated with which properties) that are broadly *extrinsic* to the production of the effect. (Wilson & Bernstein, 2016, p. 321, my italics)

This quote serves to emphasise the distinction between Bernstein’s and Wilson’s accounts of causation. As I have explained before, there are arguments supporting the claim that having a distinct causal profile is enough for causal efficacy. Of course, it is not enough for an intrinsic source of efficacy in terms of “oomph”, which is the one considered by Bernstein, but it is indeed sufficient for an extrinsic and relational one. What Bernstein should provide for her objection to be solid is an argument against extrinsic sources of causal efficacy, so that causation in a case of overdetermination could be understood *only* à la two-rocks. If this argument is not in place, then there is no reason why one could not endorse overdetermination and claim that it is nothing like two-rocks, and yet avoid reduction.

CHAPTER 3. ASSESSING SYDNEY SHOEMAKER'S NONREDUCTIVE PHYSICALIST ACCOUNT

3.0 Introduction

Wilson's weak emergentist account is designed to avoid Kim's causal exclusion argument. However, Wilson's is not the only theory aiming to overcome Kim's argument from a non-reductive physicalist stance. In Chapter 2 I pointed out that Wilson's response to Kim is actually a schema. As such, the theory aims to be general enough so that it can be argued that any successful nonreductive physicalist account is, either explicitly or implicitly, implementing it.² In particular, Wilson (2011, 2015, 2021) argues that the three most popular responses to Kim's argument: functional realization, mereological realization, and the determinate/determinable relation, as well as constitutive mechanism, are just four different implementations of her schema.³

Sydney Shoemaker has been the main advocate of mereological realization to solve Kim's Causal Exclusion argument. Shoemaker proposes a subset relation between powers similar to Wilson's. Moreover, he argues that given that properties are individuated by their causal powers, higher-level properties, which have a subset of the powers of their realizers, are *parts* of their physical realizers. According to Shoemaker, it is in virtue of this parthood relation that higher-level properties are causally efficacious vis-à-vis their physical realizers.

3.1 Advantages of Wilson's Schema over Shoemaker's Account

Shoemaker's mereological realization and Wilson's weak emergentism have several points in common, the most important being that Shoemaker's solution follows a subset strategy similar to Wilson's. Actually, historically many have referred to the subset condition as Shoemaker's

² The term "implement" is widely used by Wilson throughout her work. It is used to refer to her view as the common core shared by all the other solutions. Although Wilson has favoured the determinate/determinable relation as the most successful implementation of her view, I will not engage here with this discussion.

³ I will not assess Wilson's arguments to defend that those solutions are implicitly implementing her schema, and I will assume for the sake of the discussion that her arguments are successful. See Wilson (2011, 2021) for a detailed discussion.

condition, omitting Wilson's contribution.⁴ Another similarity is that he also argues that, as we can read in the quote below, the determinable/determinate relation and functional realization are implementations of his mereological realization account:

The instantiation of the determinate entails the instantiation of the determinable, and can quite naturally be said to include it. It seems natural to me to say that being scarlet is in part being red. Likewise, the instantiation of a realizer property entails, and might naturally be said to include as a part, the instantiation of the functional property realized. The conditional powers conferred by the instance of the determinable or functional property are a proper subset, and in that sense a part, of the conditional powers conferred by the instance of the more determinate property that realizes it on a particular occasion. (Shoemaker, 2001, pp. 80- 81)

Strikingly, Shoemaker has never discussed Wilson's view in any of his publications, although, as I have said, she does argue that Shoemaker's mereological realization is an implementation of her schema. There is therefore a one-sided debate on the subset view in which Wilson argues that mereological realization implements her schema and Shoemaker provides no answer as to whether this is so. In this chapter I pursue what I think is a more interesting way –one that maybe would prompt a response– to assess the views: I will consider them competitors for the ultimate subset view. This is justified by the fact that both Shoemaker and Wilson defend a subset strategy and both of them argue that the determinable/determinate relation and functional realization are implementations of their views, but their strategies differ in that they require different commitments. As I will explain in what follows, Wilson manages to remain neutral about several controversial issues throughout her works, in what I call Wilson's Neutrality Theses, while Shoemaker's view requires taking a stance on some of these issues before it can get off the ground. Therefore, given that their strategies and their target are the same but they differ in the commitments required for the strategy to be successful, assessing them as competing views will be an illuminating discussion. On the face of it, Wilson's theory should be preferred, as she needs fewer commitments for her strategy to work. However, if it turns out that Shoemaker's commitments are necessary conditions for a subset view to work,

⁴ Although the first published work introducing the subset strategy is Wilson's 1999, Shoemaker often presented himself as the main proponent of the subset view, which he published for the first time in Shoemaker (2001). Even more strikingly, in his Shoemaker (2013, p. 41) when introducing the origins of the view he only mentions Clapp (2001) and Watkins (2002). To know more, see Wilson (2015, p. 357, footnote).

then Wilson's theory would be lacking important insights and Shoemaker's account should be preferred. But first let me introduce Wilson's Neutrality Theses.

3.1.1 Wilson's Neutrality Theses

In *Metaphysical Emergence* (2021) Wilson mentions several points of non-commitment, but for the purposes of this section there are only three Neutrality Theses I need to introduce. The first one is Wilson's stance on what causal powers are. Wilson's definition of causal powers is broad enough to capture a wide range of positions on this issue: causal powers of a property are just "what causal contributions possession of a given property makes (...) to an entity's bringing about an effect". This is a general definition that arguably most philosophers would endorse, independently of what specific theory of powers they hold. Therefore, defining causal powers in this neutral way allows her to avoid the objections and shortcomings of more specific theories of causal powers and moreover its neutrality makes her weak emergentist account attractive to supporters of diverse views on powers. As this fragment claims, Wilson's strategy is valid independently of what theory of powers one advocates.

Philosophers differ over the nature of powers, their role in causation, and whether they are essential to the features that have them. (...) No commitment on any of these issues is required in order to implement the powers-based strategy to follow. (...) As we will see, even a contingentist categoricallist Humean can implement the powers-based strategy. (Wilson, 2011, p. 127)

Wilson argues that her view can be implemented independently of what one takes powers to be and what one considers to be their relation with properties. For instance, if one has a categoricallist Humean understanding of powers and therefore takes powers to be (in some way or another) mere regularities between categorical features, then Wilson's conditions on powers can still be applied: the regularities involving a higher level feature are a subset of those involving its realizer (Wilson, 2015, p. 355, footnote). Given that the notion of power is intimately linked to those of property and law, two further neutrality theses stem from Wilson's neutral stance on powers, which she acknowledges and embraces:

- (1) Features need not have their powers essentially, but they could have them contingently,
and

- (2) Features need not be individuated by their powers but might also have non-causal aspects such as quiddities or phenomenal features and be further or exclusively individuated according to them (Wilson, 2015, p. 355, also in 2021, p. 43).

For the purpose of this chapter, these two are the only neutrality theses I need to endorse to favour Wilson's account against Shoemaker's. I will not discuss Wilson's neutrality thesis about the nature of powers. Contrary to Wilson's neutrality theses, Shoemaker's subset view argues that a subset strategy requires a commitment to a certain theory of powers and properties. In section 3.1.2, I will first introduce Shoemaker's view and what those commitments are. Then I will argue that, in fact, Shoemaker's view needs stronger commitments than he claims. Once I have spelled out the stronger commitments Shoemaker's view needs to succeed, I will move on to discuss why Wilson's Neutrality Theses are preferable. A trivial answer would be that if everything can be done with fewer commitments, then so be it. But I will provide a further argument, The Individuation Argument, to support Wilson's theory. The argument shows that not only she can achieve the same goals as Shoemaker with fewer commitments, but that she can actually do more than Shoemaker with less.

3.1.2 The Advantages of Neutrality

Shoemaker's mereological approach to realization was first introduced in Shoemaker (2001). Although in his book *Physical Realization* (2007) he made some changes to avoid objections to his previous account, in recent papers he has offered a newly revised version of the theory that goes back to the original view (Shoemaker, 2011, 2013). For this reason, regarding those aspects of the theory that have differed over the years I will focus only on the first and recently re-endorsed work. Shoemaker explicitly proposes a subset strategy grounded in a condition on powers similar to Wilson's subset condition. In the following quote we can see how close their views are, while their key difference is also apparent: Shoemaker proposes a parthood relation between realizer and realizee.

I hold against this [Kim's Causal Exclusion Argument] that the causal powers of the P instance [realizee] are a proper subset of those of the R instance [realizer], that this makes the former instance a part of, rather than identical with, the latter, and that this allows the causing of an effect by the P instance to be a manifestation of the causal profile of P. In such a case both instances can be said to cause the effect; but the R instance will cause it because it contains the

P instance as a part. This is not overdetermination of an objectionable sort. (Shoemaker, 2007, p. 53)

Shoemaker's full account of realization is divided in two parts: (1) an account of property realization, and (2) an account of microphysical realization, which is an account of how properties are realized in microphysical states of affairs. Most of the literature on realization relations focuses only on (1), including Wilson's account. However, Shoemaker argues that a physicalist theory of property realization also needs to deal with how properties are realized in microphysical states of affairs. I will engage with this kind of realization in section 3.2 and 3.3. For now, I will focus on property realization.

According to Shoemaker, property realization is a constitutive relation between instances of properties. One key feature of this relation is that the realizer is metaphysically sufficient for the realisee. This metaphysical sufficiency clause is, according to Shoemaker, required for physicalism: only if the realizer is metaphysically sufficient for the realisee can we say that the realisee is nothing over and above its realizer (more on this below). The subset view he endorses has the same condition on powers as Wilson's: "property P has property Q as a realizer if the forward-looking causal features of P—its aptness to contribute when instantiated to the production of certain effects—are a subset of the forward-looking causal features of Q." (Shoemaker, 2013, p. 41, also in 2011). The main difference from Wilson's view, as we can read in the previous quote, is that he claims that realization is a mereological relation. In Shoemaker's account, realized properties are parts of their realizers: "[A higher-order property] is *part* of the latter [its realizer]. And that seems the right conclusion to draw from the fact that the causal powers of the former are a proper subset of those of the latter." (Shoemaker 2013, p. 46). Given the subset condition on powers, and on the assumption that properties are individuated by their powers (I will assess this assumption shortly), then realized properties are part of their realizers. It is precisely in virtue of this parthood relation that Shoemaker vindicates the causal efficacy of higher-level properties and avoids overdetermination. Roughly put, the argument for higher-level causal efficacy is the following: a realized property, say M, is a part of its realizer, say P; the causal powers of M, which are a proper subset of those of P, are the causal powers responsible for a certain effect E. Shoemaker claims that it is only because P has M as a part that it causes E, so M is the "proportional cause" of E, in Yablo's terms. "So while it is true that the instance of C-fibre firing "does the causal work," it does not do so in a way

that leaves the instance of pain with no work to do; on the contrary, it does the causal work *because* it includes as a part the instance of pain” (Shoemaker, 2007, p. 47, my italics).

From this brief introduction to Shoemaker’s account, we can already point out at least three key differences between his and Wilson’s views: 1) Shoemaker’s approach requires properties to be individuated by their causal profiles, 2) Shoemaker claims that higher-level properties are parts of their realizers, and 3) Shoemaker requires realizers to be metaphysically sufficient for the instantiation of their realizees. Most of this section will be devoted to discussing the first and second difference, and how they are related.

In all of the aforementioned works Shoemaker has argued that a straightforward way to guarantee that properties are individuated by their causal profiles is to endorse his Causal Theory of Properties (CTP), which is a theory that he first introduced in Shoemaker (1980) and further defended in Shoemaker (1998). The CTP is a theory about the nature of those properties that play a role in causation –it says nothing about, for instance, the property of being a prime number. It argues that such properties are exhaustively individuated by their causal powers in the strong sense that causal profiles are everything that is essential to such properties. This entails that properties possess the same causal profiles across different worlds, that the instantiation of such causal profiles is sufficient for the instantiation of their associated properties, and that there are no two different properties with the same causal profile.

Knowing that the CTP is controversial, Shoemaker has argued that his subset strategy does not require such a strong thesis, but that a weaker version of property individuation in terms of powers is sufficient. I will call it property individuation* and is the following: “each such property is individuated by a causal profile in the sense that it and it alone has that profile in the actual world and worlds nomologically like it.” (Shoemaker, 2007, p. 142). Also put as “in the actual world, and worlds nomologically like it, having that causal profile is sufficient for being that property.” (Shoemaker, 2007, p. 5). Property individuation* is not as controversial as property individuation because it only claims that having a causal profile is necessary and sufficient for the instantiation of its associated property in our world and worlds nomologically like ours, and remains neutral about worlds with different laws. So in worlds nomologically like ours properties are individuated by their causal powers. Shoemaker accepts that any given property might have a different causal profile in worlds with different natural laws. Property individuation* is therefore weaker than property individuation as per the CTP, in particular, it

is modally weaker. Property individuation* certainly seems an ad hoc proposal aimed at engaging readers unconvinced by the CTP. In fact, it looks more like a nomological correlation relation between properties and their causal profiles, than an individuation relation. My scepticism towards property individuation* motivates what I call the Metaphysical Necessity Argument. The argument shows that actually Shoemaker needs a commitment stronger than property individuation* to guarantee a parthood relation between realizer and realizee. Before that, let me explain why Shoemaker's mereological realization needs to presuppose the arguably problematic view that properties are individuated in some way or another by their causal profiles.

The reason is that the individuation of properties by their powers is one of the premises leading to the conclusion that realized properties are *parts* of their realizers (2013, p. 47). Shoemaker argues that the realizer has the realized property as a part because 1) given the subset view, the instantiation of the realizer is sufficient for the instantiation of the causal profile of the realizee; and 2) given that properties are individuated by their causal profiles in the sense (discussed above) that having a causal profile is sufficient for having a given property, then the instantiation of the realizer constitutes the instantiation of the realizee. If properties were not individuated by their powers, then the fact that the realizer instantiates the causal profile of the realizee would not be sufficient for the realizee to be instantiated, so it would not be guaranteed that the instantiation of the realizer includes the realizee as a part.

His argument, let us call it The Parthood Argument, can be spelled out as follows:

P1. Subset Condition: Property P realizes a higher-level property M iff M has a proper subset of the powers of P.

P2. Property Individuation*: In our world and worlds nomologically like ours, having a given causal profile is necessary and sufficient for having its associated property.

Conclusion. Property P includes M as a part.

Arguably there should be a third premise, stating what it is for a property to stand in a parthood relation to another property. However, Shoemaker does not provide a definition of the parthood relation he proposes. He does say that “being scarlet is in part being red” (Shoemaker 2001, p. 80), and that given the subset view it “might naturally be said that [the realizer] includes [the realizee] as a part” (Shoemaker, 2001, pp. 80- 81). I will discuss this below.

In what follows I scrutinise The Parthood Argument, in particular focusing on the role of P2, Property Individuation*. On the one hand, I will argue that Shoemaker's P2 is not strong enough to guarantee a parthood relation between realizer and realizee. To do so, I introduce what I call the Metaphysical Necessity Argument. On the other hand, I will show that P2 is too strong to accommodate the possibility of properties that have essential aspects that float free from their causal powers —from now on, non-causal aspects—, such as phenomenal non-causal aspects of mental properties or quiddities. The latter is not a problem per se but given that Wilson's account does not deny the possibility of such features, the discussion will provide a further reason to prefer Wilson's weak emergentism.

I will start by introducing two claims that I consider are implicit in Shoemaker's understanding of mereological realisation. A1 is a claim about the modal strength of the parthood relation between properties, and A2 states a fairly uncontroversial feature of the relation *being metaphysically necessary for*.

A1. Whenever a property is part of another, having the part-property is metaphysically necessary for having the whole property.

A2. The relation *being metaphysically necessary for* is transitive.

Let me lay the ground to introduce the Metaphysical Necessity Argument, which requires the truth of A1. In general, it is true that parts of an object are not metaphysically necessary for the whole object, but in this case we are talking about properties, and the talk of properties that is relevant here seems to have universal application: when Shoemaker says "being scarlet is in part being red", he surely means that for absolutely anything that is scarlet, it is red. So by saying that having a whole-property is in part having a part-property, Shoemaker is in fact implying that having the part-property is part of what it is to have the whole-property, which is a much stronger thesis than the parthood relation between objects and their parts. Using Shoemaker's example again, "being scarlet is in part being red" could be rephrased as "being red is part of what it is to be scarlet". The Metaphysical Necessity Argument is the following. A1 is true, because being red is metaphysically necessary to being scarlet, so it must be preserved by a mereological realisation relation. However, does The Parthood Argument guarantee A1? Given that the Property Individuation* premise only refers to worlds nomologically like ours, we can think of a world with different laws than ours where red is not

a part of scarlet, as they could have different causal profiles. So Shoemaker's property individuation* violates A1, thus making it difficult to understand how the parthood relation would work. Therefore, Shoemaker needs a stronger relation between part-properties and whole-properties, i.e. metaphysical necessity. Hence, his Property Individuation*, P2, needs to be strengthened. To guarantee metaphysical necessity properties must be individuated by their powers in every possible world. This can be done easily by accepting CTP, given that according to this theory properties have their powers essentially. Hence, CTP guarantees that whenever a whole-property is instantiated its part-property is instantiated with metaphysical necessity.

I will now present The Individuation Argument, which shows that Shoemaker's parthood relation prevents the possibility of non-causal features of higher-level properties. As I mentioned before, this is not in itself an argument against his view, rather an argument that shows that his view is unable to account for a certain kind of properties. The argument presupposes that both A1 and A2 hold, and shows that Shoemaker's parthood relation requires that everything that is metaphysically necessary for higher-level properties is metaphysically necessary for their realizers. To do so, I will put A1, A2, and the conclusion of Shoemaker's argument to work in a particular case. Let us consider that it is theoretically possible that a given mental property has an essential non-causal feature, such as the phenomenal aspect of pain. According to Shoemaker, pain is a higher-level property realized by a physical property, say P. Given the conclusion of Shoemaker's argument, pain is a part of its physical realizer P, and given A1, pain is metaphysically necessary for P. Now, given A2, anything metaphysically necessary for pain must be metaphysically necessary for the realizer of pain. I have started by saying that pain not only has a causal profile, but it might also have a non-causal phenomenal feature. Then, given A2, if there is a phenomenal aspect necessary for pain, it would also be necessary for its physical realizer. So there cannot be any phenomenal feature, or any feature for that matter, that is metaphysically necessary for pain but not metaphysically necessary for the realizer of pain. This argument shows that actually the parthood relation prevents some of the theoretical possibilities (i.e., properties having essential non-causal aspects) Shoemaker actually intended to allow by providing a weaker version of the CTP.

Finally, I am now in a position to argue in favour of Wilson's Neutrality Theses. To start with, I have shown that Shoemaker's property individuation* is a problematic notion. I have provided The Metaphysical Necessity Argument that shows that he needs to commit to properties bestowing their powers with metaphysical necessity. One way to guarantee this is

by endorsing the CTP. This discussion is also supported by the fact that most reviews of Shoemaker's account take his mereological realization relation to be directly related to his Causal Theory of Properties, although he explicitly denies so (for instance, Wilson, 2011, pp. 132 -133; Morris, 2011, p. 366; Pineda and Vicente, 2016, p. 101). If CTP is accepted, then properties bestow their powers essentially, so The Metaphysical Necessity Argument is deactivated, and the parthood relation is guaranteed. This in itself is a reason to prefer Wilson's account. She does not need to endorse or defend such a controversial theory as the CTP to accomplish the same goals, i.e. providing a realization relation in terms of a subset condition on powers, because she does not require a parthood relation between realizer and realizee. In section 2.1 I have explained The Proportionality Argument and The Causal Joints Argument that precisely show that higher-level causal efficacy can be defended without appealing to a parthood relation. If this were not so, then proving that properties have their causal powers essentially might be worth the effort, and Shoemaker's theory should be preferred. But given that it is not, I think that this alone shows that Wilson's view is clearly more appealing. There is a further reason to favour Wilson's view. If Shoemaker requires properties to be individuated by their causal powers in such a strong sense, then he cannot accommodate the possibility of realized properties that essentially have or can be individuated by non-causal features not had by their realizers. By contrast, given that Wilson only needs to accept that properties have the same powers in relation to the same laws of nature, she has no problem in accommodating such features of properties.

There is one last difference between Wilson's view and Shoemaker's worth discussing. Shoemaker's view, for physicalist reasons I will not delve on here, requires that the realizer must be metaphysically sufficient for the realisee, rather than nomologically sufficient as Wilson does. So following the example, something being scarlet is metaphysically sufficient for it being red. This condition complements A1, which stated that in cases of mereological realisation having the part-property is metaphysically necessary for having the whole property. However, I believe that requiring metaphysical sufficiency without endorsing CTP is problematic. Unless one accepts the Causal Theory of Properties, that is that properties have their powers essentially, it is very difficult to see how realizers alone are metaphysically sufficient for their realisees. Let us show how the CTP accommodates the claim that the realizer must be metaphysically sufficient for the realisee. If one accepts the Causal Theory of Properties then one holds that powers are essential to properties, so for any property, say P, to be instantiated in a possible world, say *w*, P must bestow the same causal powers as in the

actual world. From this alone it follows that the instantiation of P is sufficient for the instantiation of its realisee, say M, because in any given world P will have the same causal powers necessary for M to be instantiated.

However, if one holds that properties bestow powers with nomological necessity (as Property Individuation* does), in a possible world *w*, P may bestow different powers, since it is possible that those related to M are lacking and, therefore, M is not instantiated. Hence, the instantiation of P is not metaphysically sufficient for M (Shoemaker, 2007, p. 6). This means that property individuation* plus parthood with nomological necessity is not sufficient for realization, as they cannot guarantee that the realiser is metaphysically sufficient for the instantiation of the realisee. It should be noticed that Shoemaker suggests that there is another way to argue for the sufficiency claim that does not require accepting the Causal Theory of Properties. What he suggests is that in this case one needs to take “the realizer to include the causal laws that hold in the worlds in which the realization relation exists” (Shoemaker, 2007, p.14). It is striking that he does not provide further details about how this might be done. How can a realizer include such laws? Would it not make more sense to say, with Wilson, that the realizer is sufficient for the realisee only in those worlds with the same laws as our world, that is, with nomological sufficiency? If that was the case, though, it would be difficult to argue that there is a parthood relation between realiser and realisee. Suppose that nomological necessity is sufficient for parthood. If that were so, we could end up with absurd cases such as the following. Let us consider the property of being an alive fish and the property of being wet. Given the laws of nature fish must be in an environment with water to survive in our world, so being wet is nomologically necessary for the property of being an alive fish. However, it does not make sense to say that being wet is in part being an alive fish. Therefore, Shoemaker needs a stronger relation between part-properties and whole-properties. As long as it is not clear how this stronger relation is achieved, this is yet another reason to favour Wilson’s account. I do not have space to pursue this further here, although see McLaughlin (2007) for a development of this worry.

3.2 Against Shoemaker’s Microphysical Realization (I)

As I have mentioned above, Shoemaker’s realization relation has two parts. In the previous section I have focused on Shoemaker’s account of property realization. However, he argues that property realization is not sufficient to provide a complete non-reductive physicalist

account but that a further realization relation must be posited —i.e. microphysical realization. His argument in favour of microphysical realization has two premises. The first one is the following. Property realization can only account for the realisation of properties by realisers that are properties of macroscopic objects, but we should be able to account for the metaphysical dependence relation between the many properties of constituents and a property of the composite.

The subset view and the standard higher-order property account have been held to be unsatisfactory because they limit property-realizers to properties of macroscopic objects, and normally to properties of the very macroscopic object that has the realized property, and make no room for realizers that are properties of constituents of the things having the realized properties. (Shoemaker, 2007, p. 32)

The second premise has to do with Shoemaker’s account of physicalism: microphysical facts fix all the facts. In other words, all God had to do was to create microphysical entities propertied and distributed in a certain way, and the laws that govern them. The second premise can be found in the following quote, where “states of affairs”⁵ is used to refer to ways things are in the world, and in general states of affairs are constituted by entities propertied and related in certain ways (a property instance is a special case of this, but microphysical realisation focuses on those states of affairs that are not property instances themselves, but that contain many property instances):

If physicalism is true, all of the facts are determined, constitutively, by the microphysical facts—by how fundamental physical micro-entities are distributed in the world, and how they are, as I will say, propertied and related. And this means that instantiations of properties in macroscopic entities will be realized in, will have as realizers, microphysical states of affairs. (Shoemaker, 2007, p. 3)

If microphysical states of affairs fix all the states of affairs (I will use “states of affairs” from now on instead of “facts”) there are, and property realization cannot account for the realization

⁵ Shoemaker prefers to use the term “state of affairs” because there is a sense of the term “fact” that takes facts to depend in some way on our knowledge about the world, i.e. “Superman is a superhero” expresses a different fact than “Clark Kent is a superhero”, while the state of affairs —namely the referent of those propositions— that would make them true is the same (Shoemaker, 2007, p.34).

relation needed between properties and microphysical states of affairs —as property realization only deals with realizers that are properties—, then an extra relation is needed. That is, we need a realization relation that can account for how the properties of macroscopic entities are realized in microphysical states of affairs, which to reiterate consist of physical entities propertied and related in a certain way. So following this argument it seems that microphysical realization is necessary for a successful non-reductive physicalist account. If this is so, then *prima facie* Wilson’s theory would be lacking an essential insight. I will not challenge here the argument itself or its premises, because I will do so in the next section, 3.2.3. Instead, I will explain how microphysical realization works, focusing on its role in the realization of mental properties. That will facilitate the discussion about the extent to which Wilson’s theory might be incomplete without a microphysical realization relation.

Before explaining what it is for a state of affairs to realize a property it will be helpful to introduce the overall picture Shoemaker is proposing, which is shown in Figure 2. Every property is realized in a lower-level property, and in turn this lower-level property is realized in a property at an even lower level, until we reach the maximally determinate property that is not further realized in any property, but only in a maximally determinate state of affairs. Shoemaker calls such properties microphysical-state-of-affairs-embedding properties, or MSE properties for short. In turn and given the physicalist commitment that microphysical states of affairs fix all the states of affairs there are, every property is not only realised in a lower-level property but is also microphysically realized in a more or less abstract microphysical state of affairs. Although these more and less abstract states of affairs also microphysically realise corresponding MSE properties, which serve as the base for any property-realisation that state of affairs has, Shoemaker reserves the term “MSE” to those properties that are microphysically realised in maximally determinate state of affairs.⁶ As I have said, MSE properties are the ones realized in maximally determinate states of affairs, so properties of this kind will have the least abstract microphysical realizers.

⁶ “for every type of microphysical state of affairs whose members can realize a property instantiation there will be a property something has just in case its career embeds a state of affairs of this sort. One can speak of these as MSE properties, for microphysical-state-of-affairs-embedding properties. But here I will reserve the term “MSE property” for cases in which the embedded state of affairs is maximally determinate.” (Shoemaker, 2011, p.5)

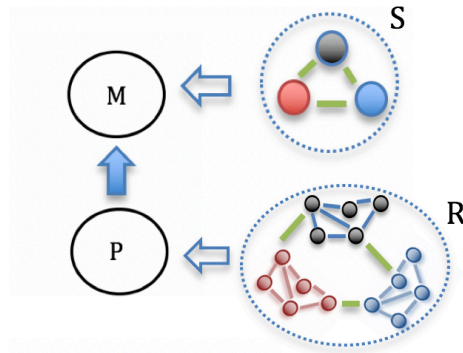


Figure 2. Representation of Shoemaker's view on property realization and microphysical realization.

In Figure 2, the dots represent entities, and the straight lines represent relations between entities. Entities have different colours depending on their properties. Property P realizes (blue arrow) higher-level property M. In turn, state of affairs S microphysically realizes (arrow with blue contour) property M, and state of affairs R microphysically realizes property P. This is actually a crucial and, in my understanding, controversial feature of Shoemaker's view: there must be different microphysical realisers for P and M because property-realisation of M by P requires M to have a subset of the powers had by P. If P and M were realised by the same microphysical state of affairs, given that states of affairs have matching powers with the properties they realise —this is thoroughly explained in the next page—, satisfaction of the subset condition on powers would be impossible. At the same time, Shoemaker is committed to the idea that a microphysical state of affairs that realises P must also be the microphysical realiser of M. At this point it is helpful to introduce Shoemaker's example:

Suppose again that pain [M] is realized by C-fiber firing [P]. This formulation allows an instance of pain to have as a realizer a state of affairs [R] that is a microphysical realizer of an instance of C-fiber firing, and so is of a type whose causal profile matches that of the property of C-fiber firing rather than that of the property of being in pain. But it will also have a microphysical realizer [S] that is of a state of affairs type, a more abstract one, whose causal profile matches that of the property of being in pain. Such more abstract states of affairs will be realized by more concrete ones. (Shoemaker, 2011, p. 6).

State of affairs S is an abstraction of state of affairs R (represented in the figure by the fact that S has three entities that are an abstraction from each of the bundles of entities in R). In his 2007, Shoemaker has a slightly different picture, where a concrete state of affairs has different

parts each being the “concrete core” of the realiser of a different property. So all properties of an object are virtually microrealised by the same state of affairs but by different parts of it. For instance, the concrete state of affairs that realises scarlet, will also be the microrealiser of red, but given that red and scarlet have distinct causal profiles, they are realised by different concrete cores.

Putting together the parts that contribute directly to the instantiation of a property, we get what I call the concrete core of the realizer of the instance of that property. So one part of the concrete state of affairs will be the concrete core of the realizer of the instance of P, another will be the concrete core of the realizer of the instance of Q, and so on. (Shoemaker, 2007, p.43)

The notions of “abstraction” and “concrete cores” are actually quite different from each other, although their goal is the same: to explain how P and M can have different causal profiles, even if they are realised by the same microphysical state of affairs. Ultimately, in Shoemaker’s picture, every single property is realised by the same maximally determinate global state of affairs. The notions of abstraction and “concrete cores” are not sufficiently justified at this point. So let me move on to introducing how microphysical realisation explains the metaphysical dependence relation between states of affairs and the properties they microphysically realise. One last detail is that the dotted circles around R and S represent microphysical-state-of-affairs-embedding properties, the one corresponding to S being more abstract than the one corresponding to R.

Shoemaker explicitly claims that for an account of microphysical realization to get off the ground we should accept that properties are individuated by their causal profiles. This is against Wilson’s Neutrality Theses, so if this relation really was needed to hold between properties and states of affairs, then one should reject the neutrality thesis that properties need not be individuated by their powers. The core of microphysical realization is that the causal profile of the realized property “corresponds to” or is “isomorphic with” the causal profile of the microphysical state of affairs (Shoemaker, 2007). Shoemaker has proposed slightly different versions of what this correspondence relation means, which he has also referred to as causal profiles “matching” in some way (Shoemaker, 2011, 2013). However, what all those different phrasings mean is that a given microphysical state of affairs realizes a property if that state of affairs is embedded in the career of the same thing in which the property is instantiated and it makes, with metaphysical necessity, the same causal contribution as the property it micro-

realizes. To put it in a more visual way let me introduce an example, such as the property of being circular. The property of being circular will be micro-realized by a member of a type of state of affairs which consists of entities related in such a way that they are all at the same distance, say R , from a given point. Whatever a type of state of affairs consisting of entities related in this way is able to cause, the property of being circular will also be able to cause. In this way, the instantiation of a microphysical state of affairs constitutes the instantiation of the property it realizes. *Prima facie* it is odd to say that the state of affairs type consisting of a collection of entities related such that any entity is at distance R from a given point has the same causal powers as the property of being circular, given that such a state of affairs would be able to cause many more things than what the property of being circular might cause. For instance, considering that such entities have sufficient mass, they would be able to attract other entities, while the property of being circular does not bestow this causal power. Shoemaker's (2007) way of dealing with this is to argue that microphysical states of affairs have cores, and those cores involve only the entities and properties of those entities that are involved in the powers of the realized property. In our case, having mass is a property of the entities making up the state of affairs that is not relevant to the powers of the property being circular. So the core of the microphysical state of affairs that realizes the property of being circular does not include the property of having mass, nor the powers this property endows. In Shoemaker's terms, states of affairs realizing different properties have different degrees of abstractness in virtue of being made up only of the entities and relations relevant to the causal profile of the realized property. This is illustrated in Figure 2, as S is depicted as an abstraction of R . This is going to be relevant to the discussion below about how microphysical realization deals with Kim's causal exclusion argument.

Given that all properties are microphysically realized in microphysical states of affairs, a version of Kim's argument reappears. Let us start by considering an object, say a table. This object is constituted by a microphysical state of affairs, and it has several properties, some of which are independent from each other. For instance, it has the property of being white, of having a rectangular shape, of being made of wood, etc. It could be said that all those properties are microphysically realized in one microphysical states of affairs, the one that constitutes the table. However, this would entail that they all have the same microphysical realizer. If that were so, given that the causal profiles of states of affairs must match the causal profiles of the properties they realise, then it would be difficult to say on what grounds those properties are different. In general, consider the microphysical states of affairs in Figure 2, S and R , and the

properties they realize, M and P accordingly. If the microphysical state of affairs, S, realizing a given higher-level property, M, was identical to the microphysical state of affairs, R, realizing the lower-level property, P, that realizes M, then it could be argued that the instance of M just is the instance of P. This is because given that microphysical realization is based on the causal profile of the realizer matching the causal profile of the realizee, then M would have the same causal powers as P, not a subset of them, which is the only source we have to differentiate both instances. Kim's threat would now be framed in terms of there being different properties but identical instances, so the only way for M to be causally efficacious would be in virtue of being identical to P (Shoemaker, 2007, p. 46).

