UNIVERSITÉ DU QUÉBEC À MONTRÉAL

BIOLOGICAL EXPLANATIONS OF BEHAVIOUR: THE PITFALLS OF EXPLANATORY PLURALISMS, AND A DEFENSE OF TEMPERED INTEGRATIVE MONISM

DISSERTATION
PRESENTED
AS PARTIAL REQUIREMENT
OF THE DOCTORATE IN PHILOSOPHY

BY
ERIC MUSZYNSKI

JUNE 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

EXPLICATIONS BIOLOGIQUES DU COMPORTEMENT : LES ÉCUEILS DES PLURALISMES EXPLICATIFS, ET UNE DÉFENSE DU MONISME INTÉGRATIF TEMPÉRÉ

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN PHILOSOPHIE

PAR
ERIC MUSZYNSKI

JUIN 2024

ACKNOWLEDGEMENTS

This project would not have been possible without the unconditional help and support I have received from the part of my supervisor, Christophe Malaterre. Through the many years which this has taken, he has been an unwavering presence, always there to listen, to read, to contribute, and to take my ideas seriously, all the while giving me enough liberty to pursue what I thought was best. Never once through all the challenges I have faced have I felt him lose confidence in my ability and willingness to bring this project to term, and for that I will be eternally grateful.

Christophe Malaterre also contributed in other ways, namely by creating the Laboratoire de philosophie des sciences (LAPS), a place promoting collaborative work among graduate students and faculty. Not only did this give me a dedicated workspace, but it also allowed me to create a community of researchers and break the isolation that can so easily become the norm during dissertation writing. The people I met during those years have radically changed my life for the better.

Luc Faucher and Denis Réale provided valuable feedback every step of the way, constructively highlighting the weakest parts of my argumentation, allowing me to constantly move forward.

This research was also made possible with help from the SSCHR Joseph-Armand Bombardier Canada Graduate Scholarship, as well as financial support from Christophe Malaterre's Chaire de recherche du Canada en philosophie des sciences de la vie, and the Centre interuniversitaire de recherche en sciences et technologie (CIRST).

I need to give special thanks to Cloé Gratton, who is a constant inspiration and a model of hard work and dedication. Her patience and encouragements made the final stretch not only possible but also exciting, and our discussions which continually rehashed the same ideas helped me put everything into perspective, and put everything on paper.

I would also like to thank all the people with whom I have had interesting and constructive discussions regarding my research over the years, including Helen Longino, Eric Hochstein, J.C. Cahill, Stéphanie Ruphy, Kenneth Waters, Eric Turkheimer, Cristian Larroulet, Sarah Arnaud, Baptiste Bedessem, Nicolas Bernier, Anne-Marie Gagné-Julien and all the participants who accepted our invitation to the Biology of Behaviour workshop held at UQAM in 2018. I would also like to thank the students and staff of the CIRST, and specifically Martine Foisy, for welcoming me into the community with open arms. There are also many others I have met at conferences and elsewhere, who contributed through incisive remarks or questions, whose names I have unfortunately forgotten.

DEDICATION

To my mother and father, who modeled not only how to question, but also how to find answers.

This is also dedicated to Cloé, Misha, and Louka, who are an infinite source of motivation and joy.

TABLE OF CONTENTS

LIST OF FIGURES

RÉSUMÉ

Le pluralisme scientifique est l'idée que la science est, et, pour certain-es au moins, sera toujours, caractérisée par la pluralité plutôt que l'unité. Les défenses récentes du pluralisme se sont concentrées sur le pluralisme explicatif, en particulier dans le domaine des explications biologiques du comportement. Il peut cependant être difficile de comprendre précisément à quoi la pluralité est attribuée. Pour cette raison, je propose dans cette thèse une nouvelle typologie des pluralismes explicatifs et deux exemples de ce pluralisme sont examinés de manière critique. Sandra Mitchell défend une forme de « pluralisme de fragmentation », qui propose que, bien qu'il soit possible d'intégrer certains aspects de la recherche scientifique, il n'y a pas d’unification globale possible des explications scientifiques de la biologie du comportement, et il y a des limites à ce qui peut être intégré. De son côté, Helen Longino défend une forme de « pluralisme insulaire », qui propose qu’un phénomène donné peut être expliqué par de multiples explications incompatibles et être toutes considérées comme adéquates malgré leur incompatibilité. Je montrerai que ces deux défenses du pluralisme explicatif échouent, chacune pour leurs raisons respectives.