Hence, Shoemaker needs to make the case that the microphysical states of affairs realizing the different properties of an object are different, and in particular that the microphysical state of affairs realizing the mental property M is different from the microphysical state of affairs realizing the physical property P. This is of course easy given the definition of microphysical realization and is related to the discussion above about the property of being circular. The causal profile of a given state of affairs matches the causal profile of the property it realizes. So given that S microphysically realizes M, and R microphysically realizes P, the causal profile of S matches the causal profile of M, and the causal profile of R matches the causal profile of P. Given that the causal profiles of P and M are different, so are the causal profiles of S and R. There are as many states of affairs as properties of an object. But how can so many different states of affairs be part of a single object? Shoemaker's answer to this has actually been sketched above regarding how the microphysical state of affairs realizing the property of being circular only has the causal powers associated with the property of being circular, and not those powers associated with the property of having mass: the core of the state of affairs realizing P is different from the core of the state of affairs realizing M. However, this case is slightly different from the one involving the property of being circular or the one involving the different properties of a table: P and M are not independent properties but P is M's realizer, so the causal profile of M is a subset of the causal powers of P. There is, then, a parthood relation between the cores of the microphysical states of affairs, R and S, realizing P and M accordingly.

“[The causal work of pain] is indeed done by the C-fibre stimulation instance, but this is because the core of the C-fibre stimulation instance realizer has the core of the pain instance realizer as a part. It is only because the C-fibre stimulation instance realizer contains the pain instance realizer that it has the relevant effects.” (Shoemaker, 2007, p. 48)

In this way Shoemaker is able to harmonise his view on microphysical realization with his account of physical realization.

At this point, a question arises: what does Shoemaker's microphysical realization add to the subset view's solution to the Causal Exclusion Problem? If it were to add something that cannot be achieved by means of the subset view, then Wilson's view would clearly lack a crucial insight. However, I think that from our discussion so far it is clear that it does not add much. On the contrary, microphysical realization is defined in such a way that it avoids any interference with the solution provided by the subset view. At most, we could say that it provides a different way, one in terms of microphysical realizers, to put the claim that pain is part of its realizer. Remember that the two main reasons why Shoemaker introduces microphysical realization are 1) to account for property realization where the realizers are not macroscopic properties, and 2) to deal with his definition of physicalism in terms of microphysical facts. However, neither of the two is directly related to, or affects, the Subset View's solution to Kim's Causal Exclusion Argument. On the one hand, Kim's argument is framed in terms of physically acceptable properties, such as c-fibre firing, that nonetheless are clearly not microphysical properties. This means that there is no need to introduce microphysical states of affairs to propose a realization relation able to deal with Kim's argument. Moreover, Shoemaker's definition of physicalism in terms of microphysical states of affairs is not the one used in Kim's Causal Exclusion Argument, who uses Physical Causal Closure (for instance, see Kim, 2005, p. 15). Of course, it might turn out that Shoemaker's definition is preferable. However, this is another discussion that, although related, does not affect Wilson's solution to Kim's argument as it stands. Hence, Wilson's subset view is rich enough to solve Kim's argument without appealing to microphysical realization. This has as a consequence that Wilson's Neutrality Theses, in particular the one about property individuation in terms of powers that was denied in Shoemaker's microphysical realization, do not affect her solution to Kim's problem.

So far, I have argued that microphysical realization does not add anything to Wilson's solution to Kim's problem. However, it is not straightforward how the subset view can account for realization relations where property realizers are properties of constituents of the things having the realized properties. Given that Wilson wants her view to apply not only to mental properties but to special science properties as well, and at least some special science properties are realized

by properties of constituent entities (for instance, the properties of molecules are realized in the properties of the atoms that compose them), then the scope of the subset view hinges on whether the many-one structure of realisation can be subsumed under the one-one structure that characterises the subset strategy. Both Shoemaker and Gillett argue that the property realisation in place in the subset strategy cannot achieve this. In Chapter 4 I introduce Gillett's argument that concludes that realisation, in the case of the realisation of special science properties, must have a many-one structure. The argument is presented as an objection to the subset view. But before that, I discuss whether Shoemaker's positing of a new realisation relation —microphysical realisation— is truly justified.

3.3 Against Shoemaker's Microphysical Realization (II)

Remember that Shoemaker introduces his microphysical realization relation for two main reasons: 1) to “make room for realizers that are properties of constituents of the things having the realized properties”, and 2) to preserve physicalism. I will start by discussing the first reason. Shoemaker accepts the need for a realisation relation that has a many-one structure, where the relata are not single properties at both ends, but many properties of components at the lower level, and a single property of the composite at the higher level. However, I question that Shoemaker's microphysical realisation is making room for realisers that are properties of constituents in a way not already present in the subset strategy. Gillett (2016, p. 80) mentions the exegetical difficulty of making sense of whether Shoemaker's microphysical realization relation is a one-one or many-one relation. As I explained in the previous section, but I need now to explain in a bit more detail, Shoemaker takes microphysical realization to be a realization relation between a state of affairs and a property. States of affairs consist of entities propertied and related in a certain way, in Shoemaker's own words:

As I conceive of states of affairs, the existence of a state of affairs simply consists in some proposition's being true. It can consist in particular items in the world having certain properties and standing in certain relations, or in the truth of some quantified proposition about how things are, or in some negative truth about how things are, or in some combination of these. (...) The microphysical states of affairs that primarily concern me will have particular micro-entities as constituents. But the specification of such a microphysical state of affairs can also include positive and negative existential propositions. (Shoemaker, 2007, p. 33-34)

So a state of affairs just is a way things are, including both positive and negative existential propositions. Shoemaker claims that every property there is has a microphysical realiser. As we can see in Figure 2 above, this means that middleman properties, such as P, also have microphysical realisers, in this case R. State of affairs R, which consists in some lower-level entities propertied and related in a certain way, realises middleman property P insofar as their causal profiles match in the appropriate way. At the same time, microphysical realization also holds between an abstraction, S, of those same lower-level entities propertied and related in a certain way (or a concrete core of that state of affairs) and higher-level property M, where M is also property-realized by P. Even more, given that S is just an abstraction from R, Shoemaker accepts that microphysical realization also holds between the core of the more determinate state of affairs consisting in some lower-level entities propertied and related in a certain way, R, and M. In this way, he can preserve the subset relation between M and P, in virtue of the microphysical realisation of M by R, where R has more powers than M. In conclusion, property M is realised by P and microphysically realised by S and by R. The ontology resulting from Shoemaker's picture is definitely crowded.

That said, is this crowded ontology justified by its ability to make room for realisers that are properties of constituents? Does microphysical realisation do something that the subset strategy cannot achieve? It is true that in a derived or indirect sense Shoemaker is explaining how many lower-level properties of constituents realise a single higher-level one: they do so in virtue of being part of a state of affairs that is able to realise the higher-level property. However, what he is really making room for is for realisers (states of affairs) that are constituted by the constituents, not realisers that are themselves the constituents. A note should be made here. It is true that one might understand states of affairs as "consisting of" rather than as "being constituted by" many lower-level entities/properties and their relations. As Gillett points out, there is an exegetical difficulty in understanding the nature of Shoemaker's relation: the former reading supports the many-one structure of microphysical realisation, while the latter supports a one-one reading. However, there is one feature of the relation that I believe supports the "being constituted by" or one-one reading: the isomorphism of powers between the realised property and the realiser state of affairs. Shoemaker's proposal uses the causal profile of a state of affairs to determine the properties it realises. The causal profile of the state of affairs is a function of, or is derived from, a combination of the causal profiles of the individual entities constituting it and their relations. Those are exactly the kind of causal profiles Wilson attributes

to middleman features.⁷ The fact that Shoemaker talks of states of affairs having powers, as a whole, brings him closer to a one-one picture. If this is what is required to subsume a many-one structure, then the subset strategy is also able to achieve that by means of a middleman property instead of a realisation relation. In both cases, the realiser is constituted by the lower-level non-relational entities/properties related in a certain way. All things considered, Shoemaker's crowded ontology of properties and realisation relations does not achieve anything substantially different than Wilson's middleman strategy. I will briefly come back to this in section 4.3.

Let me move to the second reason why Shoemaker considers microphysical realisation is needed. He introduces the relation to preserve physicalism: he argues that given that all of the facts are determined by the microphysical facts, then properties of macroscopic entities must be realized in microphysical states of affairs. However, we can argue that microphysical facts fix all the facts without appealing to microphysical realization: microphysical states of affairs are constituted by fundamental micro-entities propertied and related in certain "ontologically lightweight" ways, so they are just aggregates of entities that have certain features, i.e. Wilson's middleman properties. Those features are not physically problematic, because they are features of fundamental entities aggregated in non-problematic ways, or in Wilson's terms "ontologically lightweight". We can then argue that any property of a macroscopic entity will be ultimately realized in such physical features of a microphysical state of affairs. Therefore, we can still claim that microphysical facts fix all the facts, without postulating another realization relation between every macroscopic feature and microphysical states of affairs. Microphysical realization was introduced for two reasons. Regarding the first, I have shown that it is not clearly doing the work it was supposed to do, regarding the second, I have shown that it is actually not needed. Hence, the reasons Shoemaker provides to introduce a new realization relation are unconvincing.

⁷ I discuss Wilson's understanding of middleman features in more detail in Section 4.3.

CHAPTER 4. CARL GILLETT'S ACCOUNT OF REALISATION AS A COMPOSITIONAL RELATION

4.0 Introduction

Carl Gillett started developing his theory of realisation in the early 2000's, and he has written several papers in which the core ideas of the theory can be found. Nonetheless, it is in his 2016 book *Emergence and Reduction in Science and Philosophy* that we find the most advanced and thorough version of the account. For this reason, I will mainly refer to this book. There are mainly two reasons why I consider that Carl Gillett's account of realization is a key theory to put in dialogue with Wilson's Weak Emergentism. First, Gillett provides an account of metaphysical dependence that competes with the subset view of realization. Gillett's account of the dependence relation is stated in terms of composition and is similar to Shoemaker's microphysical realization. Moreover, he defends his account explicitly against the Subset View. Second, Gillett uses his account of composition to introduce a downward determinative relation between higher-level entities and their components, thus proposing a new version of physicalist emergentism, so-called Mutualism or scientific emergentism. Mutualism presents a different version of physicalist emergentism than Wilson's because higher-level features are not causally autonomous in virtue of having a distinct causal profile, but they are causally autonomous because they can affect—I discuss in due time how—their lower-level components.

4.1 Gillett's account

Before discussing Gillett's view, some preliminary remarks are needed. Gillett's view is conceptually powerful and resourceful, and inevitably this means that the definitions of some key notions are not straightforward. Two key concepts are “determination” and “production”. These concepts are used to provide a more fine-grained classification of the processes by which entities⁸ can affect each other. An entity is determinative if “it makes a difference to the powers of individuals”. However, as I will explain below, Gillett's view suggests that entities may make a difference to the powers of individuals without themselves exercising a power. That is,

⁸ Gillett uses the term “entity” to refer to properties, individuals, powers and processes. So far in this thesis, however, the term has been used to refer only to beings such as particles and objects. From now on I will use the term in Gillett's sense, unless I state otherwise.

determination need not be a causal process. When an entity makes a difference to the powers of an individual by indeed exercising a power, Gillett calls this process “production”. Production involves “the manifestation of the powers of an individual to result in an effect”. Gillett prefers to use the term “production” rather than “causation” on pragmatic grounds: given that it is a much more delimited term, it does not evoke the debates and controversies around causation. However, in the framework of this thesis, production equals causation. Productive processes are a subset of determinative processes: every productive process is determinative since every productive process makes a difference to the powers of individuals. However, not all determinative processes are productive (causal).

Importantly, Gillett claims that the crucial feature of any ontologically robust entity is being determinative, rather than productive. This contrasts with the usual formulation of Alexander’s Dictum I introduced in Chapter 1: to be real is to have causal powers (or to be causally efficacious). Gillett is much less strict when it comes to what it takes to be real. We can see this in his atypical formulations of the Eleatic Principle (i.e., Alexander’s Dictum) and of Physical Causal Closure (also known as the Completeness of Physics):

(Eleatic Principle) The only entities that exist make a difference to the powers of individuals.
(Gillett, 2016, p. 108)

(Completeness of Physics) Microphysical entities are determined, in so far as they are determined, solely by other microphysical entities according to the laws covering their behaviour in isolation or the simplest collectives. (Gillett, 2016, p. 117)

The former is atypical in that the Eleatic Principle is commonly phrased in terms of entities themselves having causal powers, rather than in terms of entities making a difference to the powers of individuals. In this definition Gillett is clearly appealing to the determination/production distinction: whether an entity exists is determined by its determinative role, not by its causal powers. For the moment, it is vague how an entity could make a difference to the powers of individuals without exercising any of its causal powers, but I will come back to this below when I further explain Gillett’s account. The latter is, again, an unusual formulation of the Completeness of Physics. In Chapter 1 I already pointed out that there is not a single undisputed formulation of this principle. There I introduced Kim’s own definition –the most common one, I should say– on which his Causal Exclusion argument

depends. Also, I reformulated it so that it explicitly referred to causal powers, rather than causation. However, in Gillett's formulation there is no reference to powers or causation. Again, this is because he is using the notion of determination rather than causation to formulate the principle. Finally, the distinction between determination and production has another important field of application: Gillett's understanding of the laws of nature. Although usually laws of nature are related to causal powers, and in fact it is in this way that I have been discussing them in the previous chapters, Gillett holds that where entities are determinative there should be natural laws referring to them. Of course, we have the trivial case of entities that are determinative in virtue of being productive. Moreover, Gillett introduces the controversial case where entities that are determinative yet not productive require natural laws referring to them. Those natural laws are fundamental, in the sense that we would not be able to get a complete description of these entities and the processes they enter in, without appeal to these laws.

I say that his view is controversial because Gillett does not provide any definitive case studies where an entity is determinative in this way, although the view is *prima facie* theoretically compelling. He does try to explore some case studies, but they are inconclusive: "at the end of a long book, I have not produced a definitive conclusion." (Gillett, 2016, p. 353). Throughout this chapter and the two case studies that follow it, I assess the theoretical coherence of his view, but most importantly, I explore empirical evidence that could potentially support or refute his view. Despite Gillett's self-admitted lack of definitive conclusions, the framework he provides can be of substantial value when applied to diverse scientific case studies. The inconclusive nature of his case studies does not inherently undermine his theory, but instead highlights the complexity of the subject matter and the need for more comprehensive studies that bridge the gap between philosophical and empirical research. Therefore, as I delve deeper into these cases, the aim will be to illuminate the intricacies of his arguments—as well as Wilson's—, evaluate their applicability, and determine whether they can withstand empirical scrutiny, ultimately contributing to the ongoing debate about physicalist emergence.

A final stop before explaining Gillett's account is needed to state his view on the relation between properties and the causal powers they have. Surprisingly enough, Gillett's stance on powers is exactly the same as Shoemaker's weaker version of CTP, as he makes explicit in the following quote: "[I] use Shoemaker's (1980) account of properties/relations under which what I term a "powerful" property or relation is individuated by the powers it contributes to

individuals in this world under certain conditions.” (Gillett, 2016, p. 63). However, in this case Gillett’s stance on powers will not invite the objections I raised to Shoemaker’s account because Gillett’s metaphysical dependence relation does not involve parthood in the relevant way.

Now I am ready to introduce Gillett’s view of the metaphysical dependence relation between higher-level entities and their lower-level constituents. To start with, one key distinction from the subset view is that Gillett grounds his account on the notion of composition, which holds between many entities and a single composite —namely, it has a many-one structure. He asserts this kind of many-one structure is the most common in scientists’ work, rather than a one-one relation between a lower-level relational property and a higher-level property, which characterises the subset view. One of the main motivations of Gillett’s view, which he also uses as a justification for it, is the idea that he is taking scientific notions and working intuitions and conceptualising them to provide a philosophical theory. Thus, he claims that his view is grounded in the scientific understanding of nature. In this vein, he claims that the composition relation present in the sciences can be defined as a non-productive (non-causal) determinative relation between many working lower-level features (working features are those individuated by their roles, i.e. powers) and a higher-level feature. So instead of a one-one relation, Gillett takes as a starting point a many-one relation. So higher-level entities are composed by lower-level components. As Figure 1 shows, there is a plurality of composition relations, depending on the nature of the composed entities. Interestingly, Gillett refers to the relation between lower-level non-relational properties and higher-level properties as realization. In this case realization is a many-one relation. This difference between Wilson’s and Gillett’s accounts will become central to this thesis from the next section onwards. Most of the subsequent chapters of this thesis will question whether a many-one account should be preferred over a one-one account. But before that, I continue introducing Gillett’s account.

Type of Entity	Compositional Relation
Powers	Lower-level powers together <i>comprise</i> higher-level powers
Properties	Lower-level properties together <i>realize</i> higher-level properties
Individuals	Lower-level individuals together <i>constitute</i> higher-level individuals
Processes	Lower-level processes together <i>implement</i> higher-level processes

Figure 1. Table with the different compositional relations and the kinds of entities they relate. Notice the term “together” before each compositional relation, which refers to their many-one nature. Extracted from Gillett (2016, p. 68).

How can we explain that lower-level entities *together* compose higher-level ones? Gillett answer to this question requires introducing a new notion: “joint role-filling”, which contrasts with the more usual notion of “role-playing”. According to Gillett, when in certain spatial and temporal relations and under certain circumstances, working components jointly fill the role of the higher-level entity (be it a property, power, process or individual). The structure of “joint role-filling” captures the many-one character typical of compositional relations. Role-playing, on the other hand, is a one-one relation in which the lower-level entity plays the causal role of its realizee, so each causal contribution of the higher-level entity makes is inherited from and identical to a causal contribution made by its lower-level realiser. Role-playing is widely used in the non-reductionist physicalist literature, for instance in functional accounts and subset accounts. By the end of this chapter I will thoroughly discuss the arguments Gillett provides to favour joint role-filling over role-playing, so I just mention a couple of them very briefly here. First, joint role-filling is commonplace in the sciences. So even if role-playing has been successful so far in providing accounts of realisation of special science entities, it is problematic that scientific accounts of such entities display many-one relations. Second, Gillett claims that role-playing is unable to capture a key trait of scientific composed entities: that they are qualitatively different from their components. By “qualitatively different” Gillett means that the higher-level entities do not share properties nor powers with the component entities that realise them. This will become clear when I introduce the diamond example in the next paragraph. At this point I want to note that neither qualitative difference nor joint role-filling per se entail anti-physicalism, as I explain later in this section.

Now that the notion of joint role-filling is in place, it is easy to understand Gillett’s view of how individuals at a higher level are composed by a set of individuals at a lower-level: just as

a Lego house is composed by Lego bricks, the lower-level individuals together compose the higher-level individual. However, joint role-filling also applies to properties, powers and processes. How composition relations work in those cases, and in particular between properties of the higher-level individual and properties of many lower-level individuals, is not straightforward. To justify and explain composition between properties, Gillett provides the following argument based on a particular case: the realization of a property of diamonds, that of being hard, by the many lower-level properties of its components. This is my reconstruction of the premises and conclusion of his argument:

1. Scientifically we can explain how several microscopic processes compose the process of a diamond scratching glass.
2. A process simply consists in “a power of some individual being triggered and manifested to produce an effect” (Gillett, 2016, p. 64).
3. Therefore, each microscopic process corresponds to the manifestation of the powers of a carbon atom, and the macroscopic process corresponds to the manifestation of a power of the diamond.
4. Composition between processes entails composition between powers: the power of the diamond to scratch glass is comprised by the powers of carbon atoms.
5. Natural properties are individuated by their powers.
6. Therefore, the composition relation between powers entails a composition relation between properties.

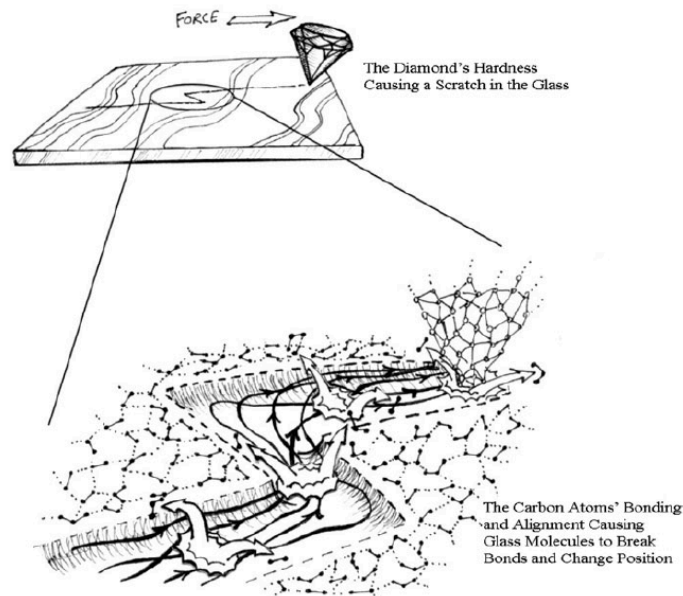


Figure 2. Figure showing the process of scratching a glass at two different levels: processes involving carbon atoms together implement the process of scratching glass at the diamond level. Figure extracted from Gillett, 2010, p. 17.

The upshot is that the properties of the carbon atoms, in particular the property of forming strong bonds with the neighbouring atoms in the lattice, *jointly* realize the properties of the diamond, in this case the property of being hard. Gillett takes as a starting point the fact that scientists mechanistically explain higher-level processes by identifying them with an array of lower-level processes and concludes that lower-level properties of the carbon atoms together realize the properties of the diamond (Gillett, 2010, p. 171). It is not the aim of this chapter to assess the premises of this argument, nor the validity of the conclusion. However, I should mention that even if all the premises were accepted as true, one could challenge the logical jump from “composition between powers” to “composition between properties”. While it may seem intuitive, it's not logically necessary. It presupposes that powers and properties are so tightly connected that because powers are composed in a certain way, properties — since they are individuated by those powers — must also be composed in the same way, which may not necessarily be true. Still, let's assume for now that the conclusion follows from the premises.

So far I have not discussed whether those compositional relations involve emergence of any kind. According to Gillett, compositional relations are compatible with both reduction and emergence. Let us discuss Figure 4 in order to clarify what has been said so far, and also to introduce Gillett's physicalist emergentist view. This figure shows a picture that resembles in

some ways the views discussed in the previous chapters —Wilson’s and Shoemaker’s— but differs from them in some important ways. Crucially, the white upwards arrow from P1-Pn to G does not represent a one-one relation between a lower-level property and a higher-level property, but a many-one composition relation between n lower-level entities that jointly fill the role of the entity they compose. So G is a higher-level qualitatively different entity composed by n component entities. So far, this description matches with the features of Gillett’s view I have introduced in this chapter.

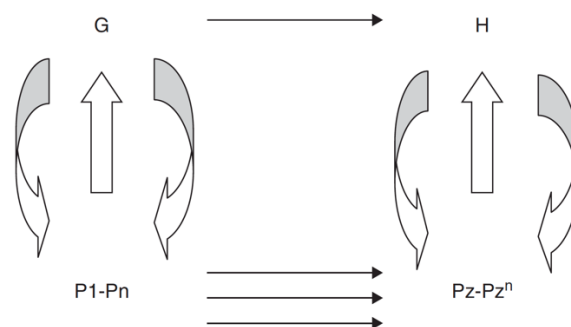


Figure 4. Gillett’s structure of relations between microphysical components and composed features. Upwards arrows represent composition, while downward arrows represent machresis. Horizontal thin arrows represent causation. Figure extracted from Gillett (2016, p. 241)

But what about those curved downward arrows? According to Gillett, entities coming together may give rise to a new determinative thus emergent entity depending on the nature of aggregation (aggregation can take place in two radically different ways, one that leads to emergence and one that does not, as I will explain shortly). That is actually the case in Figure 4: what the curved downward arrows represent is the determinative (yet non-productive) power of the higher-level emergent entity G, and the arrow from G to H represents a determinative and, in some instances, productive power (remember that as per Gillett’s formulation of the Eleatic Principle the fact that the emergent feature is determinative is a necessary condition for its genuine existence). The determinative power of G is therefore twofold: on the one hand, it productively or non-productively determines entities at its same level; on the other hand, it non-productively determines its own components. However, not all cases of composition lead to emergent features. Compositional relations and joint role-filling are also compatible with reduction. What then causes some compositional relations to give rise to emergent features? There are two different ways lower-level components could aggregate, one that leads to reductionism, and another that leads to emergentism. The former Gillett calls The Simple View of Aggregation, and the latter he calls The Conditioned View of Aggregation.

According to the Simple View of Aggregation, aggregation is continuous. Composed entities are not determinative, and the powers of the components when isolated or when in a collective follow the same laws (Gillett, 2016, p.118). An example could be helpful here. Take two entities with the same mass, m , that aggregate to compose an entity with mass $2m$. The laws governing the aspects of the entities that involve their mass, will be operating in the same way when the entities together compose the bigger entity than when they were independent. Given that this kind of aggregation is not relevant to understanding Gillett's view of emergence, I will not further explain it. I believe it will become clearer when the contrast is made with The Conditioned View of Aggregation. For more details, see Gillett, 2016, p.118.

According to The Conditioned View of Aggregation, which Gillett defends in an emergentist way, aggregation is discontinuous and entails downward determination:

In certain complex collectives (...) we have discontinuities where components contribute differential powers not accounted for by the laws and/ or principles of composition holding of such components in simpler collectives. (...) [T]he new powers and behaviours of the components are not those that they would have if the laws applying in the simpler collectives were the only laws applying to the complex, composite systems. (...) We thus have the space for a novel variety of determination between composed and component entities. (Gillett, 2016, p. 205-206)

That “novel variety of determination” is what Gillett has named *machresis*. Machresis is determinative because higher-level features make a difference to the powers of its realizers, but it is non-productive because they do so without involving the manifestation of powers. How can an entity make a difference to the powers of another entity without this being a productive process? According to Gillett, some higher-level properties are determinative because they contribute differential powers to their realizers (Gillett, 2016, p. 207). Composed entities do not have new downward causal powers, rather they contribute to the powers of their realizers.

Let me introduce an example, which can be found in Gillett (2016, p. 213), to illustrate how machresis works. Surprisingly, the example he provides and which I will now follow does not involve scientific entities but imaginary ones. Gillett provides several reasons to justify the fact that he does not introduce his view with a real scientific case, the main one being that his goal

is not to argue that there is machresis in nature, but to provide a theoretical framework for emergence. Although I will not discuss this now, it would certainly be a problem for his view if it turns out that there are no empirical cases of machresis. In particular, given the strong emphasis Gillett puts on following the scientific literature to build his view.

The example involves the following imaginary entities: blobs, which are lower-level entities, and globules, which are higher-level entities composed by a collection of blobs. Figure 5 shows the two ways blobs can aggregate to compose globules. Let me first consider the left side of Figure 5, where blobs assemble following the Simple view of aggregation. Each blob has property P, which bestows the power to attract other blobs. Globules have property F, which results from the simple aggregation of blobs with property P. Property F bestows the power to attract globules in all directions. This power is comprised by the powers the component entities, blobs, have to attract other blobs, but it is different from them because it attracts globules, not blobs. In this way, Gillett explains what he believes is a key feature of scientific entities: higher-level entities are qualitatively distinct from their component entities. Meaning that scientific entities have properties and powers that, although realised in and comprised by, respectively, their lower-level components, are different from them. This is therefore meant to be a schema that fits any non-emergent composed scientific higher-level property. For instance, Gillett would consider that the structure represented on the left side of Figure 5 perfectly fits the diamond example. Importantly, according to Gillett this is not a case of emergence, as the behaviour of the globules can be explained by the collective behaviour of blobs. Gillett claims that even a reductionist could accept a picture similar to this one, simply adding that globules are not ontologically distinct from their parts given that their behaviour can be reduced to the behaviour of the collection of blobs. To conclude, cases of composition in the sciences that fit the representation on the left of Figure 5 are compatible with reduction (Gillett, 2016, p. 176).

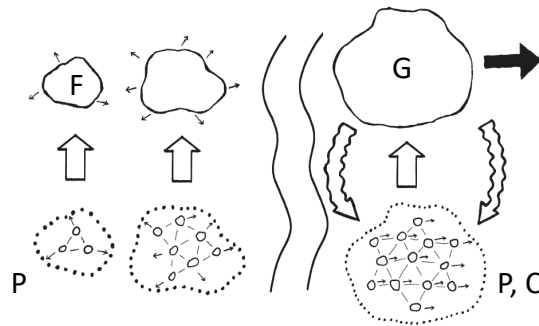


Figure 5. The drawings show the two different kinds of aggregation. On the left, there is simple aggregation. Entities aggregate following the laws of composition and therefore the composite entity has powers that are comprised by the powers of its components. On the right, the upward arrow represents the conditioned view of aggregation. The downwards curvy arrows represent machresis. Figure extracted from Gillett (2016, p. 213) and modified.

Let us move on to the right side of Figure 5. It represents a scenario where blobs assemble following the Conditioned view of aggregation, which entails that higher-level globules are determinative towards their components in virtue of there being a machresis relation between them and the collection of blobs. Let me show what this all means in this example. Gillett claims that when the Conditioned view of aggregation holds, globules have a new property, *G*. This property bestows the power of strongly attracting other globules in a certain direction (not in all directions, as *F* did). There are two ways in which we can account for this scenario: 1. we could argue that the power bestowed by *G* to attract globules in a certain direction is strictly higher-level, so the powers blobs have remained the same, and they could not attract blobs in a certain direction. In this case, globules would attract other globules in a certain direction, which in turn would imply downward causation given that globules are made-up by blobs. Therefore, Causal Closure of the Physical would be violated, and physicalism would be at stake. 2. The other way to account for this scenario, which according to Gillett does not present a problem for physicalism, is the following:

I stipulate that when a globule has an instance of the property *G*, then the instances of property [*P*] in its constituent blobs continue to contribute the power of attracting blobs with the same magnitude they have in simple collectives, but *also* contribute the further power, [*C*]. (Gillett, 2016, p. 213).

This is Gillett's preferred view to explain the scenario in the right side of Figure 5. The idea is that when entities assemble following the Conditioned view of aggregation, new powers

emerge at the level of globules. However, given that Globules are made-up by blobs and physicalism must be preserved, it is posited that blobs also have a new power, C. Blobs determine (because they non-causally constitute) globules, and in cases of machresis, globules in turn determine blobs. Only in this way can we understand that the powers of G are comprised by the powers of blobs, and that G's features are realized by the features of blobs, while at the same time G determines the differential powers, C, of blobs. This is Gillett's conclusion:

Thus we have an overlooked way, or set of ways, to potentially defend the further claim that "Wholes are more than the sum of their parts" because "Parts behave differently in wholes." In such a situation we would have properties that are realized, but which still play a determinative role in nature both by determining the powers of their realizers, their "parts," and also by their own contributions of powers producing effects at their own level. When the Conditioned view of aggregation holds, we would thus potentially have composed properties that we ought to accept as determinative despite being realized – and hence have S-emergent [in the sense of scientific emergence] properties! (Gillett, 2016, p. 195)

In what remains of this section I will explain how Gillett deals with Kim's Exclusion Argument and with the threat of anti-physicalism.

Gillett deals with Kim's Exclusion Argument by drawing on his distinction between production and causation, and on his joint role-filling account. The joint role-filling account allows that higher-level properties have different powers than their realizers, or in Gillett's terms, that they are qualitatively different. Their powers are comprised by the powers of their components, yet they are different from them. For this reason, higher-level entities do not directly *produce* effects at lower levels because they do not have the powers needed to do so, but in cases of machresis they can be joint *causes* of those effects, so there is what he calls "mediated downward determination". Consider a case of machresis in which we have lower-level entity, P_z , and $P_1 - P_n$ lower-level entities that realize a determinative higher-level entity G. P_z is produced by the differential powers that $P_1 - P_n$ have when in a machretic relation to G:

It is not true that the lower-level realizer instances $P_1 - P_n$ would produce the instance of P_z without the Strongly emergent property instance G – so G is plausibly a joint cause of the instance of P_z . (Gillett, 2016 p. 242).

G and P₁-P_n, given their interdependent nature, are joint causes of H—a higher-level entity—, and also of P_z. However, G and P₁-P_n do not independently cause P_z (or H for that matter). G does produce H, and P₁-P_n do produce P_z. So there is no overdetermination at any of the levels. While avoiding overdetermination, Gillett claims that his view can account for a kind of downward causation that he calls “mediated downward causation”. Although higher-level entities cannot directly cause any lower-level effects, they are causally related to those effects because they are their joint causes, together with lower-level powerful entities. In this way, Gillett avoids the problems arising from direct downward causation, while keeping some causal role of higher-level entities in lower-level effects, which is key in robust accounts of emergence.

Regarding the threat of anti-physicalism, Gillett’s response is based on his own definition of the Completeness of Physics, which has been formulated considering the production/causation distinction. Usually, the anti-physicalist threats involve arguing that there is a violation of Completeness. However, in this case given that Completeness is formulated in terms of determination, rather than causation, Gillett claims that given the possibility of machresis Completeness does not hold. And yet he claims that his view is not anti-physicalist. The main reason he offers is that Completeness is suited only for reductionist physicalist accounts, where the only fundamental laws are the laws holding between microphysical entities. However, according to Mutualism there are further emergent yet fundamental laws corresponding to determinative composites and the machretic relations they have with their components. As long as mutualism remains a live theory, as Gillett believes, we cannot characterise physicalism by appeal to Completeness: “the reigning philosophical orthodoxy is that the truth of physicalism entails the truth of the Completeness of Physics. However, when one stops to look, explicit arguments supporting such a necessary connection between physicalism and the Completeness of Physics are surprisingly hard to find.” (Gillett, 2016, p. 257). Although it might be difficult to come up with an argument supporting the connection between Completeness and physicalism—for positive arguments about why physicalism should not be linked to Completeness see McLaughlin (1992)—, it is also true that empirical cases of machresis may be even harder to find. This is quite perplexing, given Gillett’s commitment to a view that draws as much as possible from current scientific knowledge. Still, mutualism remains, at least theoretically, a live view.

At this point, I am in a position to assess Gillett's Scientific Emergentism, Mutualism, versus Wilson's Weak Emergentism. Both views present a way to account for Physicalist Emergence, that is, emergence within a physicalist framework. However, what it takes for a property to be emergent is very different in both accounts. According to the subset view, both the diamond and the globules formed following the Simple view of aggregation are weakly emergent entities: the diamond (globule) has a subset of powers of the powers had by the aggregate of atoms (blobs) composing it. The aggregate of atoms forming the diamond not only has the power to scratch glass, but for instance it also has the power, in virtue of how the lattice is structured, to keep all its electrons localised. This means that all electrons must always remain bonded to a carbon atom (by contrast, in the graphite structure, where we have the same components but a different kind of lattice structure, each carbon atom has a non-bonded delocalised electron). Hence, the collection of carbon atoms that compose the diamond has at least one power more than the diamond: the power to keep all of its electrons bonded. As I explained in Chapter 2, Wilson's view does not require new determinative powers for a property to be weakly emergent, rather she just requires a different causal profile. This contrasts with Gillett's requirement for scientific emergence: the realised property must be determinative towards the powers had by its components. Therefore, nor the diamond nor the globules formed following the Simple view of aggregation are emergent under Gillett's view of emergence. Emergence requires the globules being formed following the Conditioned view of aggregation. Although lacking empirical support, I have claimed that Mutualism is nonetheless a live view. However, there are a few reasons why I consider it an inferior nonreductive physicalist view than the subset account. Some concepts used and some claims made are idiosyncratic and this is a burden for the view. For instance, the distinction between production and determination leads to the consequent rephrasing of the Eleatic Principle and the Causal Closure of the Physics (which his view violates) in terms of determination. Gillett also introduces the distinction between Simple aggregation and Conditioned aggregation, and his nonreductive physicalist view requires the existence of the relation of machresis. For all these reasons, I consider the subset view's reply to Kim's CAE is much more appealing: it is able to provide a solution in Kim's own terms; it does not need a new set of concepts and a new relation to do so. Hence, I believe the subset view is much better positioned to explain higher-level causation. However, that hinges on whether Gillett's Flat Objection, which I explain in the following section, is able to undermine the subset view.

4.2 Gillett's Flat Objection against the Subset View

To motivate the need for an account of metaphysical dependence that takes the realisers to be features of components –i.e., many-to-one– rather than features of collectives –i.e., one-to-one, Gillett introduces what I call the Flat Objection directly against the Subset View. This objection is a rather powerful one and is taken by Shoemaker at face value: the conclusion of Gillett's objection is one of the two reasons Shoemaker provides to motivate the introduction of a new realisation relation: the microphysical realization relation. Gillett has presented his Flat Objection in several publications. He claims that this objection provides enough grounds to reject the Subset View. In her *Metaphysical Emergence* (2021) Wilson tackles the objection extensively in section 3.4.2 for the first time. She had briefly engaged with Gillett's work by providing a brief analysis of the many-one or one-one nature of the realization relation in her *Wilson* (2015), only recently did she acknowledge its power and try to overcome it. Regarding Shoemaker, as I mentioned above, he accepts the conclusion of the objection without further consideration of the argument itself, and the objection is used to motivate his microphysical realization relation to supplement the deficiencies of property realisation. But is introducing a new realisation relation the only way to overcome the objection? Shoemaker's strategy clearly is a workaround, but an ideal solution for the subset strategist would be to show that the objection does not work.

The Flat Objection is explicitly intended against Wilson's view (Gillett, 2010, p. 168), amongst other subset strategies. Although Gillett acknowledges that Shoemaker's microphysical realization is intended to overcome the Flat Objection, he is unsure whether it is sufficient to do so. Gillett's argument is based on a premise about the nature of the relation between special science's higher-level entities and lower-level entities. He argues that this premise simply reflects the scientific understanding of higher-level and lower-level entities, and that for this reason it is uncontroversial: given that scientists account for higher-level entities —i.e., individuals, powers, processes, or properties— by providing compositional explanations⁹ between them and some set of lower-level entities, any philosophical account of the relation between scientific entities must have the same features composition has. For instance, Gillett

⁹ In some works, Gillett used the term “mechanistic explanation” instead. However, in his most recent book he reserves this term for a particular case of compositional explanation: that between a higher-level process and lower-level ones.

mentions that we now understand “why crystals bend light in certain ways in terms of the properties and relations of the atoms or molecules taken to compose them” (Gillett, 2016, p. 33), or the hardness of a diamond in terms of the “bonding and spatial alignment” of the carbon atoms that compose it (Gillett, 2010, p. 170). Thermodynamic systems also provide some examples, although Gillett does not directly refer to them, such as the pressure of the gas being understood in terms of collisions of the tiny particles of the gas with the walls of the volume that contains it, or temperature in terms of kinetic energy.

Gillett claims that this kind of compositional explanation is commonplace in the sciences. So any philosophical account of the dependence relation between scientific entities must be compatible with the features of compositional explanations. One key feature is the qualitative distinctness between the many lower-level entities and the composed entity. Qualitative distinctness refers to the fact that the entities of lower-level entities have different powers and properties than those had by the entities they compose. In particular, Gillett focuses on powers. What lower-level entities can do is different from what higher-level entities can do: “the carbon atoms do not contribute any powers in common with the property of the diamond —the diamond has the power to scratch glass, but no carbon atom has such a power” (Gillett, 2010, p. 174).

Given that qualitatively distinct properties must have different powers by definition,¹⁰ then higher-level properties cannot be realised by properties with which they share all of their powers. This is an argument against the subset view, as Wilson’s account requires that higher-level properties’ powers be a subset of the powers of their lower-level realisers. Having a distinct causal profile, which is the source of causal efficacy of higher-level properties in Wilson’s account, is not enough for qualitative distinctness. So far, we have a negative argument against a one-one account, such as Wilson’s. If we add physicalism as a premise, then we obtain a positive argument for a many-one account: if one accepts physicalism then one must accept that higher-level properties cannot have new causal powers, as that would violate the causal closure of the physical. So, in order for higher-level features to have different powers than their lower-level realisers, i.e., in order for them to be qualitative distinct, there must be a many-one relation between realisers and realisee. In this way, a power of the higher-level

¹⁰ One might argue that this only holds if the CTP is true, but on Gillett definition of “qualitative distinctness” it must involve different powers. And certainly, that is what happens in the scientific examples presented.

feature is comprised by the powers of the lower-level features but is not identical to any of them. Remember that the relation of “being comprised by” is Gillett’s way to refer to a compositional relation between a higher-level power and the joint powers of many lower-level entities that is compatible with physicalism. So, one can preserve physicalism and yet account for qualitative distinctness if one adds a compositional relation between powers at different levels. Gillett accepts this and argues that higher-level powers are comprised by the joint powers of many lower-level entities. Bringing back the example of the diamond and the carbon atoms, the power of the diamond to scratch glass is comprised by the powers of many carbon atoms (in virtue of their strong bonds and alignment) to break the bonds of glass molecules. According to Gillett, then, qualitative distinctness can only be preserved if higher-level properties are realised by the joint properties of an organised group of lower-level entities. In this way, the realiser of a higher-level property, such as being hard, is the joint properties, such as the property of forming strong bonds each carbon atom has, of a collection of lower-level entities. A one-one relation in a physicalist framework will always entail that a higher-level property’s powers are numerically identical to the powers of its realiser, thus precluding qualitative distinctness. Therefore, Gillett contends, there is no way the subset strategist could account for qualitative distinctness, and the many-one approach must be accepted.

4.3 Wilson’s Response to Gillett’s Objection (and Why It Does Not Work)

The Flat Objection raises important problems for the Subset View. In what follows I introduce and discuss Wilson’s solution to those problems. The solution is based on the existence of micro-structural or middleman features. As I explained in Chapter 2, Wilson’s realisation relation is a relation between two properties at different levels. For instance, determinate properties realise determinable properties in virtue of the latter only having a subset of powers of the former, e.g., scarlet realises red. The subset account was first originated within study of the metaphysics of mind to show that non-reductive physicalism is a tenable view that can account for mental causation and yet avoid Kim’s Causal Exclusion Argument. Since then, the Subset View has evolved to apply not only to cases of determinate/determinable relations or mental/physical relations, but also to how special science properties can be causally efficacious.

Very generally, the problem raised by Gillett is that the special science properties as discussed in scientific literature are not—at least not straightforwardly—realised by a single lower-level

property. Rather, the scientific literature is most in line with the following: special-science property, M , of special-science entity S is realised by the properties P_i of the lower-level entities s_i that compose S . The structure of these cases is *prima facie* incompatible with the Subset View, as M does not contribute any power shared with P_i , let alone a subset of them. As I explained before, this is the core of the Flat Objection. The Subset Strategist, then, needs to recover her property-property (one-to-one) picture in order to be able to apply the subset strategy. Wilson does so by introducing a micro-structural property, referred to as a “middleman feature”. From now on I refer to this solution to the Flat Objection as the Middleman Strategy. It is important to emphasise that the Flat Objection targets cases of realisation of special science properties. Wilson can account for the relation between determinable/determinate (red/scarlet) without the Flat Objection arising, as in that case there is no clear compositional relation between the relata that could problematise the one-to-one picture. However, failing to account for the realisation of special science properties would clearly undermine the subset view, as it clearly is one of the domains subset strategists consider fundamental. To inadequately address these would drastically reduce the explanatory power of the subset view. Therefore, the Flat Objection forces subset strategists to consider more carefully the complex nature of the relation between higher-level and lower-level properties.

But what are middleman properties? Is their existence justified? If they turned out to be ontologically superfluous, how could the subset view survive? I will address these questions in this section, which has the following structure. First, I thoroughly explain the Middleman Strategy and I show in detail how it is supposed to solve the Flat Objection. Second, I introduce and assess Wilson’s view on middleman features, which slightly changed from Wilson (2015) to Wilson (2021). Finally, I discuss Gillett’s contention that even if their existence was justified, middleman features cannot solve the Flat Objection.

4.3.1 The Middleman Strategy

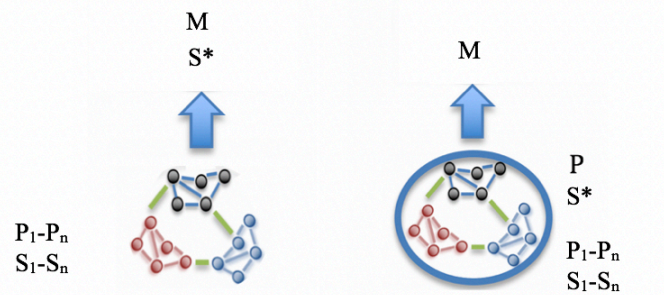


Figure 6. The blue arrow represents a realization relation, the dots represent individuals, the colour of the dots represent properties, and the straight lines represent relations between individuals. On the left we have a many-one realization relation, while on the right we have Wilson’s preferred one-one realization. On the right there is depicted the structural feature (or middleman feature, as Wilson calls it), P , that realizes M , and that is not needed in the many-one relation.

The Middleman Strategy, represented at the right in Figure 6, consists in positing the existence of an aggregate, S^* , resulting from an “ontologically lightweight” —i.e., not emergence-engendering— combination of many lower-level individuals, S_1-S_n , their properties, P_1-P_n , and the relations among them. This is what Wilson refers to as a relational aggregate, because the parts bear certain relations to each other. Structural properties such as P are properties of the relational aggregate, S^* in this case, and are the realization base of all the higher-level properties of S^* , in this case we are focusing on M . Structural properties can be properties only of pluralities or aggregates, as they always have the following form: a structural property is the property of being made-up of certain lower-level entities, propertied and related in a certain way. In previous work, Wilson calls those structural properties “middleman” properties, although in her most recent work, *Metaphysical Emergence* (2021), she drops talk of middleman features and introduces talk of base features: “the *base feature* at issue is a feature of a given (ultimately physical micro-) configuration (plurality or structural aggregate)” (Wilson, 2021, p. 73, my italics). As is clear from this quote, there is no substantial change to the notion of middleman feature beyond the change in the words used to refer to them. There is reason to believe that Wilson has decided to name them differently because calling them “middleman features” attracted unwanted attention towards them: such phrasing seems to imply that their nature differs in some way from the nature of regular features (i.e., features that apply to entities instead of pluralities). By calling them “base features”, this concern does

not arise so straightforwardly. Nonetheless, I will continue to refer to them as “middleman features”, as I think there is no need to shy away from the fact that they are features that link the features of lower-level entities with the features of higher-level entities. I now explain how middleman features can be used to solve the Flat Objection.

The reasoning behind the Middleman Strategy is that there is something different about middleman features: they are structural properties. If we go back to Figure 6, the middleman feature P would be akin to the property of being made of three sets of five particles, one blue set with entities related in a “bluey” way, one red set with entities related in a “redly” way, and one black set with entities related in a “bluey” way, with “greenly” relations among the sets. This contrasts with the property of being blue that some of the entities have, as middleman properties are properties of “being made-up by entities S_1 - S_n propertied and related in a certain way”. The causal powers higher-level feature M has in virtue of being realised by middleman feature P are exactly the same powers it would have by being directly realised by the collection of lower-level entities S_1 - S_n propertied and related in a certain way in a many-to-one structure. The main difference, then, is that middleman feature P is a single property that contains information about all the lower-level goings-on that contribute to M . Given Wilson’s understanding of physicalism, each causal power M has in virtue of being realised by P must be numerically identical to a power of P to avoid violation of CCP. On the other hand, each causal power M has in virtue of being realised by the collection of lower-level entities S_1 - S_n propertied and related in a certain way is a power not had by any of the particular entities in the collection, although comprised by them in a way compatible with reduction.¹¹ As I have shown, the fact that middleman properties are (micro-)structural properties means that they contain information about the lower-level components and their interactions. In this way, they capture the qualitative differences between the lower-level comparatively non-relational entities and the higher-level entities precisely because they are the link between them. I will come back to this in more detail shortly when I discuss Wilson’s (2021) stance on middleman features in 4.3.2.

¹¹ Except if there is machresis, which would not be compatible with reduction. But this case is not relevant to this discussion as it is not part of Gillett’s objection against the subset view. Rather, it is Gillett’s positive account of emergence compatible with physicalism.

The key difference between the two approaches is whether realisation is a one-to-one or a many-to-one metaphysical relation, which in turn has implications for how physicalist emergence should be conceived. The former means that a single lower-level relational property realises a single higher-level property. The latter means that many comparatively non-relational lower-level properties jointly realise a single higher-level property. The subset view needs realisation to be one-to-one because it is only in this way that the subset condition –i.e., higher-level properties must have a proper subset of the powers of their realisers– is made possible. So one-to-one realisation will be referred from now on as *realisation_{O-O}*, where *O-O* stands for one-to-one, while many-to-one realisation will be referred to as *realisation_{M-O}*, where *M-O* stands for many-to-one. The debate over whether realisation is one-to-one, or many-to-one is central to this thesis. In what follows I introduce and discuss Wilson’s and Gillett’s theoretical justifications for preferring one over the other. Crucially, the two approaches have different ontological commitments: the one-to-one approach requires the existence of middleman properties, and the many-to-one approach requires that a collection of lower-level powers can give rise to a single higher-level power.

4.3.2. Middleman Properties: The Controversy

An important problem for the one-to-one approach is that middleman features are controversial. A further problem is that, even if we granted their existence, it is not clear whether the solution to the Flat Objection middleman features offer is satisfactory. I will first focus on the former problem. Although Wilson did not properly engage with Gillett’s objection until Wilson (2021), in Wilson (2015) there is a brief discussion of middleman properties that can be put in relation to Gillett’s concerns. In 2021, Wilson presents middleman properties in a slightly different way and introduces more elements to the discussion. Moreover, she provides for the first time a clear response to Gillett’s objection. Therefore, in this debate there are three stances on middleman properties: 1. Wilson’s (2015) view that the one-to-one approach is equivalent in what it does to the many-to-one approach. 2. Wilson’s (2021) view that the one-to-one approach subsumes the many-to-one approach. 3. Gillett’s (2002, 2010, 2016) view that the one-to-one approach is untenable. In what follows I introduce the three stances and I discuss whether they are satisfactory.

Wilson's (2015) stance on middleman properties

I will first introduce Wilson's stance on middleman features as presented in Wilson (2015). Let me start with a quote that summarises her thoughts at the time.

Emergence is treated (multiple dependence or realizability to one side) as a one-one relation between higher-level and lower-level features. This *treatment presupposes that certain relational lower-level entities exist and have features serving as a dependence base for the associated emergent features*. The presupposition is useful, in encoding (as had by the posited lower-level relational entity) the sorts of features of complex entities that are assumed by all parties not to be emergent, in any interesting sense. Alternatively, *one might (following Gillett 2002) dispense with the relational lower-level middleman* and take the dependence base to consist in collections of comparatively non-relational lower-level features (say, features of individual molecules and pairwise relations between individual molecules), understood as combinable via certain ontologically 'lightweight' compositional principles, including additive causal combination, and certain boolean or mereological operations. (...) [I]t's clear that the 'one-one' and 'many-one' approaches target the same phenomena: the latter considers the nature of the dependence of higher-level entities/features on comparatively non-relational lower-level entities/features given certain allowable combinatorial principles, whereas the former considers the nature of the dependence of higher-level entities/features on relational lower-level entities/ features allowed by the combinatorial principles. (Wilson, 2015, p. 350 – 351)

As it is clear in this quote, Wilson does not put much effort into arguing for the existence of middleman features. That is surprising given the centrality of their role in the subset account: the existence of middleman properties is essential for the subset condition to get off the ground. As I explained above, the subset account can only work under the one-one approach, so not providing sufficient justification for introducing middleman properties is a weakness that leaves the view vulnerable to objections. In particular, it is perplexing that in the quote above Wilson claims that one could alternatively "(following Gillett 2002) dispense with the relational lower-level middleman and take the dependence base to consist in collections of comparatively non-relational lower-level features". There is no question, then, that Wilson believes that the many-to-one and one-to-one approaches are interchangeable in the sense that they both target the same phenomena, the realisation of higher-level properties, and both can account for it. In the one-to-one approach, however, the existence of a middleman feature is required. As Wilson points out, the middleman feature is non-problematic because it merely

consists in what we accept to be non-emergent features of the lower-level aggregate. However, if they are two available alternatives that target and account for the same phenomena, but the one-to-one approach's ontology is heavier because it requires the positing of a new set of properties, why should one prefer it? Even if middleman properties are lower-level and not emergence-engendering, they still are a new set of properties the existence of which needs to be justified. Wilson argues that it is on the grounds of pragmatism that one should favour the one-to-one approach: "The presupposition [the existence of middleman features] is useful, in encoding (as had by the posited lower-level relational entity) the sorts of features of complex entities that are assumed by all parties not to be emergent, in any interesting sense." At this point, Wilson offers no other arguments in favour of the one-to-one approach. A justification in terms of usefulness might be sufficient, as being explanatory useful would be enough for many to justify the existence of a property. However, in this case I argue usefulness is not enough. The aim of the subset strategy is to provide an account of nonreductive physicalism that is able to overcome Kim's Causal Exclusion Argument. However, Gillett's view, which also provides an account of nonreductive physicalism that is able to overcome Kim's Causal Exclusion Argument does not need to posit any extra property. So, if we can achieve the same goal positing less properties, clearly there is a parsimony consideration to prefer the account that needs less properties. Of course, one could go on to argue that the subset view provides a better account of nonreductive physicalism than mutualism, but this is another discussion. Moreover, there is the concern that middleman properties do not play a role in scientific theories, so the need to posit them to account for the relation between scientific properties is problematic. I will come back to this when I explain Gillett's view on middleman properties. It is not clear why Wilson would let her approach be vulnerable to objections such as Gillett's, I believe it is because she did not consider them a substantial problem for her account. In the next section I show why.

Wilson's (2021) stance on middleman properties

As I have mentioned above, in her 2021 Wilson is more thorough, and not only claims that the one-one approach subsumes the many-one approach, but she also gives an argument for why one should prefer the former.

On Gillett's approach, emergence involves the cotemporal material dependence of a single (e.g., special-scientific) feature on multiple, comparatively nonrelational (e.g., physical) features, where the latter are understood as combining in various (e.g., lower-level causal,

spatiotemporal, or mereological) ways agreed by all parties not to involve emergence, whereas the one-one approach takes emergence to involve the cotermporal material dependence of a single (e.g., special-scientific) feature on a single (e.g., physical) feature of a micro-configuration (plurality or structural aggregate), understood as already combined in various ways agreed by all parties not to involve emergence. *So understood, the many-one approach is subsumable under the one-one approach.* (Wilson, 2021, p. 11, my italics)

As we can read in the quote, Wilson's view on middleman features has not changed, and is consistent with the representation in Figure 6. In this case, however, she has moved from claiming that the two approaches are interchangeable to claiming that the many-to-one approach is *subsumable* under the one-to-one approach. As one may understand subsumability in this context, her point might be that whenever the many-to-one approach is valid, so is a one-to-one approach. If that were the case, there is a sense in which the quote would be saying something similar to the quote from 2015, namely that the many-one and one-one approaches are equivalent in what they do, and she would just be adding that the one-one approach applies to any scenario where the many-one approach is valid. The argument for that is that the one-to-one approach uses a middleman property that is able to capture the many properties plus the organisation of the many lower-level entities that jointly realise the higher-level entity, so it is able to take all the information about the "many" side of the many-one relation and encode it into a single lower-level micro-structural property, which serves as the realiser.

If that was how Wilson understands subsumability, namely that middleman properties capture the information of the "many" plus their organisation, so they can be invoked in any case of many-one realisation of special-science properties, we would again question why one should add such properties to one's ontology if everything can be done without them. The only argument she had provided for that so far is based on the utility and pragmatism of such an approach. However, when discussing constitutive mechanisms in her 2021 she provides a stronger argument for introducing middleman properties and thus for a one-one account of realisation. The argument concludes that there are instances where only the one-one approach works, so middleman features are required to account for those cases. This leads to understand her sentence about "subsumability" in the sense that one-to-one realisation can account for all cases of many-to-one realisation, and more. In what follows I explain her reasoning, and ultimately show it does not work, so it remains unclear why we should add middleman properties to our ontology.

Gillett's (2002, 2010, 2016) stance on middleman properties

Finally, I briefly discuss Gillett's stance on middleman features. I have shown that his account of realisation_{M-O} does not need them, and this fact provides an argument against middleman features due to parsimony concerns: given that we could do all the work with a many-one approach, and thus we could easily avoid positing another property between realisers and realisee, there is no compelling reason why we should do so. Not only they are superfluous, but Gillett contends that they are "ontological monsters" because they combine different ontological categories, in this case properties and relations. He considers that this kind of monster properties and entities should be avoided, as they lead to a confusing ontology. On the other hand, Gillett argues that even if we grant the existence of structural properties, they are not able to solve the Flat Objection, rather they just mask it. "The resulting picture (...) converts the relations of making-up we sought to understand into properties of being made-up—hence masking that the cases actually involve relations between levels of individuals, powers, properties and processes which are qualitatively distinct" (Gillett, 2010, p. 185, my italics). In other words, by positing a middleman property between many lower-level properties and higher-level properties the subset strategist is not providing an explanation of the key feature of scientific making-up relations: the realization of qualitatively distinct higher-level properties by a group of lower-level properties P_1 - P_n .

4.3.3. Wilson's Positive Argument for the Existence of Middleman Properties (and Where It Goes Wrong)

The argument is based on constitutive mechanisms. It concludes that in cases of constitutive mechanism realisation middleman properties are necessary, as the many-one approach alone cannot capture the nature of such realisation. Therefore, the one-one approach is not equivalent to the many-one approach, rather it subsumes it in the sense that it applies to a broader range of properties. In Wilson (2021, p. 110-112) she introduces constitutive mechanism realisation as such a case: a realisation relation where the many-one approach does not work, and the one-one approach does. To properly assess the argument, I first need to introduce the constitutive mechanism approach.

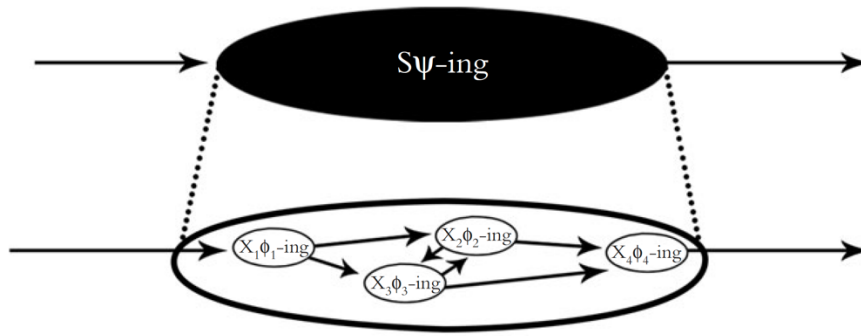


Figure 7. The entities $X_i\phi_i$ -ing and their organisation give rise to the phenomenon of $S\Psi$ -ing. Figure extracted from Craver (2007, p. 7).

According to this view, many special science entities, in particular most biological entities, are mechanisms. Mechanisms are composed of interacting organised parts. As shown in Figure 7, entities X_i compose the higher-level entity S . The components of mechanisms are causally, spatially, and temporally organised and they interact with each other. The resulting organised activity —in Figure 7 that is the joint organised activities of $X_i\phi_i$ -ing — is the mechanism, and it realises the behaviour or properties of the whole they compose — i.e. gives rise to the phenomenon of $S\Psi$ -ing.

According to Wilson (2021, p. 60-62), mechanistic explanations have led to two different accounts of mechanistic realisation: a mainstream one-one account, realisation_{O-O} , and a many-one account, realisation_{M-O} , defended mainly by Gillett. Now I can finally introduce Wilson's argument for the need for middleman properties in cases of mechanist realisation, which is also an argument for realisation_{O-O} .

Her argument is based on Craver (2001). Drawing on this paper she claims that explaining a mechanism requires providing a description of the organisation of parts and how they interact, in short, a description of how they are all integrated so that their joint behaviour results in the mechanism. In Craver's words:

An analytic account for a mechanism is not just a list of entities and activities; it is a description of a mechanism. And that description involves, in addition to a list of entities and activities, a description of how they are organized together actively, spatially, and temporally in S 's ϕ -ing. (Craver, 2001, p. 61)

Wilson translates this into metaphysical terms roughly as follows: “mechanisms need to be explained by descriptions of how their parts are organised and integrated” is translated to metaphysical terms as “mechanisms are not realised by a plurality of features of their parts”. And Wilson continues, in what is a key fragment regarding this debate:

Nor would Gillett deny that in order for mechanistic realization to occur, the parts and associated features need to be appropriately spatiotemporally and causally integrated. But if a many-one account of mechanistic realization is feasible only if the many parts and their many features are appropriately integrated, there must be some metaphysical locus of this integration; and the natural thing to say here is that this locus is something like a micro-structural feature. (Wilson, 2021, p. 111)

Wilson's claim here, in simpler terms, is that the workings of mechanisms cannot be fully explained by solely considering their individual parts. Rather, it is crucial to also include how these parts are arranged and interact with each other. Translated into metaphysical terms, this implies that there exists a micro-structural feature that plays a crucial role in the functioning of the mechanism. Essentially, the organisation and interplay of parts are just as important as the parts themselves, and this requires a unique feature at the micro-structural level. Looking back at Figure 7, the oval encircling the X_i and their organised ϕ_i -ings is a lower-level middleman feature that gives rise to the phenomenon, $S\psi$ -ing. This middleman feature is something along the lines of the feature of having X_i parts, related spatiotemporally and causally in a certain way, engaging in ϕ_i -ing activities. From this, Wilson straightforwardly concludes that middleman features exist in this case and that the one-to-one approach is best suited to account for constitutive mechanism realisation.

Now I turn to showing that Wilson's argument has important shortcomings. The main reason why the argument is flawed is because she has cherry picked the features of the constitutive mechanist approach that best fit her view. Her interpretation of the constitutive mechanism account and her reading of Figure 7 are not the whole picture. In particular, Wilson overlooks what I consider to be important nuances discussed in Craver (2007) and Povich & Craver (2017) and she also misses the work of Bechtel (2017), who being a mechanist himself criticises Craver's approach in a way that should not be overlooked by a non-reductionist physicalist. In what follows I show that once we take into account the wider literature on constitutive mechanisms, the argument becomes –at least– controversial.

First, I introduce nuances discussed in Craver (2007) and Povich & Craver (2017) and then I show how they undermine Wilson’s argument that constitutive mechanist realisation only works with a one-to-one approach, so constitutive mechanist realisation is of the kind realisation_{O-O} . From the brief introduction to constitutive mechanisms above it might not be straightforwardly clear, and it is not a topic vastly discussed in the constitutive mechanist literature, but Craver (2007) and Povich & Craver (2017) introduce a nuance relevant to the topics discussed in this thesis:

“[Figure 7] is ambiguous: Two distinct relations are represented in both the top and bottom (corresponding roughly to what Gillett (2002) calls dimensioned and flat realization). The first is a part-whole relation between S’s ψ -ing and each of the X’s individual ϕ -ings [realisation_{M-O}]. We use the term “mechanistic level” exclusively for this relation. The second is a whole-whole relation [realisation_{O-O}] between two ways of describing the same thing. We can talk about S’s ψ -ing or we can talk about the organized collection of ϕ -ing Xs. These are two ways of describing one and the same object.” (Povich & Craver, 2017, p. 5)

They distinguish between two kinds of levels: levels of mechanisms —or, in my terminology, realisation_{M-O} — and levels of realisation —i.e., realisation_{O-O} —, as shown in Figure 8 below.

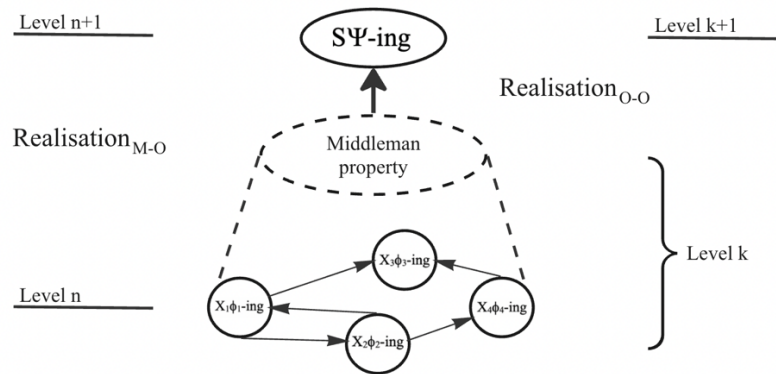


Figure 8. Modification of Figure 7 above to show the two kinds of levels already present, although hidden, in that figure. Realisation_{M-O} bridges levels of mechanism (n and n+1) as an interlevel relation of organisation, and realisation_{O-O} bridges levels of realisation (k and k+1).

Mechanistic levels consider part-whole relations in mechanisms, where parts are always at a lower level than the whole they compose. The entities X_i are at a lower mechanistic level than S because they are components in the mechanism for SΨ-ing . They are components because

the organised X_i and their corresponding activities, ϕ_i -ing, constitute the phenomenon of $S\Psi$ -ing. S is also referred to as the mechanism as a whole. Mechanistic levels, therefore, go from level n —i.e. level where one finds parts and their properties and activities— to level $n+1$ —i.e. level where one finds mechanisms as a whole. But as I explained before, explanations of mechanisms require reference to how those parts are organised. Remember that Wilson translated this into metaphysical terms as a requirement for the existence of a middleman property. However, there is no such property needed in Figure 8. As Craver explains:

Organization is the interlevel relation between a mechanism as a whole and its components. Lower-level components are made up into higher-level components by organizing them spatially, temporally, and actively into something greater than a mere sum of the parts. (Craver, 2007, p. 189).

The integration between parts and their activities, that is, the bridge from level n to level $n+1$, does not require positing a micro-structural property —and thus going through another level, k . It can be explained by an interlevel *relation* of organisation, in Craver’s terms. The relation that obtains is exactly Gillett’s many-one realisation_{M-O} relation. On the other hand, there is also a one-one relation between the collection of parts $X_i\phi_i$ -ing plus their organisation and the mechanism as a whole, $S\Psi$ -ing —that is, between the white oval encircling the collection and the black oval in Figure 7. To illustrate this, Craver uses a common example:

the ability to generate action potentials [$S\Psi$ -ing] *is not at a higher level of mechanisms* than the property of being a particular organized collection of acting Na^+ channels, K^+ channels, and a membrane [collection of $X_i\phi_i$ -ing]. Rather, the parts-plus-organization realize [realise_{O-O}] the generation of action potentials. (Craver, 2007, p. 214)

The white oval is what Wilson has in mind when she argues that there is a locus of integration among the parts and is what she identifies as the middleman property: “the feature of having such-and-such spatiotemporally related parts engaging in such-and-such causal interactions”. The middleman property, although at the same mechanistic level as $S\Psi$ -ing, is at a lower realisation_{O-O} level. Levels of realisation_{O-O} are the kind of levels discussed in the Metaphysics of Mind literature. Given that the white oval realises_{O-O} the black oval, according to levels of realisation_{O-O} the white oval is at a lower realisation_{O-O} level (k) than the black oval ($k+1$). However, if one considers levels of mechanisms they are at the same mechanistic level ($n+1$).

My claim here is not that the Craver (2007) and Povich & Craver (2017) line of reasoning is incompatible with Wilson's view about constitutive mechanism, rather I want to point to the fact that the debate on the nature of mechanistic realisation is not settled. It is true that constitutive mechanists accept that integration of the parts is key to mechanisms. However, they do not accept that integration must entail having a micro-structural property, as it can be alternatively accounted for by considering organisation as an interlevel *relation* (rather than a *property*). In this respect, Craver & Robert Wilson, 2007, section 4.1 *Must Realization Be Decompositional?* provide important insights. Craver & Wilson (2007) review the concept of realisation and reiterate the concerns that there are two kinds of levels in play, and that in some contexts realisation is invoked as realisation_{M-O} and in some others as realisation_{O-O}. They reach no conclusion as to whether realisation should be many-one (what they refer to in the paper as “decompositional”) or one-one, and argue that scientists are concerned with the former, while metaphysicians are concerned with the latter, and there seems to be no agreement in sight.

Scientists, unlike metaphysicians, spend little time worrying about whether wholes are greater than their parts plus their organization. It is a working assumption that the wholes just are the parts organized together. (...) When the whole is greater than the sum, this is because the parts are organized in a particular way, they interact with one another, and they do things together that they could not do alone. But these assertions are entirely consistent with the idea that wholes just are the parts and their organization. (Craver & Robert Wilson, 2007, p. 95)

Metaphysicians use realisation relations in a different context than scientists: they use them when questioning or defending physicalism, or in contexts where physicalism is at stake, because they are concerned about whether the whole could be greater than the parts plus the organisation. However, scientists take physicalism as a premise. Their enquiry is focused on knowing how parts X_i of an object S and their features or activities (ϕ_i -ing) are organised so that they realise features or activities such as ψ -ing of that object S . Given that constitutive mechanists are inspired by scientific explanations and accounts of phenomena, what they find most interesting is a many-one relation between the parts of a mechanism and the mechanism as a whole. This many-one relation between the parts and the whole cannot be disregarded in favour of a one-one relation between the parts-plus-organisation and the phenomenon. As per my previous discussion about kinds of levels, constitutive mechanists accept a many-one realisation_{M-O}, and their project is independent of whether there is or is not a one-one realisation_{O-O}: “my view of explanation is committed only to the existence of higher levels of

mechanisms” (Craver, 2007, p. 218), which, to bring this back to my discussion, implies that his view is only committed to the left side of Figure 8, namely to realisation_{M-O} . In relation to this point, Craver & Povich refer to the levels k and $n+1(=k+1)$ as “two ways of describing one and the same object”. Although this does not necessarily preclude the existence of realisation_{O-O} between k and $n+1(=k+1)$, it certainly shows that the constitutive mechanists do not consider such a relation their main concern.

At this point the reader might wonder why constitutive mechanists introduce levels of realisation if they are not their main concern. The main reason is that by distinguishing levels of mechanisms from levels of realisation they are able to easily deal with the alleged threat to their picture by Kim’s Causal Exclusion Argument (Craver, 2007, pp. 214-217). The gist of the argument is that Kim’s CEA is framed in terms of levels of realisation_{O-O} , however once one moves to levels of mechanisms, the CEA cannot be applied. I do not need to explain this argument in any more detail, as the point here is simply to stress that levels of mechanisms are what matter to the constitutive mechanist. This is also clear in that most work on constitutive mechanism literature only deals with levels of mechanisms. For instance, Craver (2007), Craver & Bechtel (2007) and Bechtel (2017). So far, I have shown that Wilson’s claim that constitutive mechanists can be divided in those that defend a mainstream one-to-one approach to mechanistic realisation and those that defend a many-one fringe account is not true. Constitutive mechanists are actually more interested in a many-to-one approach to how mechanisms are built from their parts, and the idea that there might be a one-one relation between the collection of parts plus organisation and the higher-level phenomenon is mostly brought about to contrast it with the mechanist many-one picture and to show why this picture is not affected by Kim’s CEA.

So far, I have shown that Wilson’s claim that there is a “locus of integration” of parts of mechanisms is not supported by the constitutive mechanism literature. Constitutive mechanists do not need a *locus*, a middleman property to account for higher-level organisation, they invoke organisation as an interlevel *relation*, which corresponds to Gillett’s realisation_{M-O} . This is an important problem for the justification of middleman properties. We have seen that in a sense realisation_{O-O} and realisation_{M-O} are equivalent in what they do, as they can both account for mechanisms: one by means of a middleman property, the other directly. But given that realisation_{O-O} requires positing a new property it remains unclear why one should prefer it.

In her 2021 (p. 111), Wilson provides two arguments against the many-one approach to constitutive mechanisms that I discuss in the next paragraphs. If the previous argument was merely based on reviewing the literature and trying to show how it supports the one-one approach, these next arguments directly tackle the many-one approach and aim to expose its shortcomings. At the same time, the arguments are meant to show that Gillett's Flat Objection does not work. My main conclusion here will be that both Wilson and Gillett are talking past each other, and that the arguments they give fail to engage with the point the other was intending.

Wilson divides Gillett's Flat Objection in two claims, which although related I will consider separately: 1. Given that according to the subset view realised properties inherit all their powers from their realisers, the subset view cannot account for qualitative distinctness between realisers and realised features, and 2. the subset view cannot accommodate realisation between individual features of components and aggregate features of composites. Wilson replies to the first claim with a rather strong position: according to physicalism qualitative distinctness must be accounted for by intra-level relations, so it is not a matter of realisation. This is because qualitative distinctness implies new powers, and new powers at higher levels violate the causal closure of the physical. Middleman properties, which are found at the same level as components, capture qualitative distinctness and new powers, but given that they are at the same level there is not a violation of physicalism. About the second claim, she argues that the metaphysical dependence relation between individual features and the features of the aggregate cannot be realisation, as realisation must involve different levels and that is not the case between individuals and aggregates. This line of reasoning relies of course on the first one: it is because qualitative distinctness must be accounted for by intra-level relations, that the relation between individual features and the aggregate features (which are qualitatively distinct) must be at the same level.

With the understanding we now have of the constitutive mechanism approach, it is clear why Wilson's reply to Gillett is not satisfactory and why I previously said that they are talking past each other. If we consider the two kinds of levels I introduced before, neither Gillett's objection nor Wilson's replies are on point. Gillett's objection applies only if one considers levels of mechanisms, but this is not made explicit in his discussions. On the other hand, Wilson's replies only work if one's framework is that of levels of realisation. Let me start with the first claim and reply.

Gillett's Claim 1: Given that according to the subset view realised properties inherit all their powers from their realisers, the subset view cannot account for qualitative distinctness between realisers and realised features.

Wilson's Reply to Claim 1: According to physicalism qualitative distinctness must be accounted for by intra-level relations. Given that realisers_{O-O} and the features they realise are at different levels, they cannot be qualitatively different. Namely, they cannot have new powers that would violate the causal closure of the physical. In Wilson's view, the qualitative differences one can find between parts and whole and between entities and the collections they form are captured by middleman features.

Gillett is asking the subset view to account for qualitative differences because he is interested in levels of mechanisms. Wilson's reply misses the point of his objection, because according to constitutive mechanism qualitative distinctness can be accounted for by inter-level relations, as I have explained before, without endangering physicalism. The parts of the mechanism and the mechanism as a whole are at different mechanistic levels and the former constitutes (or in Gillett's terms realises_{M-O}) the latter. I propose that a reply available to Wilson is the following.

Proposed Reply 1: The subset view is not in the business of accounting for qualitative differences. Considering levels of realisation, there are no qualitative differences between realisers (i.e. middleman properties) and realised properties. Qualitative differences between realisers and realised properties only appear if considering realisation as a relation between entities at different levels of mechanisms.

Moving on to the second claim and Wilson's reply:

Gillett's Claim 2: The subset view cannot accommodate realisation between individual features of components and aggregate features of composites.

Wilson's Reply: The metaphysical dependence relation between individual features and the features of the aggregate cannot be realisation, as realisation must involve different levels and that is not the case between individuals and aggregates.

In this case, Gillett's claim misses the point because the subset account is not a theory about how individual features of components realise aggregate features of composites. It is a theory about how lower-level relational features of aggregates realise higher-level features of that

same aggregate. Wilson's reply also misses the point because if one takes the levels of mechanism approach, it is actually true that individual features and the aggregates they compose are at different levels (and this need not violate physicalism). A reply available to Wilson is the following.

Proposed Reply: The subset view is not in the business of explaining how features of aggregates arise from the features of their parts. Rather it is concerned only with how lower-level relational features of aggregates realise higher-level features of that same aggregate.

Although these are not sophisticated replies, they are a starting point for a proper discussion about the many-one approach and the one-one approach. On the other hand, it is true that Wilson does not explain how features of aggregates arise from the features of their components: introducing a structural property does not, for instance, explain which features of the lower-level entities and relations among them must be considered. Should middleman features include all of them? Or only those that are relevant to the higher-level features of the resulting composed entity? For instance, mechanist accounts only consider the activities and interactions between lower-level components that are relevant to the activity of the higher-level mechanism. The conclusion from this section is that so far there has not been a proper discussion of the virtues and shortcomings of the many-one versus one-one accounts in the literature. Many-one accounts proliferate in scientific contexts and in the constitutive mechanist literature, while one-one accounts dominate debates on the metaphysics of mind. I have shown that so far the main contenders in the debate have been talking past each other, and that the burden of the proof is on Wilson's side, as she still needs to show why the subset view is preferable to the many-one account in scientific contexts.

CHAPTER 5. TAKING STOCK AND MOVING FORWARD

In the previous chapters I have introduced three different views of interlevel metaphysical dependence relations. Wilson's and Shoemaker's framework for nonreductive physicalism are similar with regards to property realisation, although I have argued that Wilson's account is superior. For this reason, in the remainder of this thesis of the two I focus solely on Wilson's subset strategy. Gillett's account, on the other hand, proposes a different kind of physicalist emergence, mutualism, which is a matter of higher-level properties of the composite being determinative towards its components. In Chapters 6 and 7 I am going to put those accounts to work in two different case studies. It is often the case that metaphysicians, and philosophers in general, do not engage with scientific research concerning the phenomena they are trying to explain. Certainly, there are reasons to aim to provide the most universal account independently of our current best scientific theories. Universality itself is a valued feature of any metaphysical account. However, this is not an argument to disregard the scientific understanding and knowledge of the topic we are theorising and investigating as philosophers. More so if one's proposal aims to account for the interlevel relations between a *scientific* feature and its lower-level realisers. Wilson and Gillett share the view that their accounts need to be compatible with our best scientific theories. The structure of the two main books I discussed in Chapters 2 and 4 —Wilson's 2021 and Gillett's 2016 respectively— is actually quite similar, as they can both be divided in two parts. The first is devoted to providing an account that answers the question about what emergence is:

Just what is metaphysical emergence, more precisely? How is it, exactly, that special scientific and artefactual entities can coterminally materially depend on complex configurations of fundamental physical entities, while retaining some degree of ontological and causal autonomy? (Wilson, 2021, p. 7)

Wilson's answer to this question is twofold: strong emergence and weak emergence. Gillett's answer is mutualism. The second part of their books is devoted to answering the question: are there empirical cases of physicalist emergence? Is physicalist emergence actually instantiated in our world? They both engage more or less extensively with case studies to show the extent to which their proposals go beyond the theoretical realm.

However, my discussion of case studies in the next two chapters is not directly related to the second question, rather it has one main goal: to settle the dispute over whether the structure of realisation is one-to-one or many-to-one. I showed in Chapter 4 that the Flat Objection does pose a problem for the Subset View, and one way to show whether it can be overcome or dismissed is to assess our best scientific understanding of the structure of the relations between scientific higher-level properties and their realisers. To do that, I have researched two pivotal (for different reasons) case studies: the emergence of thermodynamic properties from systems of particles, and the emergence of consciousness from systems of neurons. The first case study is pivotal because there is a relatively straightforward mathematical framework, statistical mechanics, that explains how thermodynamic properties arise from collections of particles. Hence, this case study will allow me to provide a detailed and empirically informed discussion of the relation between higher-level thermodynamic properties of the system and the properties of its components. The second case study is pivotal because it involves a scientific theory of consciousness and mental properties, the Global Neuronal Workspace, which is what first fuelled the debate on emergence, as they are considered by many to be the most recalcitrant cases of emergence. Being able to account for those properties should be considered a great success by any physicalist emergentist. As I explained both Gillett and Wilson do engage with case studies, but as I will discuss in the remainder of this chapter, when it comes to testing whether their accounts can actually be applied to special science features, their discussions do not have enough depth or are inconclusive. I focus on special science features because these are the kind of properties both accounts aim to account for, and because they are taken to be the main motivation for physicalist emergence. The second question — whether there are empirical cases of physicalist emergence — is only tangential to this discussion. The aim of the case studies is to show whether they have the necessary structure to apply the subset view — that is, if there is scientific support for the genuine existence of middleman properties and realisation_{O-O}. Or if on the contrary, their structure is only compatible with realisation_{M-O} as Gillett claims. Whether the subset view can actually be applied or not would require further discussion. In the case of the Global Neuronal Workspace, I will venture to explain in what way the subset view could be applied.

The discussion of case studies by Gillett is unable to answer the second question for his view, that is whether there are empirical cases of mutualism. The first case study presented in his book, which I discussed in Chapter 4, is about globules and globs: it is a thought experiment explicitly introduced as a general case of how his view works and it does not aim to answer the

second question. The case studies he presents that do aim to answer whether there are empirical cases of mutualism are high-energy superconductivity, eusocial insects, and biochemical networks. However, they end up being inconclusive: they are compatible with Gillett's scientific emergence, mutualism, but also compatible with reduction:

Overall, although my own sympathies lie with Mutualism, my final conclusion is consequently that we are not presently justified in accepting a Mutualist hypothesis about a specific scientific case. (Gillett, 2016, p. 352)

[T]he next stage of the debates over reduction and emergence needs to be focused locally (...) that is, in the careful examination of specific scientific cases of successful compositional explanation. Such work may consist in looking for new empirical evidence that can resolve the key issues between the live positions, or seeking to show that our existing evidence, when marshalled more carefully, can resolve which position provides the best account of some case. (Gillett, 2016, p. 354)

I agree with Gillett's suggestion that research needs to focus on the examination of specific scientific case studies. But I also think that his account of composition, rather than just mutualism, needs to be scrutinised in the context of these case studies. The case study Gillett presents to illustrate the many-one relation between components and composites —the diamond case study— does not delve deeply enough into the scientific nuances that explain the connection between the relata. For example, it overlooks the potential implications of quantum mechanical phenomena within the components, as well as any statistical considerations, which are typically considered fundamental in scientific research to the understanding of any micro-to-macro relationship.

Let us move on to discussing Wilson's presentation and use of case studies. Throughout her work, Wilson discusses many case studies. She devotes four chapters of her 2021 to discussing whether complex systems, ordinary objects, consciousness, or free will are cases of metaphysical emergence, either weak or strong. She concludes that: 1. Some complex systems are at least weakly emergent, 2. Ordinary objects are weakly emergent, 3. Consciousness is at least weakly emergent, and 4. Free will can be either weakly or strongly emergent depending on how we define it. However, only in the first of those four chapters does she discuss scientific case studies. In this chapter Wilson presents the degrees of freedom (DOF)-based account of

weak emergence, which I briefly introduced in Chapter 2. Wilson argues that the DOF-based account of weak emergence is the best one to account for complex systems. This is because complex systems offer clear examples of elimination of degrees of freedom. Wilson discusses several examples of elimination of degrees of freedom (2010, 2013, 2021): flocking birds, the Game of Life, a spherical conductor, complex systems undergoing phase transitions, and ordinary objects governed by Classical Mechanics. Those kinds of entities are considered to be weakly emergent by lights of a DOF-based account. However, I believe that the application of the DOF-based account to those cases presents difficulties that are not addressed in these examples, nor in the most comprehensive presentation of the DOF-based account in Wilson, 2010. I cannot delve into each example to explain how the problems unfold, so for simplicity I only assess the case of ordinary objects. The same kind of reasoning can be applied to the other cases.

To apply a DOF-based account of weak emergence to ordinary objects we need to identify the e_i that constitute the emergent entity E , and the degrees of freedom that are eliminated. Then, we need to identify the relational entity e_r (referred to as “system of e_i ” in 2021) that is the realisation base. In this case, e_i are quantum entities that are governed by quantum laws and have quantum features, such as colour charge, or spin. Then e_r just is a system of e_i , and E is an ordinary object, say a rock, governed by the laws of Classical Mechanics.

[C]onsider DOF associated with quantum spin. Classical objects are ultimately composed of quantum entities, whatever exactly these entities may be; but the characteristic states of classical objects do not functionally depend on the spins of the quantum components of these entities. Hence (...) it remains the case that DOF such as quantum spin are eliminated (and not just reduced or restricted) from those needed to characterize entities of the sort appropriately treated by classical mechanics. (Wilson, 2021, p. 194)

The degrees of freedom needed to describe entity E —the rock— are less than those required to describe e_i , insofar as the characteristic states of the rock do not functionally depend on the spins and colour charge, among other properties, of the e_i that compose it. This is uncontroversial. However, what about e_r ? Does the system of e_i realising E have the same degrees of freedom as E or as e_i ? Given the Proper Subset of Powers Condition, if E is to be a weakly emergent entity, e_r must have all the powers E has, and more.

[C]onsider E and the relational entity e_r that singly realizes it. Given that what powers an entity has are expressed by the laws in the science treating it, the powers of E are those expressed by the laws in the extracted theory treating E , while the powers of e_r are those expressed by the laws in the more fundamental theory treating the e_i (and any associated relational entities). It follows that E has a proper subset of the powers had by e_r . For example, suppose e_r is a quantum relational entity, and E is a macro-entity singly realized by e_r . Then the powers of E include all those powers to produce, either directly or indirectly, effects that can occur in the constrained circumstances in which the quanta form macroscopic entities (in other words: in the macroscopic limit). The realizing entity e_r has all these powers, and in addition has all those powers to produce, either directly or indirectly, effects that can occur in circumstances that are not so constrained, and in which quantum physics is operative—for example, effects occurring in circumstances involving temperatures or energies in which atoms, but not molecules, can exist. Hence E has only a proper subset of the powers of e_r .

Let me explain how the framework presented in the quote fits the case study of ordinary objects I introduced before. The object we are considering is a rock. The rock is realised by the system of quantum entities that compose the rock. The rock is governed by Classical Mechanics' laws, and the system of quantum entities is governed by Quantum Mechanics' laws. Each token power the rock has is numerically identical to a token power had by the system of quantum entities. For instance, both the rock and the system of quantum entities have the power to break a window (this could be associated with a kinematic degree of freedom, that is the same for both E and the system of e_i). The powers the system of quantum entities shares with the rock are those powers it has under constrained conditions —namely, when the quantum entities are related in such a way that they are tied together composing the rock. Crucially, the system of quantum entities also has powers not had by the rock. What are those powers? According to Wilson, those powers are associated with the DOF of the unconstrained entities e_i . This claim, however, is controversial. When the constraints that eliminate degrees of freedom of E and make e_r a realiser of the rock are in place, e_r cannot manifest those extra powers. And when those constraints are not in place, it is unclear that e_r exists. The justification of e_r 's existence as an entity was precisely the existence of those constraints, which unify a collection of unrelated e_i into a single relational entity. I now assess Wilson's argument that concludes that e_r has more powers than E , even though e_r does not exist if the constraints are not in place.

Wilson's argument relies on there being a link between powers and laws, and between laws and degrees of freedom. I will not assess those assumptions here, as I am with Wilson that it is

plausible that laws track what powers entities have, and also that laws determine the degrees of freedom of the entities that can enter in them. So if an entity can enter into a law there is sufficient reason to consider it as having at least as many degrees of freedom as any other entity that can enter into that same law. Wilson’s argument for e_r having more powers than E can be found in 2010, pp. 300-301 & 305; 2015, p. 387; 2021, pp. 181-182 with slight variations. I reconstruct it as follows:

1. Degrees of freedom are the parameters needed to “to characterize (states relevant to) an entity or system’s law-governed properties and behaviour”.
2. The system of laws treating e_i and their e_i -relations governs the e_i when constraints are operative and when they are not operative.
3. Given (2), the system of e_i can *evolve* into states where the constraints needed for it to realise E are not in place. “A quantum system realizing a macroscopic entity may evolve, as per the quantum laws, to a (perhaps different) system of differently constrained or unconstrained quantum e_i .”
4. This entails that to characterise e_r we must include the degrees of freedom needed to characterise its evolution into a (perhaps different) unconstrained system. “These DOF will include, for example, spin DOF for each of the quantum e_i constituting the realizing system, since e_r ’s evolution, as per the quantum laws, into different systems of e_i is functionally dependent on these spins.”
5. Hence, e_r has more degrees of freedom than E .
6. Given the relation between DOF and powers, it is concluded that e_r has more powers than E .

[T]he argument here, like the argument from different DOF, does not require that e_r be able to exist without E ’s existing: For e_r to have powers to bring about effects in circumstances not permitted by the constraints on E only requires that e_r be lawfully connected to goings-on that can exist in such relatively unconstrained circumstances. The realizing entity e_r is so lawfully connected; E isn’t; hence e_r has more powers than E . (Wilson, 2010, p. 305)

The aim of this chapter is not to thoroughly assess the DOF-based subset account, but to point out some issues it may face and to motivate the need for a thorough discussion of case studies. I argue that Wilson’s line of reasoning for the claim that e_r has more powers than E is problematic in at least three ways: first, only in the unconstrained scenario, where E —e.g.,

ordinary objects— does not exist, the powers that e_r has that E does not have will be able to be manifested; second, those extra powers are not directly powers of e_r but of the quantum entities e_i in the system; third, the ontological status of e_r is unclear and thus it is vulnerable to Gillett's Flat Objection.

The first issue might not be deemed problematic by the subset strategist, but I believe it nonetheless undermines her view: e_r and E can *manifest* the same powers. The only powers that distinguish them will only be manifested in circumstances where e_r does not exist. That is certainly an undesirable consequence of the DOF-based account of weak emergence. The second problem is the following: the powers e_r has that E does not have are powers of e_i rather than e_r . For instance, Wilson talks about the power linked to the “spin DOF for each of the quantum e_i ”. One simple way to see that is that those powers can be manifested when e_r does not exist. This is problematic for a one-one approach, as the e_i themselves are not realisers, but e_r is. Wilson must therefore also attribute the powers of e_i to e_r : “the system of e_i , but not E , has token powers associated with quantum spin” (Wilson, 2021, p. 181). She needs this to be the case so that the system of e_i , e_r , has more powers than E . So clearly there is a relation between the powers of e_r and the powers of e_i . However, the nature of that relation is unclear. How does e_r get to have e_i 's powers? I believe that Wilson's case studies do not provide enough detail nor engagement with the science to be able to explain how to obtain the powers of the system of e_i , e_r , from the powers each of its component entities e_i have. This consideration takes me to the third problem.

Prima facie, an example like the following seems perfectly explainable in terms of many-one composition relations: in a ferromagnet, the spins of many individual electrons align and create a macroscopic magnetic field. The macroscopic magnetic field is realised, a la Gillett, by the joint role-filling of many individual electrons with aligned spins. There does not seem to be a need to introduce a further entity, a system of electrons, as being the single realiser of the macroscopic magnetic field. However, this precludes the possibility of the subset relation applying in this case. With thorough examination, we can provide an empirically informed view of what e_r is and what powers it has. These are questions about the nature of e_r that are left unanswered. A detailed discussion of the science can lead to a stronger justification of the existence of the relational entities posited.

I believe there are reasons to consider Wilson's discussion of case studies insufficient. The case studies most comprehensively presented are those related to the DOF-based view, yet they still lack the necessary detail to definitely determine if the subset view is applicable in these instances. In particular, I have introduced and discussed three problems for her approach. In the following two chapters, however, I leave aside the first two problems, to focus on the third: the ontological status of e_r , or, in general, the realisation base in realisation_{O-O} . The aim of the case studies is to demonstrate that, with a sufficient level of detail and thoughtful examination, the third issue can be overcome. Whether the DOF-based is a plausible implementation of the subset view remains open.

One final note. With regards to consciousness, Gillett excludes it from study, and Wilson's discussion of it is purely theoretical, in the sense that it does not assess scientific theories of consciousness. I think mental states and mental properties should, at least with regards to them being scientifically/weakly emergent, be treated as special science states and properties, so some consideration should be given to the best scientific theories we have to explain them. The aim of Chapter 7 is to examine, with a metaphysical eye, one of the best scientific theories we currently have to scientifically explain consciousness: the Global Neuronal Workspace Theory.

I have shown that the case studies as presented by Wilson and Gillett are problematic for different reasons. In the following two chapters I provide a detailed account of two case studies that will be used to test the metaphysical theories introduced in the previous chapters. Although the thorough discussion of the case studies using our current best scientific accounts requires technicalities at points, I have endeavoured to balance an in-depth discussion with clear and general explanations for a reader not familiar with the scientific literature.

CHAPTER 6. CASE STUDY I: ENTROPY

6.0 Introduction

The aim of this chapter and the following one is to assess the views on metaphysical dependence relations I have introduced in the previous chapters. In Chapter 2 I introduced and assessed Jessica Wilson's subset account. In Chapter 3 I introduced Sydney Shoemaker's version of the subset account and concluded that Wilson's account should be preferred. I also criticised his microphysical realisation relation. Finally, in Chapter 4 I introduced Carl Gillett's compositional account and his Flat Objection against the subset view. In that chapter I assessed Wilson's response to the Flat Objection, what I called the Middleman Strategy, and I explained where it goes wrong: the existence of middleman properties is not sufficiently justified. The conclusion was that Wilson's response to the Flat Objection is not compelling enough, and that to make it compelling, the positing of middleman properties needs to be justified. The goal of this chapter, and the next one, is to present and thoroughly examine two scientific case studies in order to determine whether middleman features are posited by the sciences, and if so, what role they play. If it turns out that the scientific case studies use a middleman property in the derivation or account of higher-level properties, then this will provide grounds to conclude that the one-one account is able to subsume the many-one account.

The first case study is the realization of a higher-level property of thermodynamic systems, entropy, by the properties of the particles composing the system. This is an interesting case because we can actually define (and derive) entropy, a higher-level property of a system, from the lower-level properties of its constituents. By providing a detailed explanation of this derivation, I will be able to offer an empirically informed discussion of the nature of higher-level properties of thermodynamic systems and how they are related to the properties of the particles that compose it. This discussion will go beyond testing Gillett's Flat Objection against the subset view, as it will also question the subset view itself.

6.1 From Large Sets of Particles to Thermodynamic Systems: The Case of Entropy

In this section I offer a detailed explanation of how statistical mechanics defines (and derives) entropy, which is a thermodynamic —and thus macroscopic— property of systems, from the lower-level properties of the entities composing those systems. But before I start, what is

entropy? A simple answer is that entropy is a measure of randomness or disorder in a system. The higher the entropy, the more disordered the system is. The definition of entropy, as stated in the Second Law of Thermodynamics, is the following:

The Second Law states that an extensive state function, the total differential $dS = dQ/T$, defined only for equilibrium states, is such that $\Delta S \geq \int \delta Q/T$. Loosely put, for realistic systems, this implies that in the spontaneous evolution of a thermally closed system, the entropy can never decrease and that it attains its maximum value for states at equilibrium. (Callender, 1999, p. 350)

An extensive state function is a property that depends on the amount of material in the system. Entropy is such a function because it depends on the number of particles in the system, among other things. The total differential equation, $dS = dQ/T$, defined only for equilibrium states, relates an infinitesimal change in entropy (dS) to an infinitesimal amount of heat added or removed from the system (dQ), divided by the absolute temperature (T).

To illustrate this principle, I introduce an example. Consider a closed system consisting of a room with a cup of hot coffee. The coffee and the room are both part of the system, which is thermally closed. This means no heat enters or leaves the system, though heat can be exchanged between the coffee and the room. At the beginning, the coffee is hotter than the room. Heat, which is a form of energy, will naturally flow from the hotter coffee to the cooler room. This is an example of energy dispersal, with the heat spreading out from a concentrated source to a larger, cooler area. According to the formula $dS = dQ/T$, the entropy change when a small amount of heat dQ flows from the coffee to the room will be dQ divided by the temperature at which the heat transfer occurs. If we sum up all these small entropy changes, we get the total entropy change for the process. The total entropy of the system will be the sum of the entropy of the coffee and the entropy of the room.

Now, the room gains entropy because it is absorbing heat, so dQ is positive for the room. The temperature T in the denominator is the room's temperature, which is lower than the coffee's temperature. Hence, a given amount of heat produces a larger increase in entropy (since dividing by a smaller number gives a larger result). The coffee, on the other hand, loses heat, so dQ is negative for the coffee. But the coffee's temperature T in the denominator is higher. So, a given amount of heat loss produces a smaller decrease in entropy (since dividing by a

larger number gives a smaller result). Because of this, the room's entropy increases more than the coffee's entropy decreases. Hence the total entropy of the coffee-room system increases. This reflects the Second Law of Thermodynamics.

As time passes, more and more heat will flow from the coffee to the room until they both reach the same temperature, a state known as "thermal equilibrium". At this point, no more spontaneous heat flow occurs. This state of thermal equilibrium also represents a state of maximum entropy for the coffee-room system. Energy (in the form of heat) is as spread out as it can be. For this reason, an increase in entropy is commonly known as an increase in disorder, as energy is more spread out the higher the entropy.

Importantly, while the fundamental laws of physics are time-symmetric, an irreversible thermodynamic process like the one described is not. Once the coffee and the room have reached the same temperature, they will not spontaneously return to their initial state. The heat will not spontaneously go back to the coffee. This gives us a way to distinguish the past (when the coffee was hot and the room cooler) from the future (when both the coffee and the room are at the same temperature). In the coffee-room system, the direction of time—also known as the "arrow of time"—is given by the increase in entropy. Some physicists propose a connection between the directionality of time and the second law of thermodynamics, specifically its stipulation that entropy increases in a system over time. According to this view, our perception of time moving from past to future is tied to the progression from states of lower entropy to higher entropy.

The behaviour of systems described by the second law has no direct parallel in the behaviour of microscopic particles, as in the mechanical description of individual particles, the paths these particles take are reversible and time symmetric. How can we reconcile the thermodynamic description of a system with the description provided by Hamilton's equations of motion for each particle of the system? Is the thermodynamic description reducible to the mechanical description? This is known as the Loschmidt paradox, or the paradox of irreversibility. The solution to this paradox, which is crucial to how statistical mechanics is able to derive thermodynamic properties from the properties of individual particles, is to conceive entropy as a function of a collection of systems, rather than as a function of the dynamical variables of an individual system. This is what the Gibbsian approach does, and what I go on to explain in the next section.

6.1.1 Gibbsian Entropy

A system composed by N particles, such as a gas, or a paramagnet, or a liquid, at the thermodynamic limit—that is, when the number of particles is so big that we can consider it to be infinite, $N \rightarrow \infty$ —is governed by macroscopic laws and defined by macroscopic properties. The macroscopic properties of the system, such as volume, entropy, energy, etc. define what is known as the macrostate of the system. A system in a certain macrostate is compatible with many different combinations of the microscopic states of the particles that compose it, known as microstates. Microstates contain all the necessary information to describe the state of each particle composing the system at a given instant (the position, momentum, spin, etc.). Roughly put, Gibbsian statistical mechanics shows that most macroscopic properties of the system can be obtained from averaging microscopic properties over all the available microstates. How do we obtain all the available microstates? For instance, if we have a free particle in one dimension, the particle is governed by two parameters, also referred to as degrees of freedom: one degree of freedom corresponding to position, x , and one corresponding to momentum, p_x . The phase space, usually represented as Γ , is a set of coordinates used to represent all the possible microstates. In this case, we can represent all the possible microstates in a space with two axes (known as a 2-dimensional phase space), x and p_x , where every point of the phase space corresponds to a microstate in which the particle can be found. If we have N particles in three dimensions, then we have a $6N$ dimensional phase space, as the position of each particle (x, y, z) and its momentum (p_x, p_y, p_z) now require three dimensions each, and we have N particles. Every point in this $6N$ dimensional phase space will correspond to a single microstate of the system. That is, every single point will provide a possible configuration of the system: it will determine the position and momentum of every single particle of the system.

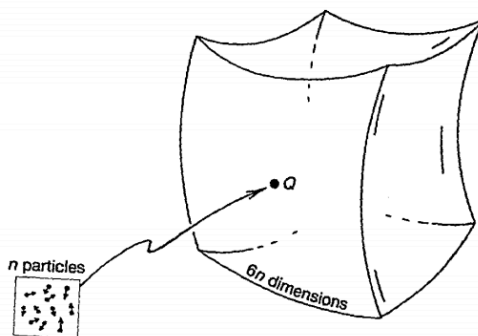


Figure 1. Represents a phase space with $6n$ dimensions and the content of each point of this phase space, which correspond to all possible microstates. Figure extracted from Penrose (1989, p. 177).

However, not all microstates are always available. Depending on external constraints of the system, such as energy or volume, a given subset of the points in the phase space, that is a subset of configurations, will be available. For instance, consider again a free particle in three dimensions with energy, E , fixed. The phase space has six dimensions. The energy of the particle is $E = \frac{|p|^2}{2m}$, so it is only a function of the total momentum, $|p|$, and the mass of the particles, m . The particle is thus constrained to be in the microstates compatible with this energy, and this is independent from its position. The number of microstates available is proportional to the surface of a sphere of radius $|p|$ in the three-dimensional phase space of momentum. Any point in the surface corresponds to a value for p_x , p_y , and p_z compatible with $|p| = \sqrt{2mE}$. An ensemble of microstates is a collection of all the microstates that are compatible with a macrostate, given a set of external constraints. In this case, the number of microstates, Ω , in the ensemble will be given by the surface of the sphere—in general, $4\pi r^2$ —, so $\Omega = 4\pi \cdot 2mE$.

There are three kinds of ensemble that are used in Gibbsian statistical mechanics, and it is assumed that a system in equilibrium will fall in one of these three cases: the microcanonical ensemble, where energy and the number of particles of the system are fixed; the canonical ensemble, where the system is allowed to exchange energy with a heat bath, so the number of particles is fixed but energy is allowed to vary; and the grand canonical ensemble, where both the number of particles and energy are allowed to vary. In this thesis I will focus only on the microcanonical ensemble and canonical ensemble for the sake of avoiding repetition, but the arguments are extensible to the grand canonical ensemble.

I now introduce the key concept of a probability distribution, usually represented as ρ , and the role of probability in Gibbsian statistical mechanics. A probability distribution is a function that gives the probability of finding the system in a certain microstate. Each of the three ensembles introduced above have a characteristic well-defined probability distribution. The probability distribution of the microcanonical ensemble is constant, given that each microstate is equiprobable. Considering a microcanonical ensemble with Ω microstates, the probability of a given microstate, say r , is: $\rho(\text{microstate } r) = \frac{1}{\Omega}$. In the canonical ensemble the system exchanges energy with the heat bath. The probability of each microstate will depend on the probability of the energy it has, ϵ . Energy is governed by the probability distribution $\rho(\epsilon) \propto$

$e^{-\frac{\epsilon}{k_B T}}$, where k_B is the Boltzmann constant, and T is the temperature. This is known as the Boltzmann distribution or the canonical distribution (Blundell and Blundell, p. 37). We can also calculate the probability of finding the system in a microstate r , with energy ϵ_r :

$$\rho(\text{microstate } r) = \frac{e^{-\frac{\epsilon_r}{k_B T}}}{\sum_i^{\Omega} e^{-\frac{\epsilon_i}{k_B T}}}$$

The denominator in this equation is the sum of the Boltzmann factors of each possible microstate, and is known as the partition function Z . In this case, it is used to normalise the probability of microstate r , i.e. to make sure that the sum of the probabilities of all microstates equals one.

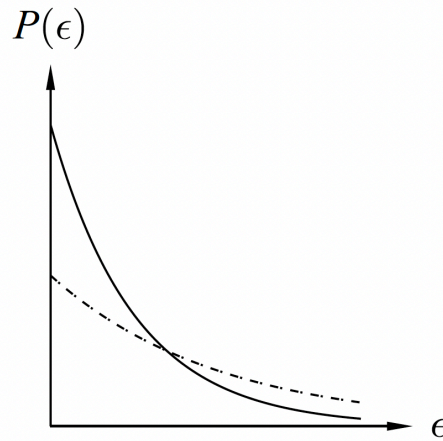


Figure 2. Graph showing the Boltzmann distribution for a system at two different temperatures. The probability of finding the system at a certain energy depends on temperature: at higher temperatures it is more probable to find the system at higher energies. Figure extracted from Blundell & Blundell (2006, p. 37).

One of the key results of Gibbsian statistical mechanics is that some of the properties of the macrostate of a system can be known by calculating the average values of the property at all the available microstates, x , of the phase space Γ . This is known as an ensemble average. Some properties, such as energy or magnetisation, are properties both of the macrostate of the system and its microstates, so we can directly calculate their average values, that is their value for the macrostate, from their value at each microstate and the probability of finding the system at a given microstate. The ensemble average of a macro-variable f , such as magnetization or internal energy, for an ensemble with a probability distribution ρ , where ρ is stationary, x represents a microstate, and Γ is the phase space, can be represented as (Frigg & Werndl, 2021, p. 6):

$$\langle f \rangle = \int_{\Gamma} f(x) \rho(x) dx$$

It is worth emphasising that, for instance, the energy of the system does not correspond to the energy of any of its particular microstates, but it is equal to the mean energy of all possible microstates. To further illustrate the statistical nature of the energy of the system I now briefly introduce a particular case (see Blundell & Blundell (2006, p. 41) for more detail). Consider a system that can only be in two microstates, one with zero energy, and the other with energy ϵ . At low temperatures, the mean energy of the system will be approximately zero, because there is very low probability to find it in the state of higher energy. However, at higher temperatures there is the same probability, $\frac{1}{2}$, to find the system in any of the two available levels, so the mean energy of the system is $\langle E \rangle = \sum_i p_i \cdot E_i = \frac{1}{2} 0 + \frac{1}{2} \epsilon = \frac{\epsilon}{2}$. This is illuminating because the energy of the occupied microstate at any instant is either zero or ϵ . However, the mean energy of the system, which corresponds to the macrostate, is $\frac{\epsilon}{2}$.

There are some other properties, such as temperature, pressure, chemical potential or entropy, that are statistical in nature. This means that we cannot attribute them to a single microstate: they are attributed to the macrostate by means of averaging over all possible weighted microstates. Of course, the fact that they can be attributed only to the macrostate does not mean that they are not derived from the microstates. Entropy is commonly known as the measure of disorder, as it is a function of the number of microstates available to a given macrostate. The properties of pressure, temperature, and chemical potential are deeply interconnected with entropy in thermodynamics. These properties can be expressed through partial derivatives involving entropy and other variables. Temperature, in the context of thermodynamics, is defined as the partial derivative of the internal energy with respect to entropy, holding volume and the number of particles constant. Pressure is the derivative of the entropy with respect to volume, keeping energy and the number of particles constant. Chemical potential, which is the change in a system's potential energy concerning a change in the number of particles of a specific type, can be expressed as the derivative of entropy with respect to particle number, while holding energy and volume constant (Swendsen, 2019, pp. 99-108). Technicalities aside, this means that all these thermodynamic properties have a very strong connection with entropy. They all involve entropy as a part of their mathematical expressions. For this reason, from now on I will focus on entropy.

Gibbsian entropy, S_G , is obtained by the following expression, where the integral is over the phase space Γ :

$$S_G(\rho(x)) = -K_B \int_{\Gamma} \rho(x) [\ln \rho(x)] dx \quad (2)$$

For the microcanonical ensemble, given that $\rho(x) = \frac{1}{\Omega}$, the solution of the integral is $S_G(\rho_{microcanonical}) = K_B \ln \Omega$, where $\ln$ is the natural logarithm.

Callender (1999), and Frigg & Weindl (e.g. 2021) discuss several objections against the idea that Gibbs entropy can be the reduction base of thermodynamic entropy. There are several refinements and alternative definitions of S_G that try to solve the objections. A recent approach is Wallace (2018). For the purpose of this chapter those modifications do not alter my overall discussion, as they still use the notion of ensemble. However, there is another approach to deriving entropy that could potentially affect my discussion: the Boltzmannian approach to entropy. In the next section I introduce Boltzmannian entropy.

6.1.2 Boltzmannian Entropy

Although the Gibbsian approach is the main one used in scientific contexts, as it allows for easier calculations, there is a growing interest within the philosophy of physics literature in Boltzmannian entropy. Advocates of the Boltzmannian approach argue that although the Gibbsian approach might be more useful in a scientific context where one needs to be able to obtain results and calculations, the Boltzmannian approach is conceptually more appealing. The main problem of the Gibbsian view according to Boltzmannians is the use of the notion of ensemble it requires, as they claim it is a purely mathematical notion that does not represent a physical property of the system: “In contrast to the Gibbs entropy, we define the Boltzmann entropy S_B for the actual microstate of an individual system” (Callender, 1999, p. 355).

To explain the Boltzmannian approach to entropy I draw on some of the concepts introduced in the previous section. Again, we need to consider Γ to be the phase space of a system, and x to be a microstate in Γ . As macrostates supervene on microstates, there is one and only one macrostate, say $M(x)$, that corresponds to x . In turn $M(x)$ can be realised by other microstates. The first step to obtain Boltzmannian entropy is to know the volume of microstates corresponding to $M(x)$ in Γ . To do so, we partition Γ into compartments with macroscopically

indistinguishable properties, so that each compartment contains all the microstates compatible with those macroscopic properties. Take $|\Gamma_{M(x)}|$ to be the volume of the compartment containing macroscopically indistinguishable microstates compatible with $M(x)$, which is the macrostate corresponding to microstate x . The Boltzmannian entropy is given by:

$$S_B(x) = K_B \ln |\Gamma_{M(x)}| \quad (4)$$

Gibbsian entropy and Boltzmannian entropy not only are defined using different mathematical tools, but also have different conceptual underpinnings: S_G is a function of a probability distribution over an ensemble of microstates, while S_B is a function of a microstate of a system.

6.2 Thermodynamic Properties Are Causally Efficacious

That some thermodynamic properties are powerful is usually intuitively true without problem. For instance, touching water at high temperature burns our skin, and low pressure causes our boiler to malfunction. However, the causal efficacy of entropy is not so easy to grasp intuitively. The aim of this section is to show that there are sufficient reasons to consider entropy a causally efficacious property of thermodynamic systems.

Below I provide a particular example where entropy is causally efficacious, and then I go on to provide a generalisation. It has a parallel with the example of the cup of coffee cooling down, but heat is not directly involved in explaining the phenomenon:

- **Perfume in a room:** When you open a bottle of perfume, the perfume molecules are not evenly spread out in the room. Instead, they are initially concentrated in one small area (inside the bottle). As time passes, the perfume molecules begin to spread out and mix with the air in the room. This can be seen as a result of the relatively lower entropy state of the perfume initially, which causes the perfume to spread out. The process of diffusion, which involves the perfume molecules spreading out and increasing the system's entropy (where the system is the perfume + the air in the room), continues until the system reaches a state of maximum entropy, i.e., the perfume is evenly spread throughout the room.

- Generalisation: For any system R composed of two substances A and B not in contact, when brought into contact, there will be a diffusion process (X) from the region of low entropy (say, A) to the region where the system's entropy is greater, which is to diffuse into B. We can say that A having lower entropy than A+B caused the diffusion process X of A into B. This is because the combined system of the two substances not mixed has lower entropy than it would in a state of complete mixture. The diffusion process increases the entropy of the combined system until it reaches a maximum, corresponding to a state where A and B are evenly mixed throughout.

Using a difference-making account of causation, or a counterfactual account of causation, this case shows that entropy is causally efficient: if the system where A and B are mixed had had lower entropy than A not being mixed, the particles would not have moved from A's region into B's region. It is because A's region has relatively lower entropy that A's particles diffuse into B. If the reader is not persuaded that entropy is causally efficacious, I want to mention that the crucial aspects of the discussion that follows in section 6.3 apply to the case of temperature, or to any other property of a system whose derivation from the properties of the system's components requires using a statistical ensemble (with regards to Boltzmannian entropy, I will also explain that it does require something similar to an ensemble). To strengthen this point: under the microcanonical ensemble, temperature is derived from entropy and internal energy. However, due to the added mathematical as well as conceptual complications to understanding temperature and how it is derived from the properties of the components of the system (as its derivation requires understanding how entropy is derived), I have restricted the case study to entropy.

6.3 Testing Wilson's Subset View and Gillett's Mutualism Against the Statistical Mechanics Case Study

More often than not, when discussing metaphysical dependence relations philosophers do not take into account the actual derivations and definitions of higher-level properties in terms of their lower-level realizers that the sciences provide. As I have explained, thermodynamic systems offer a range of interesting yet sufficiently simple derivations that should at least be considered when discussing dependence relations. In this section I will use the case of entropy to assess Gillett's and Wilson's views. In Section 6.3.2 I argue that the case of entropy introduced above provides enough grounds to question Gillett's Mutualist view. On the other

hand, in Section 6.3.3 I argue that the case of entropy seems to favour a version of Gillett's Flat Objection against Wilson. In this section I therefore explain the objection and how it differs from Gillett's original worry, and I present a new account of Wilson's middleman properties that is able to overcome it.

As I have explained, Gillett's many-one view argues that a higher-level property of a system, entropy in this case, is realised by lower-level properties of the entities composing the system. On the other hand, Wilson's one-one view argues that a higher-level property of a system is realised by a single lower-level relational property of the aggregate of entities composing the system. There are four scenarios to consider: 1. both the Boltzmannian approach and the Gibbsian approach favour only one and the same view (be it mutualism or the subset view), 2. the Boltzmannian approach favours one view and the Gibbsian approach favours the other, 3. none of the approaches favour any of the views, and 4. one and only one of the approaches favours one and only one of the views.

In the literature, there is no clear resolution as to what approach is better, the Gibbsian or the Boltzmannian, and there are as many objections as solutions on either side. In this thesis I will not pick sides or try to settle the debate, rather I would like to discuss which of the four scenarios above is the case. Depending on the conclusion of this discussion, it might turn out that the case study is not helpful in deciding which of the two views, Gillett's or Wilson's, is most promising to account for the relation between thermodynamics and statistical mechanics. However, I can advance now that the case study is not neutral.

A brief note: someone might think that probabilities and statistical considerations are not objective features of SM, rather subjective or due to our incomplete knowledge of the system. Discussion of the objective nature of probabilities in SM is a very interesting debate that I cannot have here. However, I want to mention that there is a strong argument from quantum mechanics that supports the objective nature of probabilities and the ensemble.

Probability is already inherent in the underlying quantum mechanics: it is not an additional ingredient. This is because the state of the system in quantum mechanics is not represented by a point in phase space but by a density matrix $\hat{\rho}$, (Baierlein, 1971, Ch. 12). (...) [P]robabilities in SM stem from probabilities in QM. (Robertson, 2019, p. 75)

The density matrix $\hat{\rho}$ represents the state of a system in quantum mechanics and is inherently probabilistic, as it considers all the possible states the system can be in and their probabilities. Therefore, if quantum mechanics were introduced in this case study to interpret SM, it would make it even more appealing: it would provide an explanation of the nature of ensembles and SM probabilities as objective features of systems.

Before diving into a discussion about metaphysical dependence relations, I need to bring back the previous debate to the metaphysical realm. Literature in the philosophy of physics does not usually tackle the metaphysical questions I am concerned with in this thesis, so given all the concepts and derivations that have been discussed so far, I need to offer an account of what the realisation base of entropy is according to the Gibbsian account and the Boltzmannian account.

6.3.1 An Empirically Informed Account of the Realisation Base of Entropy

The first step to decide which scenario is the case is to translate into metaphysical terms, or in terms closer to Wilson and Gillett's views, the relation between thermodynamic properties and statistical mechanics properties that both the Gibbsian and Boltzmannian approaches suggest. Given that the Gibbsian and Boltzmannian approaches differ in some important respects, I will discuss them separately here. It will turn out, though, that they posit the same or a similar kind of relation between thermodynamic and statistical mechanical properties. In this section I will continue focusing on entropy instead of looking for a general description of the relation between thermodynamic and statistical mechanical properties.

Realisation base of Gibbsian entropy

As I introduced in section 6.1.1, Gibbsian entropy is a function of a property of the ensemble, the probability distribution, ρ .

$$S_G(\rho(x)) = -K_B \int_{\Gamma} \rho(x) [\ln \rho(x)] dx$$

I think that one clear candidate for being the realisation base of Gibbsian entropy is the ensemble, as entropy is obtained from integrating a function of the probability distribution over the phase space —i.e. the ensemble. The realiser is then the property of being an ensemble with probability distribution $\rho(x)$. This is compatible with the intuitive understanding of entropy as a measure of disorder: entropy measures the number of microstates that are possible within certain macroscopic constraints. Including the weight of the microstates in the ensemble is

motivated not only because of the Gibbsian definition, but also because it is intuitively compelling: two ensembles containing the same number of microstates would have different measures of disorder if, say, in one of them the probability to find the system in one microstate is much bigger than the probability to find it in any other microstate, while in the other the microstates are equiprobable. The latter will be much more disordered than the former, even if they have the same number of possible microstates. An ensemble is a particular system of a certain type, which is defined partially modally. For those that consider that treating the ensemble as an entity is problematic, there is another candidate for being the realisation base of entropy that is a property of microstates themselves: the property of being a member of the ensemble with probability distribution $\rho(x)$. In this case, entropy is realised by a property which is defined in terms of the whole system, and which all microstates have. So even if the ensemble is not considered a particular, we still need the global property that refers to it: microstates can be the relevant particulars but still they need to have the property of being a member of an ensemble.

Realisation base of Boltzmannian entropy

Boltzmannian entropy is defined differently. The expression, as introduced in section 6.1.2, is the following:

$$S_B(x) = K_B \log |\Gamma_{M(x)}|$$

Crucially, Boltzmannian entropy is a property of microstates, and can be obtained without introducing the notions of ensemble and of the probability of microstates. However, although Boltzmannian entropy is defined as a property of microstates, that is, a property of the aggregate of entities composing the system, the key term $|\Gamma_{M(x)}|$ in the above expression refers not to properties of the microstate but to a volume of microstates. It contains all the microstates compatible with macrostate $M(x)$, which is the macrostate corresponding to the microstate whose entropy is calculated. Given the definition, the most plausible realisation base for Boltzmannian entropy is the property of being in the volume of microstates corresponding to the macrostate $M(x)$, namely $|\Gamma_{M(x)}|$. Although the Boltzmannian approach does not explicitly use a probability distribution, it is implicitly assuming that all microstates within this volume are equiprobable. Moreover, the realisation base of Boltzmannian entropy has an important similarity to the realisation base of Gibbsian entropy. They both have a modal element: both realisers are properties that refer to a global feature, be it an ensemble or a volume of microstates, that involve *possible* microstates compatible with a given macrostate. To

strengthen the point that both approaches are very closely related, I would like to mention that in the literature there is some agreement that the Gibbsian approach subsumes the Boltzmannian approach.

The Gibbsian and Boltzmannian approaches are not rival approaches to the statistical mechanics and thermodynamics of the same systems: rather, the Gibbsian framework is a more general framework in which the Boltzmannian approach may be understood as a special case. (Wallace, 2018, p. 2)

6.3.2 Testing Gillett's Mutualism

To decide which of the four scenarios is the case, we need to consider Gillett's Mutualism in relation to Gibbsian entropy and Boltzmannian entropy. Entropy is a higher-level property of thermodynamic systems, which are composed by particles, and is derived differently under the Gibbsian and Boltzmannian accounts. I will begin addressing whether Gillett's view can account for Gibbsian entropy. As I have introduced in the previous section, one plausible candidate to be the realisation base of Gibbsian entropy is the property of being an ensemble with a probability density $\rho(x)$.

6.3.2.1 Can Gillett's View Account for the Realisation of Gibbsian Entropy?

Let me emphasise that the ensemble takes into account all the possible microstates in which the system can be found, and the probability distribution assigns a probability to each possible microstate. Therefore, the ensemble and the probability distribution cannot be defined in terms of the actual properties of the individual particles composing the system at a given time—for instance, it cannot be defined in terms of the positions and momenta of all the particles of a given microstate. Let me illustrate this with an example where the actual properties of the object at a time do realize the object at a given time.

Consider a table. The parts of the table —say its four legs, some screws, and the top—constitute the table. Arguably the properties of the parts of the table realize the properties of the table. For instance, it is in virtue of the legs being 75cm high that the table is 75cm high, it is in virtue of the legs being able to hold 5kg each that the table can hold 20kg, and so on. At any instant, the properties of the table can be obtained from a combination of the properties of

its components. For this reason, we say that the properties of the parts of the table synchronically realize the properties of the table. Now, we can see the contrast between this case and the case of Gibbsian entropy.

Under the Gibbsian interpretation, to explain a certain thermodynamic system in terms of the particles that compose it we need to introduce the notion of ensemble. So there is no single combination of the properties of the particles of a thermodynamic system at a given time that will result in its Gibbsian entropy. Rather, we need to consider a close to infinite number of possible variations of that same system and their associated probability. The notion of ensemble contains an element of *modality*: to obtain a certain thermodynamic property of a system, say entropy, we cannot just consider the actual system, but we need to consider a close to infinite number of possible microstates the system could be in.

As I thoroughly discussed in Chapter 4, Gillett's account describes realisation in the sciences as a relation between many property instantiations at the lower level and a single higher-level property, so realisation has a many-one structure similar to the example of the diamond introduced in section 4.1, or to the example above, where a property of the table is realised by the properties of its parts. The case study of entropy provides a counterexample to this view for two reasons: 1. Gillett only deals with actual properties of the lower-level entities composing a higher-level entity. However, in the case of Gibbsian entropy, the realisation base is not an actual property of entities, but a modal property of the system. 2. Gibbsian entropy is realised by a single property, rather than by many properties.

In what follows I discuss how Gillett's view could avoid these two problems. First, we could introduce a modification in Gillett's view to account for the realisation of Gibbsian entropy: modal properties of entities can also serve as property realisers. However, this modification alone would not be sufficient, as the realisation relation involved in Gibbsian entropy is a one-one relation: a single lower-level property, namely the property of being an ensemble with a probability density $\rho(x)$, realises the property of having Gibbsian entropy S_B . This is incompatible with his many-one approach. To avoid this problem, Gillett would need not only to modify his account to include modal properties of entities but he would need to argue that there is a many-one structure underlying the realisation of Gibbsian entropy —i.e. that the single realisation base proposed can actually be put in terms of many properties of individuals so that we recover a many-one relation. Once this argument is in place he could further argue

that for the sake of ontological simplicity and given that it adds nothing to the realisation of Gibbsian entropy, there is no reason to postulate the existence of the single realiser property – the ensemble—, so a one-one approach fails.

How could we put the property of being an ensemble with a probability distribution $\rho(x)$ in terms of properties of the many components of the system? One way to answer this question is to argue that the components have modal properties that contain information about their possible states, just as the ensemble contains information about the possible configurations of the system. For instance, a particle could have a modal property that contains information about its possible energy states and their probabilities. If every particle of the system has a modal property of this kind, Gillett could argue that we can actually obtain the entropy of a system from the collection of modal properties of its components. That is, instead of being realised by a single property determining all the possible configurations of the system and their probability, Gibbsian entropy is realised by modal properties of the particles composing the system, where each modal property refers to the possible states of the particle it is a property of. Let me show how this would work in a toy example, extracted from Blundell & Blundell (2006, p. 328).

Imagine a particle that can be in two different states, $|0\rangle$ and $|1\rangle$. Now let us consider a system of two such particles, say A and B, with its states described as the product state $|0\rangle|1\rangle$. For instance the product $|0\rangle|1\rangle$ describes a system in which the first particle is in state $|0\rangle$ and the second particle is in state $|1\rangle$. Before moving on with the example I need to introduce a distinction between particles that is used in mechanical statistics. Although drawing this distinction has not been necessary so far, it will be key to decide between a one-one or many-one approach. The distinction is the following: particles can be either distinguishable or indistinguishable. Distinguishable particles are either different from each other by their nature or physically similar but still distinguishable due to other reasons. Indistinguishable particles, on the other hand, are identical in every aspect and cannot be differentiated from one another. If you have a state with two indistinguishable particles, you can swap the two particles and it will not change anything physically observable from that state –i.e., the two states are physically equivalent.

This distinction will prove key to undermine Gillett's many-one approach, as I will argue that although his view can actually account for the case of distinguishable particles, it is unable to

account for the realisation base of entropy in terms of the properties of indistinguishable particles. Continuing with the toy example, we know now that the particles can be either distinguishable or indistinguishable. If they are distinguishable, and assuming as I have through this thesis that they are classical particles, there are 4 possible configurations of the two-particle system, which are the following:

$$|0\rangle|0\rangle \quad |0\rangle|1\rangle \quad |1\rangle|0\rangle \quad |1\rangle|1\rangle$$

In this case, each particle has the property of being in possible states $|0\rangle$ or $|1\rangle$, and we can obtain the 4 possible configurations by combining A's property of being in possible states $|0\rangle$ or $|1\rangle$ with B's property of being in possible states $|0\rangle$ or $|1\rangle$. Therefore, there is indeed a combination of many properties of particles that yields all the possible configurations of the system. Assuming that we could obtain entropy for this toy example using the Gibbsian framework, a one-one approach, which is the one I have introduced as the most plausible, would argue that Gibbsian entropy is realised by the ensemble of possible configurations and their probability. A many-one approach would argue that Gibbsian entropy is realised by the combination of modal properties of the particles that contain information about their possible states. In this case, both approaches work.

Let us now move on to the case where particles are indistinguishable. In this case, the states $|0\rangle|1\rangle$ and $|1\rangle|0\rangle$ are indistinguishable from each other. If the particles are indistinguishable and classical, the configuration $|0\rangle|1\rangle$ of the system is the same as the configuration $|1\rangle|0\rangle$. So there are only three possible configurations:

$$|0\rangle|0\rangle \quad |0\rangle|1\rangle [= |1\rangle|0\rangle] \quad |1\rangle|1\rangle$$

In this case, a combination of each particle's properties about the possible states it can be in does not yield the expected result. The particles have the same properties as in the previous case: A has the property of being in possible states $|0\rangle$ or $|1\rangle$ and B has the property of being in possible states $|0\rangle$ or $|1\rangle$. However, as we have seen if we combine those properties, we get four configurations rather than three, so the ensemble is not a simple combination of the possible states of each particle. For this reason, Gillett's many-one view cannot account for cases where the particles are indistinguishable: there is no way to obtain the ensemble and its

properties only considering the properties of particles. On the other hand, I believe that the one-one approach can account for this scenario because in this case the realisation base of Gibbsian entropy is the property of being an ensemble with a probability distribution, as I have established, and it does not really matter how the ensemble is obtained as long as it is a non-emergent combination of lower-level properties and relations.

That said, I have accepted that Gillett's many-one view can account for the case of distinguishable particles. However, this is a pyrrhic victory because once we move to quantum mechanical statistics, all particles are indistinguishable. There are two kinds of particles: bosons and fermions. Bosons are indistinguishable particles, and they have the particularity that several particles can be in the same state. Fermions are indistinguishable particles, and they obey Pauli's exclusion principle, so there cannot be two particles in the same state. If we go back to our toy example, the possible configurations for each case are the following (I will not offer details of how these results are obtained, as it is not necessary to understand the implications, and it is beyond the scope of this thesis).

$$\text{Bosons: } |0\rangle|0\rangle \quad \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) \quad |1\rangle|1\rangle$$

There are three possible configurations of the two-particle system. In $|0\rangle|0\rangle$ and $|1\rangle|1\rangle$ the particles are in the same state, and in the configuration $\frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle)$ the system is in a superposition of two configurations. This points to a further feature of this case: the quantum-mechanical entanglement of particles. Particle A and B are each in a superposition of states $|0\rangle$ and $|1\rangle$, but when the state of A is determined for instance by means of a measurement, this instantaneously determines the state of B. Again, a combination of each particle's properties of being in possible states $|0\rangle$ or $|1\rangle$ does not result in the three configurations corresponding to bosons.

$$\text{Fermions: } \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle)$$

Finally, if the particles are fermions, there is only one possible configuration of the system. This is because the states $|0\rangle|0\rangle$ and $|1\rangle|1\rangle$ are excluded, as particles cannot be in the same

state. Just as before, there is no single combination of each particle's properties of being in possible states $|0\rangle$ or $|1\rangle$ that results in the configuration corresponding to fermions.

So far, I have shown that even introducing a modification in Gillett's view to allow realisers that are modal properties of entities, Gillett cannot account for the realisation of Gibbsian entropy. Moreover, I would like to offer one last argument against the possibility of Gillett's view accounting for Gibbsian entropy. Even if the previous arguments fail, and there was a way to combine lower-level modal properties of entities to obtain Gibbsian entropy, this would go against Gillett's first and most important commitment: providing an empirically informed theory of metaphysical dependence. Scientific accounts of the relation between the properties of particles and thermodynamic properties always refer to a middleman property as the realiser: the property of being an ensemble with probability density $\rho(x)$. Hence, to eliminate talk of this property would go against the motivation for Gillett's account.

6.3.2.2 Can Gillett's View Account for the Realisation of Boltzmannian Entropy?

In this section I argue that Gillett's many-one view is also unable to account for Boltzmannian entropy. Let us review the expression for Boltzmannian entropy:

$$S_B(x) = K_B \ln |\Gamma_{M(x)}|$$

In this case, entropy does not depend on the ensemble and is a property of each microstate, x , compatible with a given macrostate, $M(x)$. Each microstate compatible with a macrostate has the same entropy. As I discussed in section 6.3.1, the most plausible realisation base of Boltzmannian entropy is a single property of a microstate: the property of being in the volume of microstates corresponding to the macrostate $M(x)$, namely $|\Gamma_{M(x)}|$. Again, for Gillett's view to be able to account for this case there should be a many-one structure underlying the one-one relation I have introduced. What properties could particles have that combined could bring about the property of the microstate of being in the volume $|\Gamma_{M(x)}|$? Although on the face of it they might look very different, the volume $|\Gamma_{M(x)}|$ has a key element in common with the notion of ensemble: the modal element, which stems from the fact that the volume is obtained from calculating all the possible microstates compatible with a given macrostate. Let us again grant that particles may have modal properties. In this case, each particle of the system should have

the property of being in all possible states compatible with a given macrostate. This is very similar to the previous scenario, the only substantial difference is that instead of considering an ensemble we consider the volume of microstates, so the same objections raised in the previous section apply to this case. The conclusion, thus, is the same: given the relevance of indistinguishability, there is no way to obtain the volume of microstates only considering the properties of particles. For this reason, Gillett's view fails to account for the realisation of Boltzmann entropy.

6.3.3 Testing Wilson's Subset View

In this section I will establish whether the Subset View can account for the metaphysical dependence relations among the properties discussed in the case study. Jessica Wilson provides several cases where the Subset View works, what she refers to as cases of weak emergence. One fruitful way to find out whether the subset view can account for the metaphysical dependence relations between thermodynamic properties and their mechanical statistical realisers is to compare this case with a paradigmatic example discussed by Wilson. No case is more paradigmatic than mental states arising from c-fibre firing.¹² The subset view can explain very neatly how mental states can depend on c-fibre firing yet be causally autonomous. In Chapter 2, I provided a thorough discussion of how the subset view works in this case. For this reason, I will just briefly remind the reader of the main points of the argument, while omitting the points that are not relevant to the current discussion.

Take a mental feature M, say pain, that synchronically depends on a physical base feature P, say a particular firing in the nervous system. M has the power to cause another physical feature Q, say a yelp. Let us assume that M satisfies the Identity Condition on Causal Powers so that its power to cause Q is numerically identical to a power had by P. In this way, when M causes Q, P does cause it too. Wilson argues that M and P are distinct features because the causal profile of M is different from the causal profile of P. The argument for this is straightforward: P firing can be recorded in a neuron detector, while pain cannot. This entails that P has a power

¹² This example is problematic because it does not engage with the best scientific accounts of consciousness, so it does not provide an empirically informed conclusion. Even if this is not a story about actual relations between scientific features, the subset view can explain very neatly how mental states depend on c-fibre firing, yet be causally autonomous, and I use this as a toy model for the subset view.

not had by M, namely the power to affect a neuron detector. So we have two different properties, M and P, which are both causes for Q. As I explained in Chapter 2, this does not entail any sort of problematic overdetermination. Finally, Wilson provides two arguments, one using natural laws and the other appealing to difference-making considerations, to show that M is causally relevant vis-à-vis P. Again, I will not stop here to discuss this claim, which I did in Chapter 2, and also it is not relevant to the main discussion of this section. A point worth emphasising, that has been thoroughly discussed in Chapter 4, is the status of P as a middleman property. P is a property of an aggregate of lower-level entities, in this case plausibly a property of a group of neurons.

Going back to our case study, I will show now how this schema could apply to thermodynamic properties and their statistical mechanics realisers. Take the example I introduced above. We have a higher-level property, entropy, that arises from a lower-level, statistical mechanics property. So, this lower-level, statistical mechanics property is in Wilson's terms a middleman property, and is the realiser of entropy. A system with relatively low entropy has the power to cause a system that is put in contact with it to diffuse, and this power is numerically identical to a power had by the aggregate of particles. Moreover, the aggregate of particles also has a power not had by the higher-level property, namely the power of being able to trigger a particle detector. So the same general schema applies to this case. *Prima facie*, we could therefore conclude that entropy is a causally autonomous weakly emergent property, and that Wilson's view clearly accounts for it.

However, there is at least one way to problematise the above analysis: from my previous discussion of Gibbsian and Boltzmannian accounts of entropy, it is not clear that the realiser of entropy is what Wilson refers to as a "lightweight relational property". Under both the Gibbsian and Boltzmannian interpretation to obtain entropy from the lower-level components of the system we cannot only consider the actual microstate of the system, which would correspond to a lightweight relational property —e.g., the property of being an aggregate of particles with certain velocities and positions—, but we need to consider an array of possible microstates. Under the Gibbsian interpretation, to explain a certain thermodynamic system in terms of the particles that compose it we need to introduce the notion of ensemble. As I have explained before, an ensemble is a collection of systems of the same kind. Each copy of the system represents one possible microstate compatible with a given set of macroscopic constraints, and the collection contains all the possible microstates of the system compatible with such

constraints. I have also explained before that to obtain the value of any thermodynamic property of a system, except for temperature and entropy, we need to calculate the ensemble average. On the other hand, to obtain the value of entropy and temperature we need to use other mathematical expressions, but nonetheless these expressions use the notion of ensemble as well. The ensemble is a modal notion, so to obtain a certain thermodynamic property of a system, say entropy, we cannot just consider the actual system, but we need to consider an infinite number of possible microstates the system could be in.

Let us move on to the Boltzmannian approach. Although the Boltzmannian approach, as I have explained, does not need ensembles to obtain the thermodynamic properties of a system, it does introduce a magnitude, the volume of microstates corresponding to a given macrostate that, just as the notion of ensemble did in the Gibbsian approach, introduces a modal element to the explanation of entropy in terms of the entities composing the system and their properties. So everything that will be said about the realiser of Gibbsian entropy is valid for the realiser of Boltzmannian entropy, more so if, with Wallace, we consider the latter to be a special case of the former.

The fact that the realiser has a modal element has implications for Wilson's understanding of middleman properties. As I introduced them in Chapter 4, according to Wilson middleman properties are lightweight mereological resultants that are obtained from combining the components propertied and related in certain ways. However, this case study shows that we need a new definition of the middleman property, as the middleman property is not a mere combination of components. To motivate this new definition, in Figure 3 I represent the relations and properties of this case study in a version of the schema presented in Chapter 4. There are two kinds of realisation relations possible: 1) a realisation_{M-O} between the joint role-filling of the particles propertied and related in a certain way (level n) and the system's property of having entropy (level $n+1$); and 2) a realisation_{O-O} between the property of being an ensemble with a probability density $\rho(x)$ at level k (the middleman) and the system's property of having entropy (level $k+1$). In section 6.3.2.1 I have argued that there are reasons to believe that Gillett's realisation_{M-O} cannot account for the realisation of entropy, as I argued we cannot by-pass the ensemble. Entropy cannot be directly obtained from the joint role-filling of particles propertied and related in a certain way as it is a property defined in terms of possibilities at the level of the system: the property of being an ensemble with a probability density $\rho(x)$. This is a very different kind of middleman property than the one Wilson had in

mind in 2015, as it is not merely a relational property arising from “ontologically ‘lightweight’ compositional principles, including additive causal combination, and certain Boolean or mereological operations” (Wilson, 2015, p. 350). It also has a modal element at the level of the system, the ensemble. Still, in Figure 3 I have placed the ensemble, or middleman, at level k, as it does not involve any sort of emergence and it is derivable from the properties of the system. The only difference with middleman properties as understood in Wilson 2015 is that one must consider possible states the system could be in, rather than only actual states. According to this modified version of the subset view, a middleman property, or a realiser_{O-O}, is a Holistically Defined Modal property. Call this the HDM view of realisation. The realiser is holistically defined because it is defined in terms of properties and states of the system. In some cases, such as the case in this chapter, this is an important amendment to Wilson’s view: the realiser property cannot be defined in terms of properties of the system’s constituents. There are other cases where the realiser property may be further defined in terms of compositional principles and Boolean operations on the constituents, as Wilson had in mind when she introduced middleman properties as realisers. Being “holistically defined” allows for both kinds of cases. On the other hand, realisers are modal properties because we need to consider possible states the system could be in. Of course, not all cases of realisation involve such modal elements. There are cases where the realisation of a property can be explained purely through ontologically lightweight middleman properties, as envisioned by Wilson, and there are also cases of holistically defined, but not modally defined, middleman properties. An example of this will be discussed in the following chapter, Chapter 7. The HDM view is formulated to accommodate these cases –i.e., cases without modal elements– by recognising that within the possible properties of a system we find its actual properties.

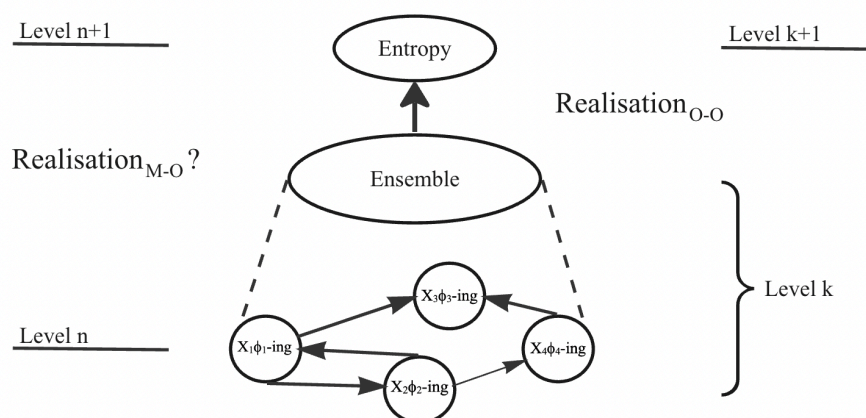


Figure 3. Schematic representation of the metaphysical dependence relations operating in the realisation of entropy.

That said, although I have provided an argument —based on the indistinguishability of particles— to show that a modified version of the Wilson’s subset view is better suited to account for this case study, I would like to offer now a weaker or less controversial interpretation of the case study that still can be used to respond to Gillett’s Flat Objection. Let us assume that entropy can be realised_{M-O} by the joint role-filling of the particles propertied and related in a certain way. In this scenario, realisation_{O-O} is not the only realisation relation possible. However, the fact that it is a possible and scientifically reputable realisation base is sufficient to defend the subset strategy. Gillett’s Flat Objection relies on the middleman property being a “monster” property and not justified empirically, and also eliminable due to parsimony considerations. This case study shows that the property of being an ensemble with a probability density $\rho(x)$ is a scientific non-monstrous property that can be used as a realisation_{O-O} base of entropy, and that moreover it is the property scientists take to be key to explain how thermodynamic properties arise from collections of particles. This is a powerful rebuttal of Gillett’s arguments against the subset view. However, we are left with the worry of parsimony. It is true that if realisation_{M-O} is operative as well, positing an extra property might raise parsimony concerns. But I believe that these concerns are much less pressing given the scientific nature of the middleman property.

6.4 Conclusion

In this chapter I have explained how statistical mechanics derives the entropy of a system from the constituents that compose it. I have explained two different approaches to the derivation of entropy, the Boltzmannian approach and the Gibbsian approach, and I have concluded that they both share a key modal element. Given that some authors consider the Boltzmannian approach to be a special case of the Gibbsian approach, I have focused on analysing the Gibbsian approach. There is sufficient reason to believe that a similar analysis of the Boltzmannian approach would yield the same results. The results are the following: 1) realisation_{O-O} is better suited to account for the realisation of entropy, 2) the realiser base in this case is a Holistically Defined Modal property, which has led me to introduce a modification of the subset view, the HDM view, to accommodate this kind of realisers. The HDM can account for the kind of realisers that Wilson envisioned – i.e., middleman properties– but also for a new kind of realisers: holistically defined modal realisers. That is, realisers that cannot be obtained directly from the properties of the constituents alone, and that require considering the possible states the system could be in.

In this chapter I have not shown that entropy is weakly emergent, as the aim of this thesis is not to demonstrate that there are actual cases of weak emergence, but to elucidate how NRP should be spelled out. In this case, it is left open whether the property of being an ensemble with a probability density $\rho(x)$ has more powers than the entropy it realises. To conclude, this chapter has shown that the subset strategy, albeit in a modified form, proves to be a suitable framework for accounting for scientific case studies.

CHAPTER 7. CASE STUDY II: THE GLOBAL NEURONAL WORKSPACE

7.0 Introduction

As I showed in Chapter 4, Wilson’s and Gillett’s accounts of the metaphysical dependence relation — realisation_{O-O} versus realisation_{M-O} — between entities at different levels are difficult to assess against each other. The main reason for this is that, although they both provide a so-called realisation relation between entities at different levels, the inter-level relations they propose aim to bridge distinct kinds of levels. On the one hand, Wilson’s view addresses the dependence relation between entities and properties at different realisation levels — k and $k+1$ — and concludes that the relation is realisation_{O-O} . On the other hand, Gillett’s view addresses the dependence relation between properties at distinct levels of mechanisms — n and $n+1$ — and concludes that the relation is realisation_{M-O} . As I argued in Chapter 4, Gillett’s Flat Objection against the Subset View does not succeed because he has failed to explain how the objection would be relevant to the case of levels of realisation. The objection only works if levels of mechanisms were the only levels relevant to any discussion about realisation.

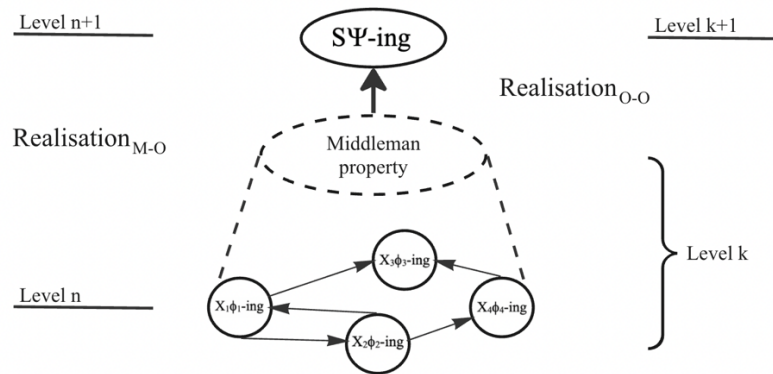


Figure 1. Realisation_{M-O} bridges levels of mechanism (n and $n+1$) as an interlevel relation of organisation, and realisation_{O-O} bridges levels of realisation (k and $k+1$).

The question I aim to answer in these case studies is whether realisation_{O-O} could be eliminated, as Gillett suggests, so that in every case of realisation the constituents (level n) could directly realise_{M-O} all the higher-level properties of a composite. In Chapter 6 I showed that in cases of thermodynamic properties like entropy there is empirical justification for the existence of a realiser_{O-O} at level k . That is a substantial finding that, as shown, can tip the balance of the

debate in favour of a modified version of the subset view, the HDM (Holistically Defined Modal) view. The case study in this chapter provides further justification for the need of a realiser property at level k . Moreover, this case study is significant because of the relevance of the phenomena under discussion: consciousness. As I explained in Chapter 1, the debate over emergence and nonreductive physicalism had in its origin and still has as a driving force the quest for understanding the nature of consciousness. In line with my aim to provide an empirically informed analysis of the properties studied, I have decided to centre this case study on the Global Neuronal Workspace (GNW). This theory is one of the best scientific theories currently available to explain the relation between consciousness and the brain, although it is not the only one. As will be clear in section 7.3, though, my conclusions apply to any scientific theory of consciousness that takes communication between neurons to be part central to consciousness. Before introducing the GNW case study, I present another case study: the Circadian Rhythms case study, following Bechtel (2017). His insights on understanding a circadian clock as a mechanism over and above the entities composing it will be crucial in my discussion of the GNW case study. However, although I draw on his analysis, my conclusions do not align with his.

7.1 Bechtel's (2017) View about Mechanisms

In this section I discuss a mechanist account that does not use middleman properties: Bechtel's (2017) account of mechanisms. Bechtel (2017) tackles the important issue of what the metaphysical relation is between the parts of the mechanism, that is, the $X_i\phi_i$ -ing, and the higher-level phenomenon, $S\psi$ -ing. And he does that bypassing the middleman property, i.e. the organised collection of $X_i\phi_i$ -ing. Importantly, Bechtel presents his account as a better framework to explain mechanisms than Craver's: the most pressing problem Bechtel detects in Craver's view is that it leaves higher-level mechanisms vulnerable to being eliminated or reduced to the organised collection of $X_i\phi_i$ -ing. For some, this might not be a problem, but if one is concerned with the structure of the mechanistic world, it seems an underwhelming response to conclude that only the bottom level of organised parts exists and is causally efficacious. The question he tries to answer is then why higher-level phenomena (or mechanisms as a whole) should be treated as higher-level rather than "just a collection of lower-level entities, with all the causality operative between these lower-level entities". I want to emphasise that what really interests me about Bechtel's proposal is the framework he provides to show that a mechanism cannot be reduced to its components.

Bechtel identifies a general problem in the mechanist literature. The problem is that the well-known mechanist schema, reproduced below in Figure 2, is actually not properly justified. In particular, he claims that mechanists provide no arguments to show that the mechanism cannot be reduced to its components, so the ontological status of the mechanism as a whole, Ψ -ing, is at stake. Craver and Bechtel (2007) tried to address this issue, by showing how mechanisms are involved in top-down causation. Certainly, one compelling way to argue for the irreducibility of Ψ -ing is to show that it has irreducible causal efficacy. Craver and Bechtel (2007) provided an account to explain top-down causation without top-down causes, introducing a new notion: mechanistically mediated effects. This notion is used to explain how mechanisms can produce top-down effects without them being top-down causes: mechanisms cause effects at a lower level because they are fully constituted by components at that level, and any power the mechanism has can be explained in terms of the powers of its constituents. So mechanistically mediated effects consist in two steps: the mechanism does not directly cause the top-down effect, but it causes it indirectly in virtue of being constituted by components that are at that same level and have the powers to cause that effect. In this way, causation can be considered to be only an intra-level relation, yet constitution relations allow us to explain how a mechanism can have top-down effects. However, although he was on board in 2007, Bechtel (2017) is not convinced by the approach. Certainly, saying that a mechanism can cause top-down mechanistically mediated effects is far from saying that the mechanism, qua mechanism, is causally efficacious: the mechanism is a cause only in virtue of the causal powers its components have, so anything the mechanism causes at the level of components can be explained by a power the components have. This is a problem, according to Bechtel, because for mechanisms to be irreducible, they need some sort of causal (or determinative) efficacy qua mechanisms. In Bechtel's words, we must answer the question: "why should one treat the whole mechanism as a higher-level entity rather than just a collection of lower-level entities, with all the causality operative between these lower-level entities?" (Bechtel, 2017, p. 254).

In his quest to show that the mechanism, Ψ -ing, is irreducible to its components, namely the collection of $X_i\phi_i$ -ing, Bechtel (2017) introduces a new way of representing the relation between components and the higher-level mechanisms they compose, as shown in Figure 2 below.

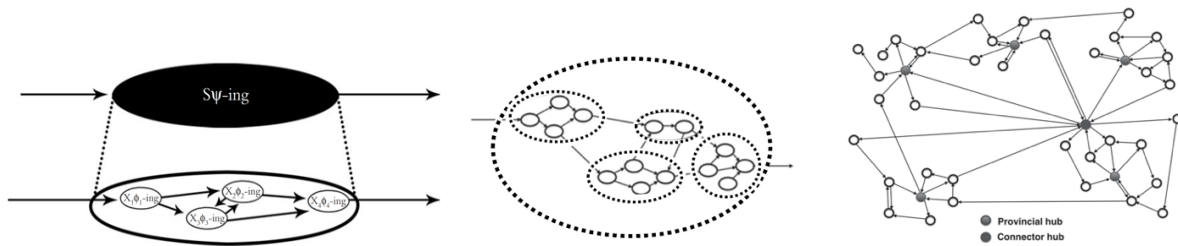


Figure 2. Bechtel flattens the figure on the left into the figure in the middle, where higher-level mechanisms (and mechanisms within mechanisms) are represented by the dotted ovals, and the bottom-level entities are the ones represented by solid ovals. If we only consider the solid ovals and their interactions, we obtain a graph-theoretic representation of a network. On the right side, there is an example of a network. Figures extracted from Bechtel (2017, p. 262 & p. 262) and Craver (2007, p.7).

On the left side of Figure 2, there is the traditional way of presenting mechanisms and their components. In the middle figure, mechanisms are represented by dotted ovals because their existence is at stake, while bottom-level entities' existence is guaranteed because they cannot be further reduced. Bechtel's first move is to provide a new model for representing mechanisms, as shown in the figure on the right. Bechtel uses graph theory to represent lower-level components as nodes interacting in a network. Given that graphs are *prima facie* one-level systems, or flat systems, it might seem striking that Bechtel uses them to argue for the existence of higher-level entities. However, using graph-theoretic analysis Bechtel can recover, and importantly justify, the hierarchy of levels present in the usual picture of mechanisms. There are three features of graph theory that I need to introduce before showing how we can introduce levels in a graph representation (Bechtel, 2017, p.263, emphasis his own):

1. *Mean shortest path length* is a measure of the average number of edges that must be traversed on the shortest path between any two nodes.
2. The *clustering coefficient* is a measure of how connected to each other the nodes linked to a given node are.
3. *Degree distribution* is a measure of how the number of connections from different nodes is varied.

If we go back to Figure 2, on the right side is represented a network that has all the features required to introduce levels. It exhibits what is known as small-world organisation. Networks with small-world organisation have the following features: 1) they have a small mean shortest path length; 2) they are highly clustered: collections of nodes are highly connected with each

other in what are known as modules; and 3) the degree-distribution is scale-free, which means that the number of connections from different nodes divides nodes in two groups: provincial hubs and connector hubs. Provincial hubs serve to integrate activity within modules, while connector hubs serve to integrate activity between modules. I can now explain how Bechtel uses graph theory to understand mechanisms. His explanation has two main points. The first point he makes, as seen in the quote below, is that the high number of connections between nodes within a module justifies positing a mechanism corresponding to that module.

The extensive interactions between the nodes constituting a module enable them to work together as a unit. This does not mean that the connections to components outside the module are not important; inputs from them may figure critically in regulating some behaviours of the module and enabling the output of the mechanism to affect other processes. But the enhanced connectivity enables greater coordination within the module, allowing the components to work together, and justifies treating them as constituting a mechanism when they produce a phenomenon we seek to explain. (Bechtel, 2017, 265)

The second point is that the graph structure generates a certain type of dynamic behaviour: components work together in a coordinated way. As I explain now, Bechtel aims to show that these two points are sufficient to demonstrate that mechanisms cannot be reduced to the collection of nodes and their interactions. For mechanisms to be irreducible, producing mechanistically mediated effects is not enough, they need a stronger causal (or determinative) efficacy qua mechanisms. The gist of his argument is that due to the dynamics between nodes, states of a whole mechanism constrain the behaviours of its parts. In other words, depending on the state of the mechanism, its constituents will respond differently to external inputs. In the following pages I explain in detail how constraints of this kind are obtained, and I will do that using Bechtel's case study of Circadian Rhythms. To put it very succinctly, he argues that top-down constraints are the result of interactions between components (nodes) within a module involving multiple feedback loops. In many cases, dynamic complex behaviour due to these feedback loops is characterised as endogenous oscillations at the level of the mechanism. Those oscillations are what constrain the behaviour of the parts. According to Bechtel, this framework is necessary to account for mechanisms constraining the behaviour of their parts, and is common in biological systems.

7.1.1 The Circadian Rhythms Case Study

To show in more detail how endogenous oscillations are obtained, and what role they play in constraining behaviour of the parts of a system, Bechtel uses a case study. He focuses on a simplified version of the phenomenon of circadian rhythmicity in mammals. Circadian rhythms are oscillations generated endogenously that govern certain behaviours and physiological functions and that have been entrained (i.e. gradually caused to fall into synchrony) by day/night cycles in the environment. Now I offer a simplified explanation of circadian rhythms once the mechanism has been entrained by day/night cycles, following von Schantz & Archer (2003).

The 24-hour circadian cycle: In the early morning, *Cry* and *Per* genes are activated by the CLOCK and BMAL1 proteins, producing mRNA transcripts. During the day, ribosomes bind to these mRNA transcripts and synthesise proteins PER and CRY. In the late afternoon, proteins PER and CRY accumulate in the cytoplasm, where casein kinase CK I (δ and ϵ) are present. When sufficient concentration of CRY and PER is found in the cytoplasm, that is in the evening, CRY, PER, and CK I molecules bind to each other, forming CRY:PER:CK I complexes, and they enter the nucleus at night. Once inside the nucleus, CRY:PER:CK I complexes bind to the CLOCK and BMAL1 proteins, inhibiting their activating effect on *Cry* and *Per* genes. As a result, *Cry* and *Per* genes stop producing mRNA transcripts, which in turn interrupts production of CRY and PER proteins. CRY and PER proteins gradually break down and disappear from the nucleus as the morning starts. This allows for CLOCK and BMAL1 proteins to recover their activator effect on *Per* and *Cry* genes, which resume activity. This is the full circadian cycle. The production of CRY and PER starts again, so the cycle starts anew.

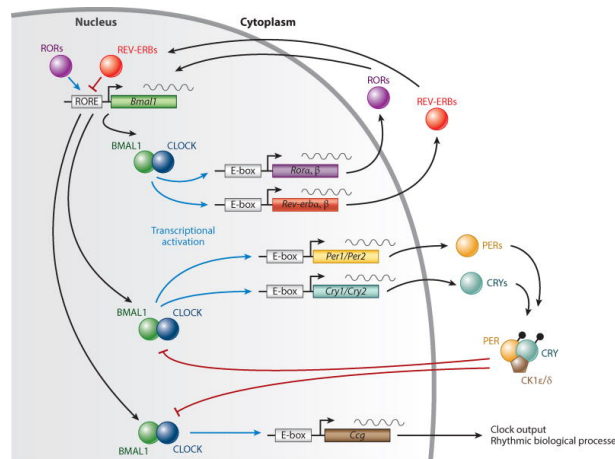


Figure 3. Representation of the 24-hour circadian cycle. Extracted from Lazzereni Ospri et al. (2017).

The mechanism for circadian rhythms, also known as the clock mechanism, arises from a negative feedback loop: the concentrations of PER and CRY inhibit their own transcription once a certain threshold is reached. This behaviour is stable, more specifically, there is a sustained waning and waxing of the concentrations of PER and CRY, characterised as an endogenous oscillation (the clock mechanism) that has 24-hour periodicity. That endogenous oscillation is at the level of the whole cell, and its main parts are genes *Per* and *Cry* and the proteins synthesized from them, PER and CRY, as well as activators BMAL1 and CLOCK. How is this used to justify the irreducibility of the mechanism?

Whether *Per* and *Cry* are transcribed and translated into the proteins PER and CRY depends on the phase of oscillation the host cell is in. *The oscillatory phase, a state of the whole cell, is determining the behaviour of its parts, seemingly in accord with accounts of top-down causation.* (Bechtel, 2017, p. 257)

In this quote Bechtel claims that the state of the whole mechanism determines the behaviour of its parts, and that this is what is required for top-down causation, which in turn is crucial to show that mechanisms are irreducible to their parts. In this case, the oscillatory phase is the state of the whole mechanism that determines the behaviour of the parts, which in this case are the genes linked to the production of PER and CRY. However, Bechtel accepts that the mechanism just is a collection of parts: “The phase of the oscillation in the cell is not distinct from the concentrations of PER and CRY (...) the state of the whole mechanism might be described in terms of the states of some components of the mechanism (e.g., PER

concentrations in the nucleus are high)” (Bechtel, 2017, p. 267). These are surprising claims, as it is then problematic to see how oscillations are irreducible and at the level of the cell. Let me emphasise the problem arising from these claims, which will become even clearer by the end of this section: Bechtel sometimes considers the oscillatory phase to be a state of the cell (that is, a higher-level state), and at times, given that it can be described in terms of the states of the components, he considers the oscillatory phase to be a part of the mechanism (that is, a lower-level state). Moreover, there is still no proof that the oscillatory phase plays any causal role. I now delve deeper into Bechtel’s view to understand his argument against reducibility, and I will show that the problem just mentioned is not solved. Not only that, but in his argument to show that there is top-down causation, Bechtel exploits the ambiguity regarding the level at which the oscillatory phase is found.

Let me explain in detail Bechtel’s view against reducibility. The presence of feedback loops is the first step in the two-step argument Bechtel uses to show that mechanisms cannot be reduced to their parts. Highly interconnected modules contain feedback loops, where the products of an operation performed by a part themselves affect the operation of the part that brought them into being. As I just mentioned, a feedback loop is present in the explanation of circadian rhythms. The genes *Per* and *Cry* cause the synthesis of proteins PER and CRY, but in turn the concentrations of such proteins inhibit or activate the genes. This kind of feedback loop, which is commonplace in biological systems, typically generates endogenous oscillations. So how does all this explain that the clock mechanism is at a higher level than the concentration of proteins from which it results? The second step of the argument is that due to feedback loops, the same external input has different effects on the parts of the mechanism depending on the state of the whole mechanism. In the case study, this means that an external input has different effects on the parts depending on the phase of the oscillation, which is considered to be at a higher level. This is how Bechtel explains this phenomenon:

(...) A light signal from the retina (...) functions to increase PER transcription. Now researchers could understand why light exposure had different effects on the phase depending on time of day. If PER transcription is already at its maximum, as it would be during anticipated daytime, light input could have no further effect. At dusk and early night, as PER concentration is diminishing due to the feedback mechanism, a light signal can counter the increasing inhibition and keep transcription going longer. This has the effect of delaying the phase of the oscillator. On the other hand, later in the night as dawn is approaching, PER concentration is again

beginning to increase due to the endogenous oscillation. A light signal at this phase will speed up transcription, resulting in reaching daytime levels in PER concentration earlier and advancing the phase of the oscillator. (Bechtel, 2017)

The same input, that is a light signal, can cause three different effects to the phase of the oscillation depending on the phase of the oscillation the mechanism is in: (1) it can have no effect, if it is in a maximum; (2) if it is decreasing, it can delay the phase of the oscillator by countering inhibition of PER and CRY transcription, or (3) if it is increasing, it can advance the phase by speeding up transcription. From this, Bechtel needs to argue that the clock mechanism determines, or constrains, the behaviour of its parts, which was, remember, his main aim with this case study. As details are important, I quote his argument supporting this claim below:

The circadian example makes clear the functional importance of the type of integration of nodes found in modules in network representations of biological systems. The operations of individual components (e.g., transcribing genes) depend on the state of other parts of the mechanism (e.g., the concentrations of the PER protein in the nucleus). As a result of the interconnectivity of the parts, especially the feedback loops, the module identified as the mechanism functions as a unit, *with the operations of the individual parts of the mechanism determined by other parts of the mechanism*. If there were no feedback loops, the parts that receive the input would not be sensitive to conditions elsewhere and would always respond in the same manner to the input. There would also be little reason to identify the parts as constituting a module. The feedback loops are responsible for *the state of the module modulating the responsiveness of its parts to external input*. Even though the input to the mechanism may affect only one or a few components, the mechanism as a whole is the relevant unit due to the interconnections that run through the mechanism. (Bechtel, 2017, p. 268, my italics)

With feedback, modules are capable of endogenous activity that results in their being in different states at different times. When this happens, some of their components are in different states. Depending on the state, the component will produce different responses to inputs. This is the sort of behaviour that has characteristically motivated talk of top-down causation. (Bechtel, 2017, p. 273)

In these two quotes the crucial aim of Bechtel's case study —namely, to show how the clock mechanism constrains the behaviour of its parts thus providing an argument for the

irreducibility of the mechanism— becomes problematic. First, he argues that when an input affects a part of the mechanism, the effect it has is determined by other parts of the mechanism due to feedback loops. Depending on the state of those other parts of the mechanism, the effect the input will have on the parts affected is different. In this case, he is treating the oscillatory phase as a part of the mechanism. In this case Bechtel is explaining why the effect of a given input can be different at different times referring only to the parts of the mechanism and how they interact. But then he goes on to argue that the state of the module is what “modulates” the behaviour of the parts. In this case, he is treating the oscillatory phase as a state of the whole mechanism, and the effect it has in the behaviour of the parts as a top-down constraint. This top-down determination is what justifies that the clock is irreducible to its parts. The explanation is, at least, confusing. However, I believe Bechtel’s proposal has an even greater problem, which I go on to explain now. After that, I will show that Bechtel could have made a stronger argument for the causal efficacy of higher-level mechanisms, although quite possibly circadian rhythms are not complex enough to engender this kind of higher-level autonomy.

First, I want to scrutinise the claim that the same input, namely a light pulse, causes different effects on the affected parts of the module depending on the state of other parts of the module. I focus on PER from now on, but the same analysis applies to CRY. It is not totally known how light affects the proteins involved in circadian rhythms, but it is known that melanopsin molecules in photoreceptive cells are affected by light, and they trigger a cascade of reactions that ends up with an increased activity of PER transcription: “photons affect the melanopsin molecules in the intrinsically photoreceptive retinal ganglion cells. As a result, when neurotransmitters are released, they initiate a signalling cascade within cells in the suprachiasmatic nucleus, resulting in increased *Per* transcription.” (Bechtel, 2017, p. 262) The effect on PER transcription, then, is the same independently of the phase of the clock. In two out of the three cases, the light pulse increases PER transcription. Although an increase in PER transcription affects the clock in different ways depending on its state — in one case it delays the phase and in the other it advances it— it has the same effect on the parts of the mechanism. The third case, where during daytime PER transcription is at a maximum, the light pulse does not increase PER transcription. But we need not appeal to the clock to explain this, as it can be explained merely by the fact that *Per* genes are already transcribing at full capacity. So in what sense does the clock constrain or affect the behaviour of the parts? I believe Bechtel’s answer to the question “why should we consider mechanisms higher level entities with their own causal efficacy?”, which is to appeal to modules exercising constraints on their parts, has not been

satisfactorily presented. However, I think there is promise in this approach, and in what follows I present a new view inspired by Bechtel's case study that shows how some mechanisms are irreducible to their parts.

The view is the following. In some cases, the oscillatory behaviour characterising a mechanism, although closely related to the behaviour of a part or some parts, cannot be attributed to any of the parts. In those cases, none of the singular parts shows oscillatory behaviour (or at least not at the same rate as the mechanism) at any given time, so the oscillation can only be attributed to the dynamic evolution of the entire system. Thus, in these cases there will be properties characteristic of oscillations that are not properties of any of the lower-level parts. For instance, oscillatory phase. The phase of an oscillation starts at 0 and each cycle finishes at a multiple of 2π —e.g., the third cycle finishes at 6π , however it has the same properties the system had at 2π , because the system is cyclical. Figure 4 below provides an illustration of the oscillatory phase: the red dot and green dot represent the same concentration of PER, but the oscillatory phase is different. The red dot has a phase of $3\pi/2$ and the green dot has a phase of $5\pi/2$, which is equivalent to a phase of $\pi/2$.

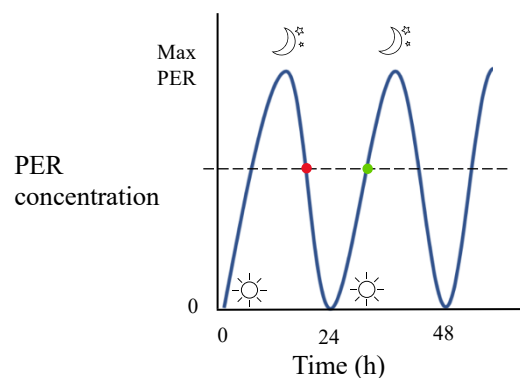


Figure 4. Sketch showing the evolution of PER concentration following the day/night cycle, where 0 hours corresponds to dawn. The sketch highlights two points where, although the concentration of PER is the same, the phase of the oscillation is different.

In cases where the oscillatory behaviour cannot be attributed to any of the parts, this makes room for a property of the overall system, the oscillatory phase, which is not a property of its components. A mechanism characterised by oscillatory behaviour can be considered a genuine entity in addition to the parts of the mechanism only if the two conditions below are true.

Condition 1: The oscillation may be found at the level of the mechanism only.

Condition 2: The properties of the oscillation determine the behaviour of the parts of the mechanism.

The second condition is needed to guarantee that the oscillation has causal (or determinative) efficacy and is not epiphenomenal, as it requires that the oscillation is necessary to account for the behaviour of the mechanism and its parts. Let me show why the circadian clock does not satisfy the two conditions. The circadian clock does not meet condition 1 because the oscillation that characterises the clock mechanism is an oscillation of the concentration of PER proteins, which are parts of the system. The circadian clock does not meet condition 2 because we can explain the behaviour of the clock mechanism—for instance, its interaction with other mechanisms in an organism—, only referring to the concentration of proteins, mRNA, and other molecules. It is true that given an external stimulus, such as a light pulse, the clock is affected differently depending on the phase it is in (it is either delayed or advanced), but as I have shown the stimulus has the same effects on the parts of the mechanism (PER transcription is increased). The oscillatory phase does not play a role in how the parts behave, as Bechtel claims it does, nor does it play an indispensable role in how the mechanism behaves. We can explain every single cause and effect in terms of the parts of the system. Therefore, Bechtel's case study does not meet either of the two conditions and he has not made a strong enough case to claim that the mechanism is irreducible to its parts.

Following my two conditions, in order to conclude that the clock mechanism is irreducible to the parts of the mechanism we should show: 1) that the phase of the oscillation is not equal to any property at the level of the parts, and 2) that the parts of the mechanism behave differently given the same input because of the phase of the oscillation. However, I do not think that is achievable in this case study. In the abstract, what we need is a mechanism that, given an input, behaves differently depending on the oscillatory phase it has, and that this phase cannot be attributed to any of the parts of the mechanism. Whether the conditions are met for any mechanism is an empirical question that needs to be answered on a case-by-case basis and following the best scientific theories about the mechanism considered. This is precisely what I have set out to do in the following section.

7.2 The Global Neuronal Workspace Case Study

In this section I introduce one of the most influential scientific theories about how consciousness arises in the brain: the Global Neuronal Workspace Theory (from now on GNWT). I focus on the central features of the view, in particular those that are relevant to my discussion of mechanisms and the metaphysical relation between entities and properties at different levels. In this case study the properties and entities in play at the lower level are spiking neurons forming networks characterised by oscillatory behaviour, and at the higher level the realised properties are mental properties. After a thorough analysis of this case study, I conclude that realisation_{O-O} is required to account for the realisation of mental properties, and that realisation_{M-O} is unable to account for it. The key move of my argument is to show that the oscillatory phase of the network of neurons meets the two conditions I introduced at the end of section 7.1.1. I then argue that given that the oscillatory phase of the network of neurons is irreducible, then according to the GNWT mental properties are realised $_{O-O}$ by a property of the network.

I start by explaining what consciousness is according to the GNWT, and then I will move on to explain the neuroanatomical basis of the view. The GNWT proposes that there are two principal areas in the brain: specialised processors and the Global Neuronal Workspace. Specialised processors are encapsulated and have evolved to process only certain kinds of information. For instance, motion processors in area MT, or the system of mirror neurons in area F5. Those specialised processors only have local and medium-range connections, so on their own they have no way to transmit their information to most areas of the brain. This is where the Global Neuronal Workspace steps in. Proponents of the GNWT suggest that the global and flexible connectivity between processors is achieved by means of a set of distributed long-range neurons, which form the Global Neuronal Workspace. Those neurons are characterised by their ability to make information available to a vast array of brain systems, which without this connection would not be able to transmit information to each other. Consciousness arises in the brain when the information being processed in a localised processor is broadly broadcasted to other cognitive processors. Let us take perceptual content as an example. Perceptual content is usually obtained by fast sensory processing in a localised area, such as V1. However, according to the GNWT this content is not yet conscious. It becomes conscious when it reaches the threshold needed to ignite the GNW. Once the GNW ignites, which is an event that triggers a self-sustained or reverberating state, the GNW broadcasts the

content globally. In this way, it becomes available to other cognitive processors, such as verbal report or motor behaviour. In the words of Dehaene & Changeux, two main proponents of the GNWT:

Conscious access would correspond to the crossing of a threshold in evidence accumulation within a distributed global workspace, a set of recurrently connected neurons with long axons that is able to integrate and broadcast back evidence from multiple brain processors. During nonconscious processing, evidence would be accumulated locally within specialized subcircuits, but would fail to reach the threshold needed for global ignition and, therefore, conscious reportability. (Dehaene 2008, p. 89)

Conscious stimuli would be distinguished by their lack of “encapsulation” in specialized processes and their flexible circulation to various processes of verbal report, evaluation, memory, planning, and intentional action, many seconds after their disappearance. [We] postulate that this global availability of information is what we subjectively experience as a conscious state. (Dehaene & Changeux 2011, p. 210)

The Global Neuronal Workspace (GNW) is the collection of neurons that, at a given time, makes broadcasting possible. There is compelling evidence that supports the existence of those neurons, and there is also extensive research into how they achieve the task of making information broadly available. However, due to the scope and aim of this chapter, I do not need to discuss that evidence (see Mashour et al. (2020) for an extensive review of the empirical evidence supporting the view). Rather, my focus is on the implications this view has for my discussion of the many-one or one-one nature of the metaphysical dependence relation between mental properties and properties of neurons. That this view is regarded by the scientific community as one of the most explanatory powerful and empirically supported views about consciousness is sufficient to justify taking it as a key case study to assess metaphysical dependence relations in this domain.

What are the global workspace neurons and where can we find them? There have been many new developments in the theory since it was first proposed by Baars (1988), who suggested that the reticular-thalamic activating system had the structure necessary to accommodate all the features of the global workspace neurons. The most up-to-date account suggests that the global workspace is a network of long-range neurons that link the thalamus with the cortex, as

shown at the top of Figure 5. For my present purposes, I do not need to give any more details about the actual anatomy of the global neuronal workspace, as I will focus on simulations of those networks (see Dehaene & Changeux (2011) for a review of the research on the anatomical substrate of the GNW). Due to ethical and practical limitations, most of the research about the GNW in humans has been made through simulations and by extrapolating findings from experimental studies with monkeys. Although simulations have limitations—for instance, they are idealisations of the real systems they are simulating—in this area of research they provide insights that would otherwise be impossible to achieve. Importantly, the simulations are consistent with available neurological findings, and are guided by them.

What is most interesting about these simulations is how they are able to capture the relations between the neurons in the network, and then show how those relations generate new network states that give rise to conscious states. I want to focus first on the simulations presented by Dehaene & Changeux (2005) and Dehaene, Sergeant & Changeux (2003).

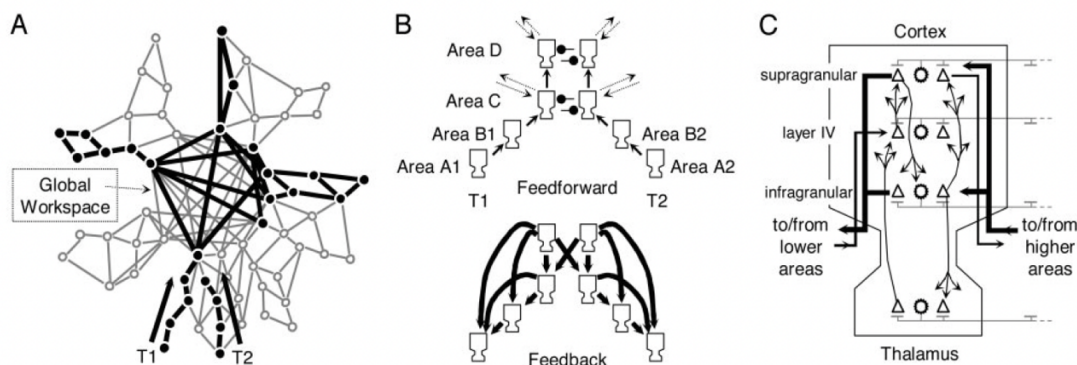


Figure 5. A and B are a schematic representation of the GNW's architecture consisting of specialised processors and a central network of high-level areas. C represents a simulated single thalamo-cortical column, where triangles are excitatory neurons and circles inhibitory ones. Figure extracted from Dehaene, Sergeant & Changeux (2003).

The simulations presented in both papers follow the same principles, techniques, and use the same model of neural network. However, they focus on distinct aspects of the GNW. The aim of the simulations is to show that the GNWT is the theory that best explains facts about models that are good enough idealisations of the known workings of the brain. The model used by Dehaene, Sergeant, and Changeux is found in Figure 5 above. The basic elements in the model

are spiking neurons. Spiking neurons are artificial neurons that are considered to closely mimic real neurons. The papers here considered use the integrate-and-fire spiking neuron model. In this model, neurons are represented by their membrane potential, $V(t)$, and spiking only occurs when there is an input current $I(t)$ above a certain threshold —see Figure 6. At that point, the neuron fires, that is, it generates a spike —also known as action potential— that in turn affects the membrane potential of neighbour neurons. In this way, a signal is transmitted through the network of spiking neurons.

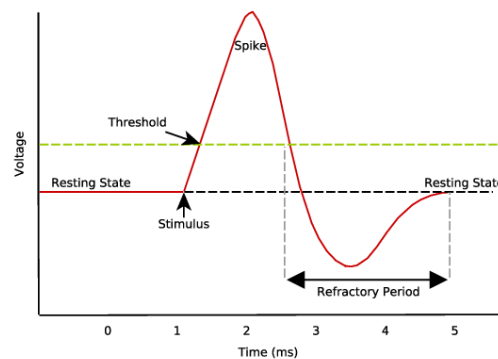


Figure 6. Model of a spiking neuron. Extracted from “Exploiting Refractory Period for Functional Multiplexing and Short-Term Memory in Spiking Neural Networks”, Zhenduo Zhai & Ismail Akturk (2020).

Groups of spiking neurons form thalamocortical columns (Figure 5, C). Each column contains 120 neurons — 80 excitatory neurons and 40 inhibitory ones— organised in three layers — infragranular layer, IV layer, and supragranular layer— and a thalamic sector. The model then takes four thalamocortical columns to represent four different brain areas and organises them hierarchically. According to the authors, these areas represent actual areas of brain anatomy: “areas A and B might correspond to a series of fast bottom-up cortical processors such as areas V1 or frontal eye field, whereas areas C and D would implement higher association areas of temporal, parietal, frontal, and cingulate cortex”. There are feedforward connections from each area to the next in the hierarchy: bottom-up projections are sent from the supragranular layer and are received in the IV layer; columns send feedback projections to all other columns lower in the hierarchy. Areas C and D, which in this simplified model represent assemblies of workspace neurons, are linked to areas A and B, as well as to other areas not represented here, by long-range excitatory axons that send top-down projections.

The simulations will help me answer a question I have not yet addressed in detail: How do GNW neurons enter the global ignition —or global broadcasting— state? And what characterises

this global state? To answer these questions, I need to introduce first a well-known phenomenon in the mammalian brain: neural oscillations. Oscillatory activity was first found in the scalp when the first electroencephalogram was recorded by Hans Berger. Nowadays, oscillatory electrical activity has also been measured much deeper in the brain, at the level of relatively small groups of neurons, as I will explain later in more detail. There are several roles attributed to oscillatory activity and there is yet no agreement within the scientific community about some of them, but the one that I will focus on, which is the role that GNW theorists attribute them, is to be the measure of consciousness.¹³ The oscillations of electrical activity found in the human brain have four main ranges of frequency: alpha, beta, gamma, and delta.

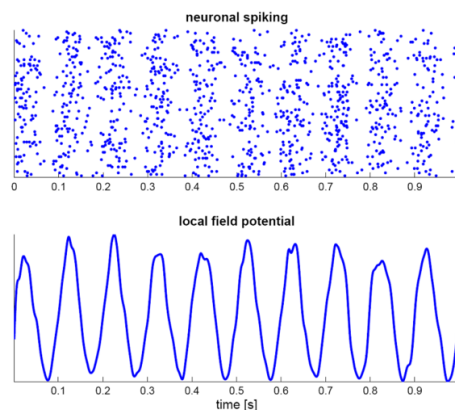


Figure 7. Representation of oscillatory activity of the local field potential. Image extracted from Yates (2016, Fig. 2).

As I mentioned, oscillatory activity was first found at the brain level, which can be measured by electroencephalogram (EEG) recording of brain activity at the level of the scalp. Oscillatory activity has more recently also been found at the level of ensembles of neurons—it is already present in ensembles of 120 neurons—. A measure used to characterise this activity is the local field potential (LFP). The LFP is the voltage potential in the extracellular inter-neuronal space. It represents the summed neuronal activity, rather than the activity of any single neuron. In the integrate-and-fire system, oscillatory activity is characterised by repetitive and rhythmic neuronal spiking, which in turn causes the LFP to oscillate at a certain frequency, as shown in Figure 7. Currently a lot of research in this area is focused on discovering the contribution of each neuron to the LFP (see Einevoll et al. (2013) for a review, and Telenczuk et al. (2017) for a recent example of this research). However, for the purposes of my case study, I only need to show that the oscillatory activity of the network need not be a feature of individual neurons.

¹³ I do not review the empirical evidence for this claim in the thesis. See Dehaene et al. (2006) for details.

Independently of how each individual neuron contributes to the LFP, as long as the network oscillation does not necessarily characterise the behaviour of the neurons themselves, my interpretation of the case study stands. This is because my interpretation requires that the characteristic waxing-and-waning behaviour appears at the level of the ensemble of neurons even when such behaviour has not been built into and is not present in any of the components of the ensemble. In other words, my interpretation requires that oscillations are at the network level, not at the level of individual neurons. The reason for this is that if that is the case then I can show how the oscillatory behaviour of the network meets the first condition for irreducibility. In doing that, I will be able to show that mental properties are realised_{o-o} by a property of the network, and that a many-one approach cannot account for that relation. The reasoning for this, of course, will become clearer later on. Before that, I need to show that oscillations are at the network level.

7.2.1 The Oscillatory Activity of Neuronal Networks Is a Network Phenomenon

In this section I offer empirical evidence for the claim that oscillatory activity of a population of neurons is a network phenomenon. There are two ways in which neuronal oscillations can be understood to be a network phenomenon: 1) individual neurons do not have endogenous oscillations but they show oscillatory behaviour once they interact with other neurons—due to feedback loops, for instance—, and 2) individual neurons do not have endogenous oscillations and do not show oscillatory behaviour—at least not with the same pattern as the network's—so neuronal oscillations of this kind are only a network phenomenon. The latter is a much stronger way for oscillations to be only a network phenomenon, but it is precisely the understanding I need to show that the mechanism meets the first condition. None of the parts in the mechanism can themselves be oscillating with the same pattern as the network, neither endogenously nor exogenously.

There are several studies that use complex simulations to show that neuronal oscillations are a network phenomenon in the first understanding above. They show that neurons inside the network do not exhibit endogenous rhythmic activity. In general, those studies use feedback connections between various kinds of neurons that, under certain circumstances and given certain parameters, are able to synchronise the spiking of individual neurons so that the network exhibits oscillatory behaviour. Bhattacharya et al. (2021) is one of the most recent studies about how activity propagates in neuronal networks. Their model does not build oscillations into the

individual neurons yet is able to obtain “a broad class of spontaneous or evoked spatiotemporal wave patterns and oscillations in the cortex and thalamus. This architecture was able to sustain smooth waves in the cortex and lurching waves in the thalamus simultaneously.” (Bhattacharya et al., 2021, p. 14359). Another interesting study is by Izhikevich and Edelman (2008). This study has influenced much of the research in the field, as it makes use of what is known as the Izhikevich model of single neuron spiking dynamics, presented in Izhikevich (2003), to build a simulation of the whole brain. This is one of the most complex simulations that have been achieved to this date, and in it Izhikevich and Edelman show that oscillations emerge in the neuronal network even if not built into any of its parts. Although I will not delve into the details of those more complex simulations, it is important to mention that in both studies there are two distinct kinds of neurons: inhibitory neurons and excitatory neurons (Izhikevich and Edelman’s simulation has moreover various kinds of excitatory and inhibitory neurons). It is through the feedback connections of these various kinds of neurons with different firing patterns and threshold values that synchrony in the firing pattern is generated, and oscillations are obtained (Whittington et al. (2000) discusses in particular the role of inhibitory neurons in brain oscillations and rhythms).¹⁴ These simulations provide evidence that oscillation at the level of the network does not need endogenous oscillators as parts, but neurons still have oscillatory behaviour once they interact with other neurons in the network, and actually that is what gives rise to the oscillatory behaviour of the network. So that empirical evidence only shows that there are cases of network-level oscillations in the first understanding mentioned above.

At this point I need to mention that Dehaene, Sergeant and Changeux’s model does include a few neurons that are intrinsic oscillators (neurons that oscillate endogenously), although most of them have random spikes, but as they explain this is just a practical matter that does not affect the results:

The basic properties of our model persist even in the absence of intrinsic cellular oscillators. In that case, oscillations are much less obvious in single-neuron firing trains, but are a network property most apparent in the local field potential. Furthermore, our simulations use only a small number of cells. As the number of neurons per cortical area is increased, oscillations

¹⁴ Some research suggests that inferior olivary cells are intrinsic oscillators (Llinás, 2014). However, this does not invalidate the point of the evidence given so far, namely that intrinsic oscillators are not needed for neuronal oscillations to obtain.

become distributed and less conspicuous at the single-cell level. Overall, our simulations, although obviously very simplified, are compatible with experimental observations in awake cats, monkeys, and humans, which indicates that oscillations do occur in awake animals and humans, particularly under conditions of focused, effortful processing, and are typically more prominent in the local field potential than in the spiking activity of isolated cells. (Dehaene & Changeux, 2005, p. 920)

So far, I have explained that we do not need intrinsic oscillators to obtain rhythms and oscillations in a population of neurons. However, even if the oscillatory behaviour of neurons is not endogenous but due to interactions between individual neurons in the network, individual neurons could nonetheless spike following an oscillatory pattern. If that was the case, given that parts would also be oscillating at the same frequency, we would be able to find the properties of the oscillation at the network level already at the lower level of its parts. This is actually the case in Figure 7, where although we do not know whether the oscillations are endogenous or caused by interactions between neurons, individual neurons fire following a rhythm. Neurons' firing patterns show rhythmicity, and this is what generates the rhythms of the network. This would be a problem for my interpretation of the case study, as my main point is that there are properties of the network —i.e., middleman properties— that cannot be attributed to any of its parts, and that these properties are the realisation base of mental properties. In order to argue that the realisation base is a single lower-level property I need to show that the existence of such a property is justified. However, insofar as oscillations are already present at the level of individual neurons, a many-one realisation relation can be accepted. As I have argued in Chapter 4, this would undermine the subset view. So now I endeavour to show that there are cases where the second understanding of network-level oscillations is true, and rhythmicity in the firing patterns of the individual neurons composing the network is not found.

As the reader might have noticed, the last sentence in Dehaene & Changeux's quote above mentions that it has been found that individual neurons' rhythms are not as prominent as network oscillations. Actually, network oscillations are sometimes observed at a higher frequency than the firing frequency of the single neurons composing the network. Moreover, network rhythms have been observed even if neurons are firing stochastically, that is, in the absence of individual neuronal oscillations. Here are some quotes showing evidence of the

existence of the phenomenon —i.e. network oscillations with different single-neuron rhythms or irregular single-neuron rhythms—:

Single-cell discharge rates is typically much lower than the LFP oscillation frequency, especially in pyramidal cells but also in interneurons (Csicsvari et al. 1998, 1999b). Indeed rhythmicity is usually not apparent in the raw spike trains of individual cells and becomes visible only after data processing of spike trains from multiple single units. Thus single-cell behavior differs markedly from the population activity during fast network oscillations. Similarly, physiological studies of primates indicate that even when the LFP signal contains a clear rhythmic component, simultaneously recorded single-unit spike trains usually appear irregular and devoid of a clear-cut oscillation (Fries et al. 2001; Logothetis et al. 2001).

Recently, oscillations have been observed in vitro (Buhl et al. 1998; Fellous and Sejnowski 2000; Fisahn et al. 1998) that resemble these characteristics: strong gamma (30–40 Hz) oscillation of the LFP, together with low (2 Hz) and irregular firing in pyramidal cells. This means that a single pyramidal cell fires only once in every 15–20 cycles of the population rhythm. (Brunel & Wang, 2003)

On the face of it, it is difficult to think of network oscillations without any rhythmic behaviour of their parts: how can we achieve oscillations at the level of the network if the firing patterns of individual neurons are not rhythmic or if they oscillate at a different rate? Researchers have been working on models and mechanisms for network rhythms without rhythmicity in the firing patterns of neurons. I now briefly survey one of those models: Brunel & Wang's (2003) model. The details of the model are not crucial for my purposes, so I only briefly explain how network rhythms can be obtained without single-neuron rhythms. Figure 8 below shows several aspects of a simulated network of neurons showing oscillatory behaviour at a frequency of around 180Hz. The network is composed of inhibitory neurons (interneurons), with connections drawn randomly between neurons but making sure that the probability of connection between any two neurons was the same.

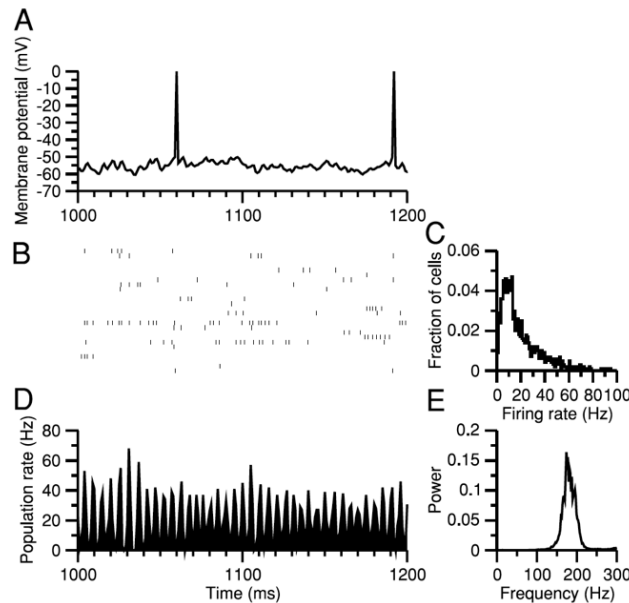


Figure 8. In a network of inhibitory integrate-and-fire neurons, there is a synchronous oscillation (D) at approximately 180 Hz (E), even if no single neuron is oscillating at this rate. The individual neurons in the network fire sporadically and irregularly, as shown in A (membrane potential of a neuron). The rastergram (B) shows low-rate and irregular spike trains from the individual neurons, with the specific neuron from the membrane potential graph shown at the bottom. The firing rate distribution (C) of single neurons across the population exhibits rates from 0 to 100 Hz. Figure extracted from Brunel & Wang (2003).

The first graph, A, shows the firing pattern of a single neuron. This neuron is not firing with any rhythm, rather its spikes are purely stochastic. B shows the firing patterns of some of the neurons in the network, which do not follow any particular rhythm and are highly irregular. We can see that the firing pattern of the neuron in A is depicted in the last row of B. C shows the distribution of firing rates of single neurons across the population, which ranges from 0Hz to 100Hz. Finally, D and E show the rhythm and frequency of the population of neurons. As can be clearly appreciated, the upshot of the neurons firing at irregular and stochastic rates is a single rate of firing of the population consistent with oscillations at 180Hz. How is this achieved? In this case, there is powerful recurrent inhibition due to the connection between inhibitory neurons. This means that neurons receive a strong inhibitory input that increases their firing threshold thus decreasing their firing likelihood, so they fire at lower rates even if they receive strong excitatory inputs from external excitatory neurons. Their firing is irregular and stochastic because the total current (excitatory inputs plus recurrent inhibitory) is subthreshold, and only fluctuations due to noise can cause the membrane current to be above the firing threshold. So where does the network oscillation stem from and what keeps it stable?

(...) the oscillation is stable because of the repetitive succession of the following events: at the peak of a cycle, there is strong inhibitory firing. After a time lag of 2.5 ms due to synaptic filtering, every neuron in the network feels a massive inhibitory input and activity goes down, hence the trough in global activity. Subsequently, 2.5 ms later, the synaptic currents decay away, the total input becomes high due to strong external stimulation, and there is another surge of activity. The period of the oscillation is therefore about two times the synaptic lag, i.e., 5 ms in this case. (Brunel & Wang, 2003, p. 422)

Briefly, even if individual neurons' firing is highly irregular and stochastic, the likelihood that they fire during an interval with high inhibition is very low. Instead, they are allowed by the network to fire only when the effect of the inhibition has faded, which means that the likelihood that the overall population fires at intervals of 5ms is extremely high. Not all of the individual neurons fire at the beginning of each cycle (only about 10%), as seen in graph B in the figure 8 above, but the key is that firing only takes place during the period of low inhibition, and this period is cyclic.

If we add excitatory neurons to the network, that is, in a two-population system of excitatory (pyramidal) cells and inhibitory cells (interneurons), we can obtain frequencies in the gamma range, which are much more common as I explained above.

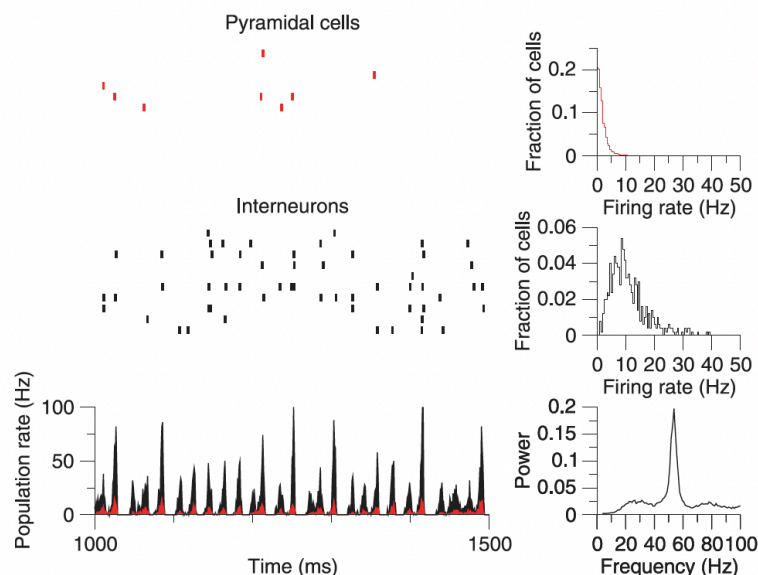


Figure 9. Pyramidal cells and interneurons display intermittent and irregular firing (as shown in the rastergrams). As can be observed in the population firing rate and its power spectrum, they nonetheless generate a coherent network rhythm. Figure extracted from Brunel & Wang (2003).

In this model, Brunel and Wang are able to simulate a network of neurons oscillating in the gamma frequency while none of the individual neurons composing the system fire at this rate. This finding is crucial in my interpretation of the Global Neuronal Workspace Case Study, as it clearly shows that the oscillatory behaviour, even if it strictly arises from the behaviour of individual neurons, may be found only at the level of the network of neurons. The key upshot of this finding is that oscillation in the network does not entail or require oscillation in the components. Bringing back my discussion of the Bechtel's case study, the relevance of this finding is clear. I have found a case where oscillations are a feature of the mechanism as a whole. If I can moreover show that these oscillations constrain the behaviour of the parts, I will be able to show what Bechtel failed to show: the mechanism cannot be reduced to any of its parts nor to a collection of them. To do that, I need to continue my exposition of the GNWT.

7.2.2 The Global Ignition State

My digression in the previous section had the aim of showing that the oscillations characteristic of conscious states are at the level of the network and cannot be attributed to any of the individual neurons. By the end of this subsection, I will have shown that there is a property of the network, namely the oscillatory phase, that is necessary to explain the interaction of the network with another network. But first, let us go back to the question about how GNW neurons enter the global ignition –or global broadcasting– state. In order to understand how GNW neurons enter the global ignition state, which is the state realising conscious states, I focus on Dehaene, Sergeant and Changeux's model and simulations, which are much easier to introduce than the studies presented above, yet they offer sufficient detail for my purposes. Remember that the model has four areas, A, B, C, and D, each constituted by a thalamocortical column made up of 120 neurons. As is shown in Figure 5, areas A and B are at the bottom of the hierarchy and correspond to bottom-up cortical processors. Dehaene & Changeux (2005) first simulate a network that only has areas A and B. The results obtained show that these areas are not sufficient to sustain activation of the network: activation quickly fades once the external stimulus disappears. However, this is not consistent with empirical findings, which show that stimulation after an external stimulus is not proportional to the duration of the stimulus, but that is sustained for longer. If we add areas C and D, the activation caused by an external stimulus can be sustained for a considerable duration. This happens because the stimulus — when above a critical duration— causes the activation of higher areas through bottom-up propagation, and these higher areas in turn produce top-down signals that reinforce the

feedforward signals that activated them in the first place. This creates what is known as the ignition of a reverberating neuronal ensemble, which is a state of the whole network that is sustained beyond the duration of the stimulus. In this model, C and D correspond to workspace neurons and show global ignition. Of course, this is a very simplified model:

In a more complete model of cortical connectivity, global ignition would propagate to cortical areas beyond areas C and D, thus broadcasting the identity of the input stimulus to many cerebral processors. We therefore propose that the simulated process of ignition provides a minimal but plausible neural mechanism for the sudden access of sensory stimuli into conscious perception. (Dehaene & Changeux, 2005, p. 915)

In the context of the GNW, what this model shows, nonetheless, is that top-down projections are necessary to account for conscious activity, and that these top-down projections cause an ignition state able to sustain activation beyond stimulus duration, thus enabling the system to broadcast the identity of the input stimulus to other processors. And this is what conscious perception is according to GNW theorists. The simulations following this model have two main findings. The first finding, which I have just mentioned, is that the simulation supports the claim that feedback top-down projections are necessary for conscious experience. The second finding is that the model is able to reproduce—and thus provide an explanation for—inattentional blindness and the attentional blink. According to the GNWT, these two phenomena are caused by how ensembles of neurons communicate with each other. For the remainder of this section, I focus on how communication between networks of neurons occurs and how this affects which information gets broadcasted, as well as on the implications this has for my interpretation of the GNW. One central feature of the global neuronal workspace is that it cannot broadcast more than one signal at a time: ignition of the GNW is all-or-none—either it enters a reverberating state that broadcasts a signal, thus there is a conscious experience, or the signal is not broadcasted at all, thus there is not conscious experience. But how does the global workspace decide which signal gets broadcasted when both of them have the same duration and are equally strong? This is a key feature of the GNWT, and I will use it to show that the oscillatory phase of a network is causally efficacious. Let me first explain how the GNWT explains the attentional blink (in this context, studies on the attentional blink are more detailed than those on inattentional blindness, so for this reason I focus only on the former).

7.2.3 The Attentional Blink

Dehaene, Sergent, and Changeux (2003) use the model introduced above to simulate the attentional blink. Roughly, the attentional blink is a real perceptual phenomenon caused when, under certain circumstances, attention is said to “blink”: subjects do not report seeing a stimulus during a period after attention was focused on another stimulus. One common circumstance, which is the one reproduced by Dehaene et al.’s model, is when subjects are asked to process two sequential targets, T1 and T2, where T2 appears only between 100ms and 500ms after T1. Crucially, T2 is capable of activating the correlates of areas A and B (for instance, areas P1, N1 and N4000 in the brain) but there is no further activation of areas C and D. According to the GNWT, T2 is unable to cause global ignition, and therefore subjects have no conscious perception of it. Let me recover the bottom part of Figure 5:

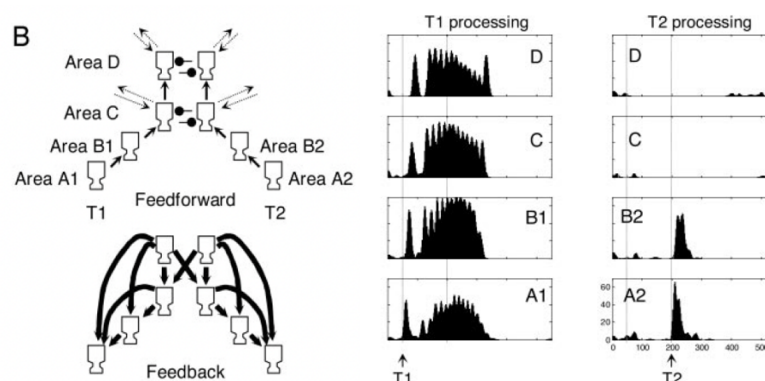


Figure 10. Representation of two targets, T1 and T2, entering the multicolumn system. The graphs represent a “blinked” trial with a 150ms lag between T1 and T2. T2 is presented while the GNW is still activated by T1 activity, so T2 fails to evoke any neural activity in areas C and D, corresponding to the GNW, and is confined in areas A and B. Figure extracted from Dehaene, Sergent & Changeux (2003).

In Figure 10 are represented the two targets, T1 and T2, entering the multicolumn system. The target T1 starts a bottom-up feedforward activation of A1 and B1 that, under certain conditions, triggers activation of further columns C and D. This causes the GNW to enter a global state of broadcasting T1, making it available to other processors. If the target T2 is shown to the subject between 100ms and 500ms after T1, it does activate areas A2 and B2, which are neuronal assemblies similar to A1 and B1. However, in cases of attentional blink, activation of areas A2 and B2 does not cause activation of areas C and D. In the model, this means that T2 is unable

to cause global ignition, thus it cannot be broadcasted, so it is not conscious. For this reason, most subjects do not report seeing T2, while they do report seeing T1. But why does T2 not trigger global ignition, if it has the same duration and strength as T1? The answer to this question is the last element I need to introduce to finally show my interpretation in terms of realisation₀₋₀ of this case study.

(...) ignition is more likely to occur when stimulus presentation coincides with the depolarized phase of spontaneous ongoing oscillations. (Dehaene, Changeux, 2005, p. 917)

The only reliable predictor [of ignition success or failure] was the phase of spontaneous oscillations in area A1 ($p < 0.045$, Kolmogorov- Smirnov test). Ignition was more likely to occur when the first stimulus-evoked spikes reaching area A1 coincided with a period of relative depolarization, thus increasing the probability of spiking activity which could then propagate to higher regions. Our simulations bear similarity to empirical observations by Fries et al. (2001), who observed fluctuating, occasionally synchronized gamma-band oscillations in cat area V1 whose phase predicted the latency of firing in response to an external stimulus. (Dehaene & Changeux, 2005, p. 916)

The way networks of neurons communicate —i.e., transfer signals to each other — is a key aspect of how the GNWT explains the attentional blink, and I want to briefly provide further insight into how communication is achieved. I use Pascal Fries' more detailed account of communication between neuronal groups. Fries has proposed a very influential model based on communication between networks through coherence of their oscillations, which he presents in detail in Fries, 2015.

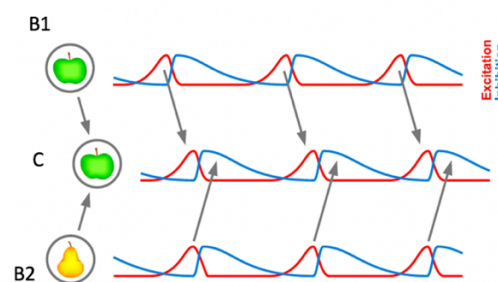


Figure 11. Communication through coherence in the multicolumn. The signal from B1 gets transmitted to C because the signal arrives to C in a depolarising moment. On the contrary, the signal from B2 arrives to C at a moment of inhibition, so it does not get transmitted. Importantly, all neuronal groups are oscillating at the same gamma frequency, and what matters to which signal is transmitted is the phase of the oscillation.

Figure 11 is really helpful to understand Fries' account of communication through coherence. I have adapted the figure to Dehaene et al.'s model discussed throughout this section. There are two different target stimuli—a pear and an apple—that, given that global ignition is all-or-none, are competing to be broadcasted. Only one of the two gets into the global neuronal workspace at a time. The deciding factor is the phase of the oscillation of the groups of neurons coding for the stimulus. Even if both groups of neurons, B1 and B2, are oscillating at the same gamma frequency, they interact with C with distinct phases. The signal from B1 gets transmitted to C because the signal arrives to C in a depolarising moment. On the contrary, the signal from B2 arrives to C at a moment of inhibition, so it does not get transmitted. The oscillatory phase, thus, is central to communication.

The explanation of the attentional blink offered by GNW theorists relies on this theory of communication. The phase of the oscillation the different areas are in determines whether a signal gets transmitted from one area to the other. The explanation of the attentional blink is the following. When T1 is received, A1 and B2 are in phase signalling T1. This signal gets transmitted to the GNW, and it causes it to oscillate at the same frequency. When T2 is received, areas A2 and B2, which are oscillating at a phase and frequency linked to T2, cannot cause the ignition of the global neuronal workspace because they are out of phase with the GNW: they arrive at the GNW at a moment of inhibition, and are not strong enough to go above the firing threshold needed for activation. This also occurs more generally whenever external stimuli interact with the spontaneous activity found in the GNW. When at rest, the GNW is in a state of continuous stochastic activity. Although this activity can serve as a catalyser of activation by external stimuli—as it randomly decreases the activation threshold—it can also block stimuli from entering consciousness: just as it decreases the threshold, it also increases it. So external stimuli are more likely to enter consciousness if their phase coincides with the phase of the spontaneous activity of the GNW. At this point, I have shown how a property of the ensemble of neurons, the phase of its oscillation, can enable or prevent a stimulus to become conscious.

7.3 The GNW Case Study and Realisation: O-O or M-O?

Now I have all the elements I need to argue that the maximally general account of realisation of conscious properties is realisation_{O-O} rather than realisation_{M-O} , as the former subsumes the latter.

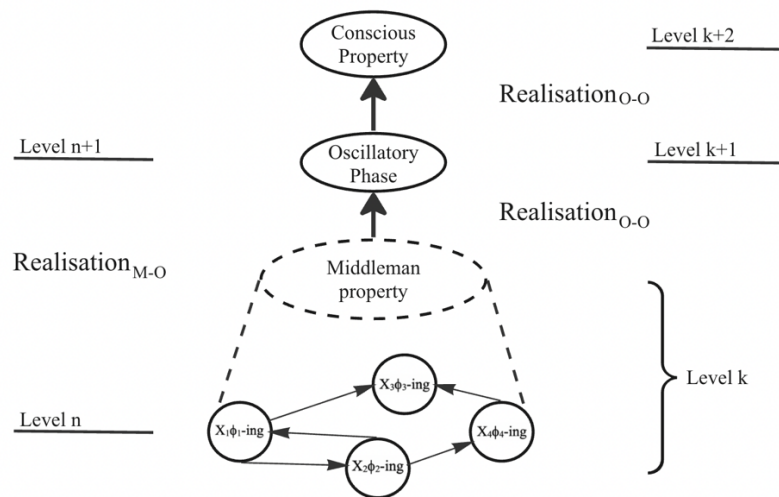


Figure 12. Representation of the properties and relations in play in the GNW case study.

Figure 12 is a modified version of the schema that guided my discussion of middleman properties in Chapter 4 and also presented at the beginning of this chapter. Now I am able to put it that schema in relation to the GNW case study. One question I can now answer is how many steps are needed to go from n to $n+1$ ($= k+1$). Gillett's answer would be only one step, as according to him in cases of the realisation of scientific properties one can go directly from the many lower-level entities composing a system to the higher-level property we want to account for, without introducing any relational middleman properties (level k). He justifies this claim by arguing that this is how the science actually works. Introducing a middleman property linking the many individual neurons and a property of the network of neurons is optional. It is true that one could do without it in this case. However, in this case study, there are reasons to deny that every higher-level property of the composite can be realised in this way: the realisation of conscious properties is, in the context of the GNWT, better accounted for as realisation_{O-O} between levels $k+1$ and $k+2$. Appealing to the operation of science does not provide reasons to deny the existence of realisation_{O-O} . On the contrary, science explains conscious experience by introducing an ignition event that triggers a reverberating state, which involves communication between neuronal networks. I have shown that communication

between neuronal networks requires them to be coherent, and this can only be explained in terms of the phase of the oscillation of the networks. In some cases, as I showed in 7.2.1, this behaviour is not attributable to any of the parts of the system. The science therefore is compatible and supports that the realiser of a conscious property is a property of the population of neurons, the oscillatory phase, at level $k+1$. This kind of property is not an ontologically lightweight middleman property, but a special science property that can be perfectly described and accounted for in scientific terms. All things considered, realisation_{O-O} subsumes realisation_{M-O} because it can account for all the cases of realisation realisation_{M-O} can account for, and more. Hence, realisation_{O-O} is the maximally general account of realisation.

Having seen the importance of communication for the overall argument, I want to mention that this argument is not restricted to the GNWT. This is a key point because there are other equally dominant scientific theories of consciousness such as Integrated Information Theory. Although GNWT is very influential, there is no agreement in the scientific community about which theory is superior. I claim that my conclusions apply to any theory of consciousness that relies on communication between groups of neurons as the central element that gives rise to consciousness.

I believe that the above interpretation of the case study might be vulnerable to the following claim. The oscillation of the network is not a genuine property, as it can be reduced to the behaviour of the neurons. If that was the case, conscious properties could be realised by the joint role-filling of the neurons propertied and related in a certain way. Even if the science uses oscillations to explain consciousness, this does not mean that ontologically we should consider oscillations to be anything else than mere neurons firing. To block this objection, I propose a stronger interpretation of the case study: oscillatory behaviour of a network of neurons is irreducible to the behaviour of the neurons themselves, so conscious properties can only be realised by a property of the network. If this is true, then not only is realisation_{O-O} compatible with the actual science to explain the realisation of conscious properties, but we could not do without it because the oscillatory properties that are the realisation base of conscious properties are irreducible to their components.

This stronger interpretation of the case study is inspired by Bechtel's account of mechanisms. Recall that Bechtel argued that mechanisms are irreducible to their parts. Although I have shown that his Circadian Rhythms case study is not convincing, in the remainder of this section

I show that the GNWT provides an example of an irreducible mechanism. In his case study, it was not clear at what level we should place the oscillations, as there was an ambiguity in the characterisation of oscillations as both properties of the parts (the concentration of PER) of the mechanism and properties of the whole mechanism. This was a problem for his account because oscillations are what provide the key to show that the mechanism is irreducible. Therefore, I concluded that his case study fails to show that the circadian rhythms mechanism is irreducible.

To reach the conclusion that conscious properties are only realised₀₋₀ by neuronal networks I need to show that in general conscious properties cannot be realised by the properties of neurons themselves. This requires showing that neuronal networks are mechanisms irreducible to their parts. Remember the two conditions I introduced in section 7.1 that must be met for a mechanism characterised by oscillatory behaviour to be irreducible to its parts:

Condition 1: The oscillation may be found at the level of the mechanism only.

Condition 2: The properties of the oscillation determine the behaviour of the parts of the mechanism.

Following Bechtel's reasoning, the mechanism as a whole in the GNW case study is the oscillatory behaviour of the network. Its parts are the individual neurons and their activities. In section 7.2 I have provided evidence for the following claims:

- Claim 1: Neuronal networks give rise to oscillations that sometimes are in a frequency not found in the individual neurons composing the network (be it because they fire at a different rate or because they fire at a stochastic and irregular rate). Therefore, in these cases oscillations of the neuronal networks are better understood as network phenomena.
- Claim 2: The oscillatory phase determines communication between neuronal networks.
- Claim 3: According to the GNWT, conscious properties are constituted by the availability of information for global broadcasting in the brain. For example, perceptual content becomes conscious when it is available for global broadcasting to many other cognitive processors brain-wide, thus making it available for verbal report, memory, etc.

Claim 1 can easily be used to show that the first condition is met. In some cases, the oscillatory behaviour of the network cannot be attributed to any of the single neurons in it, so an oscillatory behaviour at the level of the network does not require the same oscillatory behaviour in cells. Hence, it is true that, in general, the oscillation is a network-level phenomenon. Claim 2, with some elaboration, can be used to show that the second condition is met. The reasoning is as follows:

1. Whether an external input triggers global ignition of the GNW (e.g., areas C and D in Figure 5) and sustained activation of its neurons depends on whether the oscillatory phase of the specialised processors (e.g., areas A and B in Figure 5) that receive the external input and send it up the hierarchy is coherent with the oscillatory phase of the GNW (e.g., areas C and D).
2. The oscillatory phase of a neuronal ensemble is a property of the network.
3. Therefore, the behaviour of the individual neurons forming the GNW depends on a property of the mechanism as a whole, the oscillatory phase of the network.

We have here a case where the mechanism as a whole determines the behaviour of its parts: depending on whether the external input is in phase with the phase of the GNW, ignition will be triggered or not. Ignition leads to a period of sustained activation of neurons, which would not be activated if ignition had not occurred. Therefore, the oscillatory phase of the GNW determines the behaviour of its parts. Given that the two conditions are met, I conclude that a neuronal network is a higher-level irreducible mechanism.¹⁵

Now that I have shown that neuronal networks are irreducible mechanisms, Claim 3 straightforwardly leads to the conclusion that there is a single irreducible realisation_{O-O} base that cannot be reduced to the components constituting the network. The GNW case study shows that realisation_{O-O} is operative between levels k and $k+1$, and that it is the only kind of realisation operative between levels $k+1$ and $k+2$. This case study is able to block all the fronts

¹⁵ I want to note that this finding seems consistent with machresis, but I will not explore this further in this thesis. My overall conclusion, that nonreductive physicalism is a live view, and that the subset strategy provides an account of higher-level causation that can respond to Kim's CEA does not depend on and is not altered by there being cases of machresis.

of Gillett's Flat Objection. Parsimony does not apply in this case because there is no many-one equivalent to the realisation relation between levels $k+1$ and $k+2$, so there is no reasonable many-one relation that could show how network oscillatory properties should be dispensed in favour of properties of the cells as the realisation base. Also, I have shown that in realisation_{O-O} realisers are not "monster" properties, but properties that feature in the scientific theories of the phenomenon we are accounting for. All things considered, I have shown that realisation_{O-O} subsumes realisation_{M-O} . Thus, realisation_{O-O} is the maximally general account of realisation. Every case of realisation_{M-O} can be accounted for as realisation_{O-O} by positing a lightweight middleman property. Moreover, there are cases of realisation_{O-O} that cannot be accounted for as realisation_{M-O} , as the realiser is not a combination of entities propertied and related in a certain way, but a single property. I believe that this case study also requires the HDM modification of the subset view, as it is not compatible with Wilson's initial formulation of the subset view and her understanding of middleman properties as lightweight. The lightweighness of middleman properties requires that they be defined in terms of compositional principles and Boolean operations on the constituents. However, as I have argued above, the oscillatory phase of a neuronal ensemble is a property of the network, so it is a holistic property, in the sense that it can only be defined in terms of properties and states of the system in a narrow sense. The HDM view can accommodate this case: conscious properties are realised by holistically defined modal properties. The oscillatory phase is a holistically defined property because it is independent from the behaviour of individual neurons, and it results from the collective dynamics of the entire neuronal network. It is true that oscillatory properties of a network are not merely possible properties of it, as they are actual properties of it, but as I explained in Chapter 6, this is accounted for in the HDM view: actual properties of the network are still possible properties of it.

Although in Chapter 6 I have left open the question of whether entropy is actually weakly emergent, in this last chapter I will adventure to show that conscious properties are weakly emergent. My final conclusion, however, does not hinge on this, as it was not the aim of this chapter to show that consciousness is weakly emergent, but to find the best candidate for being its realiser base according to a scientific theory of consciousness. According to the GNWT, the best candidate for being the realisation_{O-O} base of conscious properties, then, is the property of being a network of neurons oscillating with a certain oscillatory phase. The oscillatory phase is what determines whether information enters the GNW and gets broadcasted to other cognitive processors. There are reasons to consider conscious properties to be weakly emergent

in light of the subset view, now that I have determined the best candidate for being their realisers. The main reason I believe this is that the property of being a network of neurons oscillating with a certain oscillatory phase arguably has more powers than the conscious property it realises.

For instance, consistent with the core claims of GNWT, conscious properties are multiply realised. A property functionally similar enough to ignition could realise consciousness as far as GNWT is concerned. Hence, we could use Wilson's proportionality argument. The property of being a network of neurons oscillating with a certain oscillatory phase is a general statement of a property that can actually be instantiated in many different ways. For instance, it can be instantiated as the property of being a network of Martian neurons oscillating with a certain oscillatory phase, P_2 . Or as a property of the human brain, P_1 . Not only that, but if we only consider the functional role of ignition and do not assume that it needs to take place in a network of neurons, even AI-based systems, P_3 , could realise consciousness. Clearly, the realisers, P_1 , P_2 , and P_3 feature in different causal relations: for example, P_2 can cause areas of the Martian brain to activate that are not present in the human brain, or in an AI-based system. Wilson's point was that when M is realised, in similar circumstances, by P_2 instead of P_1 , M will cause the same higher-level effects. This is because these higher-level effects are a manifestation of the causal powers of M and these powers are shared by P_1 , P_2 , and P_3 . The powers had by P_1 and not had by P_2 or P_3 make no difference. The fact that M has a distinctive power profile is sufficient to claim that M 's causal role regarding its higher-level effects is distinct from the causal role of its physical realisers.

7.4 Conclusion

In this chapter I have presented the GNWT about how consciousness arises in the brain. The conclusion of the chapter, which is presented in a weaker and a stronger form, is that the GNW case study supports realisation_{O-O} over realisation_{M-O} , and thus supports the viability of the subset strategy. The conclusion hinges on three findings, which roughly are the following: 1) neuronal networks' oscillations are a network-level phenomena, 2) oscillatory behaviour determines communication between ensembles of neurons, and 3) communication between ensembles of neurons is central to conscious properties. In my reading of the case study, conscious properties are weakly emergent. Although my overall conclusion does not depend on this finding, I believe it is an interesting result.

CONCLUSION

In this thesis, I have argued that a modified version of Wilson's subset view, the HDM view, is the most promising nonreductive physicalist account to accommodate not only mental causation but also other kinds of higher-level causation. Moreover, nonreductive physicalism has been reinforced as a live and tenable position. I have not endeavoured to demonstrate nonreductive physicalism's superiority against competitors, such as reductive physicalism or strong emergence. Rather, my goal has been to investigate how nonreductive physicalism should be spelled out assessing what I consider the most interesting and successful candidates for the best nonreductive physicalist account. The three candidates are Jessica Wilson's subset view, Sydney Shoemaker's mereological version of the subset view, and Carl Gillett's mutualism.

Firstly, I introduced the subset view and defended it against one of the most potentially damning and recurrent objections against it: the dilemma between powerful higher-level properties entailing overdetermination or lacking causal efficacy. If higher-level properties are causally efficacious, the objection goes, then there is problematic overdetermination. If we restrict their causal powers to avoid such problematic overdetermination, then we are left with an underwhelming kind of causal efficacy. My response to this dilemma is that the kind of causal efficacy the subset view attributes to higher-level properties to avoid problematic overdetermination is sufficiently distinct from that of its lower-level realisers. Having a distinct causal profile, although not having new powers vis-a-vis their realisers, is enough for higher-level properties to be causally efficacious in a significant way.

Secondly, I presented Shoemaker's and Gillett's accounts and concluded that Wilson's view is superior to them. It is superior to Shoemaker's account because it does not rely on there being a parthood relation between realiser and realisee. It is superior to Gillett's account because it does not require reformulating the Causal Closure of the Physical, providing a new conception of higher-level causal efficacy in terms of determination, nor introducing a new inter-level relation like machresis. However, I accepted that Gillett's Problem of Qualitative Distinctness, or Flat Objection for short, is a salient problem for the subset view that requires detailed and careful exploration. The problem targets the realisation of special science properties. Gillett

argues that science explains the behaviour of special science entities in terms of the joint behaviour of their constituents, and this is taken to show not only that special science properties of composites are realised by the properties of its constituents, but that properties at different levels do not share any powers —i.e., they are qualitatively distinct. That claim is a direct threat to the subset strategy, which requires power sharing, and according to Gillett is what makes it an untenable view.

I suggested that to assess the Flat Objection we must focus on whether realisation is a many-to-one or a one-to-one relation. Gillett argues that it must be many-to-one, given that this is the only way to preserve qualitative distinctness between realiser and realisee. Wilson's way to deal with the apparent many-to-one nature of realisation of special science properties is to provide an ontologically lightweight structural property that bridges the many properties of constituents and the higher-level realised property of the composite. She refers to this kind of properties as "middleman properties". To distinguish between these two kinds of realisation I introduced the terminology "realisation_{O-O}" to refer to the realisation of a property by a single middleman property, and "realisation_{M-O}" to refer to the realisation of a higher-level property of a composite by the properties of its constituents related in a certain way. I showed that each realisation relation bridges a different kind of level. Realisation_{M-O} bridges levels of mechanism, and realisation_{O-O} bridges what Craver calls levels of being or realisation levels. This helped me provide a response to Gillett's Flat Objection on behalf of Wilson, as I argued that the Flat Objection misses its target by focusing on levels of mechanism where the subset view is only concerned with levels of realisation. Having shown that the subset view can deal with the Flat Objection, a question arises: in the case of the realisation of special science entities, which account of realisation is superior, realisation_{M-O} or realisation_{O-O}? In the theoretical domain each view has virtues, but the burden is on the subset view to show that middleman properties are not spurious. For this reason, I have discussed two scientific case studies. Although Wilson and Gillett discuss case studies themselves, I have argued that they do not do so in an illuminating way. To show that a particular view is supported by a scientific case study, one needs to engage with the science thoroughly. I have discussed two case studies of realisation relations in the sciences: the realisation of entropy and the realisation of conscious properties.

I have argued that the realisation relation that accounts for both the realisation of entropy and conscious properties is realisation_{O-O}, whereas realisation_{M-O} can, at best, only partially account

for them. For this reason, I concluded that realisation_{O-O} is the maximally general account of realisation. That conclusion is based on one main finding: in both cases there are scientific grounds to accept the genuine existence of middleman properties, although the definition of middleman property needs to be modified. I have proposed a new definition of realisation: the Holistically Defined Modal view of realisation —the HDM view. According to this view, the realiser is a holistically defined modal property. Both the modal aspect and the holistic aspect of the view are defined in a broad sense, that also encompasses cases where a narrow definition applies. The holistic aspect of the view entails that the realiser is defined in terms of properties and states of the system, where those properties and states of the system might or might not be further defined in terms of compositional principles and Boolean operations on its constituents. The modal aspect of the view entails that the realiser has a modal aspect, namely, it takes into account possible states the system could be in. However, there are cases where the only possible states of the system considered are actual states, which nonetheless are possible states of the system. In the Global Neuronal Workspace case study, I have moreover shown that in my reading of the case study, conscious properties are weakly emergent.

REFERENCES

- Alexander, S. (1950). *Space, Time, and Deity. The Gifford lectures at Glasgow 1916-1918* (Vol. I & II). New York: The Humanities Press.
- Audi, P. (2012). Properties, Powers, and the Subset Account of Realization. *Philosophy and Phenomenological Research*, 84, 654 – 74.
- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.
- Bain, A. (1870). *Logic*. London: Longmans, Green, Reader, & Dyer.
- Baysan, U. (2016). An Argument for Power Inheritance. *The Philosophical Quarterly*, 66, 383–390.
- Baysan, U. (2018). Review: *The Multiple Realization Book* by Thomas W. Polger and Lawrence A. Shapiro. *Analysis*, 78, 177 – 180.
- Bechtel, W. (2017). Explicating Top-Down Causation Using Networks and Dynamics. *Philosophy of Science*. 84, 253 – 274.
- Bedau, M. A. (1997). Weak emergence. *Philosophical Perspectives*, 11, 375 – 399.
- Bernstein, S. (2016). Overdetermination Underdetermined. *Erkenntnis*, 81, 17 – 40.
- Bernstein, S., & Wilson, J. (2016). Free Will and Mental Causation. *Journal of the American Philosophical Association*, 2, 310 – 331.
- Bhattacharyya, K., McLean, D.L., MacIver, M.A. (2021) Intersection of motor volumes predicts the outcome of ambush predation of larval zebrafish. *The Journal of experimental biology*. 224, jeb235481.
- Blundell, S., and Blundell, K. (2006). *Concepts in Thermal Physics*. Oxford: Oxford University Press.
- Bowley, R. and Sánchez, M. (1999). *Introductory statistical mechanics*. Oxford: Oxford University Press.
- Brunel, N. and Wang, X.J. (2003) What Determines the Frequency of Fast Network Oscillations with Irregular Neural Discharges? I. Synaptic Dynamics and Excitation-Inhibition Balance. *Journal of Neurophysiology*, 90, 415-430.
- Callender, C. (1999). *Reducing thermodynamics to statistical mechanics: The case of entropy*. *Journal of Philosophy* 96, 348–373.
- Chalmers, D. J. (2006). Weak and strong emergence. In Davies, P. & Clayton, P. (eds.),

The Re-Emergence of Emergence. Oxford: Oxford University Press.

- Clapp, L. (2001). Disjunctive Properties; Multiple Realizations. *The Journal of Philosophy*. 98, 111–36.
- Craver, C. (2001). Mechanisms, and Hierarchy. *Philosophy of Science*. 68, 53–74.
- Craver, C. (2007). Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience. Oxford: Clarendon Press.
- Craver, C. & Wilson, R. (2007). Realization: Metaphysical and Scientific Perspectives. In P. Thagard (Ed.), *Handbook of the Philosophy of Science, Philosophy of Psychology and Cognitive Science*. North-Holland, 81–104.
- Craver, C. & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*. 22, 547–563.
- Crook, S. & Gillet, C. (2001). Why Physics Alone Cannot Define the 'Physical': Materialism, Metaphysics, and the Formulation of Physicalism. *Canadian Journal of Philosophy*, 31, 333 – 359.
- Dehaene, S., Sergent, C., and Changeux, J.P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences*, 100, 8520–8525.
- Dehaene, S., & Changeux, J. P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS biology*, 5, 910–927.
- Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in cognitive sciences*, 10, 204–211.
- Dehaene, S. (2010) Conscious and Nonconscious Processes: Distinct Forms of Evidence Accumulation? In C. Engel, and W. Singer (Eds.), *Better Than Conscious? Decision Making, the Human Mind, and Implications For Institutions*. Cambridge, MA: The MIT Press.
- Dehaene S., Changeux J.P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 28, 200–227.
- Einevoll, G. T., Kayser, C., Logothetis, N. K., & Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature reviews. Neuroscience*, 14, 770–785.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science (New York, N.Y.)*, 291, 1560–1563.

- Gillett, C. (2002). The Dimensions of Realization: A Critique of the Standard View. *Analysis*, 62, 316–323.
- Gillett, C. (2006). Samuel Alexander’s emergentism: or, higher causation for physicalists. *Synthese*, 153, 261–296.
- Gillett, C. (2010). Moving beyond the subset model of realization: The problem of qualitative distinctness in the metaphysics of science. *Synthese*, 177, 165 – 192.
- Gillett, C. (2016). *Reduction and Emergence in Science and Philosophy*. Cambridge: Cambridge University Press.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14, 1569–1572.
- Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 3593–3598.
- Kim, J. (1992). Downward causation in Emergentism and Nonreductive physicalism. In Bekermann, A., Flohr, H., Kim, J. (eds.). *Emergence or reduction? Essays on the prospectus of Nonreductive physicalism*. Berlin, New York: Walter de Gruyter.
- Kim, J. (1993a). The Nonreductivist's Troubles with Mental Causation. In Kim, J. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- Kim, J. (1993b). The Myth of Nonreductive Materialism. In Kim, J. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- Kim, J. (1999). Making Sense of Emergence. *Philosophical Studies*. 95, 3 – 36.
- Kim, J. (2005). *Physicalism, or something near enough*. Princeton: Princeton University Press.
- Kim, J. (2006). Emergence: core ideas and issues. *Synthese*, 151, 547 – 559.
- Frigg, R. & Werndl, C. (2021). Can Somebody Please Say What Gibbsian Statistical Mechanics Says? *British Journal for the Philosophy of Science*. 72, 105–129.
- Lazzerini Ospri L., Prusky G., Hattar S. (2017) Mood, the Circadian System, and Melanopsin Retinal Ganglion Cells. *Annual Review of Neuroscience*, 25, 539–556.
- Lewes, G. H. (1875). *Problems of Life and Mind*. Boston: James R. Osgood and Company.
- Llinás R. R. (2014). Intrinsic electrical properties of mammalian neurons and CNS function: a historical perspective. *Frontiers in cellular neuroscience*, 8, 320.
- Lloyd Morgan, C. (1923). *Emergent Evolution*. London: Williams and Norgate.
- Mashour, G. A. et al. (2020) Conscious Processing and the Global Neuronal Workspace

- Hypothesis. *Neuron*, 105, 776–798.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412, 150–157.
 - McLaughlin, B. (1992). The rise and fall of British emergentism. In Bekermann, A., Flohr, H., Kim, J. (eds.). *Emergence or reduction? Essays on the prospectus of Nonreductive physicalism*. Berlin, New York: Walter de Gruyter.
 - McLaughlin, B. (2009). Mental Causation and Shoemaker-Realization, *Erkenntnis*, 67, 149 – 72.
 - Mill, J. S. (1859). *The System of Logic*. New York: Harper & Brothers.
 - Morris, K. (2011). Subset realization, parthood, and causal overdetermination. *Pacific Philosophical Quarterly*, 92, 363–379.
 - Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World.
 - Ney, A. (2016). Microphysical Causation and the Case for Physicalism. *Analytic Philosophy*, 57, 141–164.
 - Papineau, D. (2008). Must a Physicalist be a Microphysicalist?. In Jakob Hohwy & Jesper Kallestrup (eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford: Oxford University Press.
 - Penrose, R. (1989). *The emperor's new mind*. Oxford: Oxford University Press.
 - Pineda, D., and Vicente, A. (2017). Shoemaker's Analysis of Realization: A Review. *Philosophy and Phenomenological Research*, XCIV, 97 – 120.
 - Povich, M. & Carl, C. (2017). Mechanistic Levels, Reduction, and Emergence. In Stuart Glennan & Phyllis McKay Illari (eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. Routledge. 185-97.
 - Putnam, H. (1967). The nature of mental states. Originally published as “Psychological Predicates” in W. H. Capitan & D. D. Merrill, (eds.) *Art, Mind, and Religion*, University of Pittsburgh Press, pp. 37 – 48. Reprinted in D. Chalmers (ed.), *Philosophy of Mind. Classical and Contemporary Readings*. Oxford: Oxford University Press, 2002.
 - Robertson, K. (2019). *Reductive Aspects of Thermal Physics*. PhD Thesis.
 - Sekerka, R. (2015). *Thermal Physics: Thermodynamics and Statistical Mechanics for Scientists and Engineers*. Amsterdam: Elsevier.
 - Shoemaker, S. (1980). Causality and Properties. In Peter van Inwagen (Ed.), *Time and Cause*. Dordrecht: D. Reidel.
 - Shoemaker, S. (1998). Causal and metaphysical necessity. *Pacific Philosophical Quarterly*,

79, 59–77.

- Shoemaker, S. (2001). Realization and Mental Causation. In Gillett, C. and Loewer, B. (Eds.), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.
- Shoemaker, S. (2007). *Physical realization*. Oxford: Oxford University Press.
- Shoemaker, S. (2011). Realization, Powers, and Property Identity. *The Monist*, 94, 3 – 18.
- Shoemaker, S. (2013). Physical Realization without Preemption. In S. C. Gibb, E. J. Lowe, R. D. Ingthorsson (Eds.), *Mental Causation and Ontology*. Oxford: Oxford University Press.
- Swendsen, R. H. (2019). *An Introduction to Statistical Mechanics and Thermodynamics*. Oxford Graduate Texts.
- von Schantz, M., Archer, S. N. (2003). Clocks, genes and sleep. *Journal of the Royal Society of Medicine*, 96, 486–489.
- Teleńczuk, B., Dehghani, N., Le Van Quyen, M., Cash, S. S., Halgren, E., Hatsopoulos, N. G., & Destexhe, A. (2017). Local field potentials primarily reflect inhibitory neuron activity in human and monkey cortex. *Scientific reports*, 7, 40211.
- Watkins, M. (2002). *Rediscovering Color*. Dordrecht: Kluwer.
- Wallace, D. (2018). The Necessity of Gibbsian Statistical Mechanics. [Preprint].
- Whittington, M. A., Traub, R. D., Kopell, N., Ermentrout, B., & Buhl, E. H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 38, 315–336.
- Wilson, Jessica (1999). How Superduper Does a Physicalist Supervenience need to Be?. *The Philosophical Quarterly*, 49, 33-52.
- Wilson, J. (2006). On Characterizing the Physical. *Philosophical Studies*, 131, 61–99.
- Wilson, J. (2009). Determination, Realization, and Mental Causation, *Philosophical Studies*, 145, 149 – 169.
- Wilson, J. (2010). Non-reductive Physicalism and Degrees of Freedom. *British Journal Philosophy of Science*, 61, 279–311.
- Wilson, J. (2011). Nonreductive Realization and the Powers-based Subset Strategy. *British Journal for the Philosophy of Science*, 94, 121 – 154.
- Wilson, J. (2013). Nonlinearity and Metaphysical Emergence. In Stephen Mumford and Matthew Tugby (Eds.), *Metaphysics and Science*, 201–35. Oxford: Oxford University Press.
- Wilson, J. (2015). Metaphysical Emergence: Weak and Strong. In T. Bigaj and C. Wuthrich

(eds.), *Metaphysics in Contemporary Physics: Poznan Studies in the Philosophy of the Sciences and the Humanities*. Leiden: Brill.

- Wilson, J. (2021), *Metaphysical Emergence*, Oxford: Oxford University Press.
- Yablo, S. (1992). Mental Causation. *The Philosophical Review*, 101, 245 – 280.
- Yates (2016). Demystifying Emergence. *Ergo*, 3, 809 – 841.
- Zhai, Z., & Akturk, I. (2020). Exploiting Refractory Period for Functional Multiplexing and Short-Term Memory in Spiking Neural Networks. [Preprint]