Pour répondre à ces types de pluralismes explicatifs, je propose une défense d’un type de monisme pour les explications biologiques du comportement, basé en grande partie sur la notion d’intégration. L'intégration est comprise comme la combinaison de diverses unités épistémiques qui servira à élargir ou approfondir notre compréhension d'un phénomène donné; et ces unités épistémiques (telles que des théories, des explications, des modèles, des résultats ou des données) peuvent découler de diverses approches. Le monisme proposé repose sur l'idée que la recherche scientifique dans le domaine de la biologie du comportement sera unifiée non pas par une grande théorie ou une explication de tout, mais par les liens entre les multiples explications locales des phénomènes, qui seront reliées les unes aux autres grâce à l'intégration. Je propose que l'intégration sera toujours en principe possible, offrant un remède à tout prétendu pluralisme explicatif, en brisant les barrières entre les explications et en réconciliant les approches précédemment considérée incompatibles. Ce monisme est toutefois tempéré par certaines limites, mais j’explique que ces limites ne sont pas des limitations en principe, mais plutôt dues à des préoccupations pratiques concernant les expériences qu'il est possible de réaliser.

Mots clés : pluralisme explicative, monisme explicatif, intégration, incommensurabilité, explication scientifique, biologie du comportement.

ABSTRACT

Scientific pluralism is the idea that science is, and, for some at least, always will be, characterized by plurality rather than unity. Recent defenses of pluralism have focused on explanatory pluralism, especially in the domain of biological explanations of behaviour. It can, however, be difficult to understand precisely what the plurality is attributed to. For this reason, I propose in this thesis a new typology of explanatory pluralisms and two exemplars of this pluralism are critically evaluated. Sandra Mitchell defends a form of 'fragmentation pluralism', which proposes that while it will be possible to integrate certain aspects of scientific research, there is no possible broad unification to scientific explanations of the biology of behaviour, and there are limits to what can be integrated. Helen Longino, for her part, defends a form of 'insular pluralism', which proposes that incompatible explanations for a given phenomenon can co-exist, and all be considered successful despite the incompatibility. I will show how these two defenses of explanatory pluralism are deficient, each for their respective reasons.

To answer these kinds of explanatory pluralism, I propose a defense of a type of monism for biological explanations of behaviour, based in large part on the notion of integration. Integration is understood to be the bringing together of various epistemic units (such as theories, explanations, models, results, or data) stemming from various approaches, used to broaden or deepen our understanding of a given phenomenon. The monism proposed is predicated on the idea that scientific research in the biology of behaviour will be unified not through some grand theory or explanation of everything, but through the links between the multiple, local explanations of phenomena, which will be joined one with the other through integration. I propose that integration will always in principle be possible, providing a remedy to any purported explanatory pluralism, breaking down barriers between explanations, and reconciling approaches previously considered incompatible. This monism is tempered however by certain limits, but I explain how those limits are not in-principle limitations, but rather due to practical concerns regarding the experiments it is possible to carry out.

Keywords : explanatory pluralism, explanatory monism, integration, incommensurability, scientific explanation, biology of behaviour.

INTRODUCTION

Scientific pluralism has recently become something of a default stance for many philosophers of science. Whereas early- and mid-twentieth-century philosophy of science was characterized by the search for unifying themes, methods, or types of explanations (e.g. Hempel, 1965; E. Nagel, 1961; Oppenheim & Putnam, 1958), the last decades have seen these projects get abandoned in favour of a recognition of the purportedly evident pluralism in the sciences. One need only take a cursory look at the state of contemporary scientific research to see a dazzling array of methods, of interests, of explanations, and an apparently ever-growing number of disciplines and sub-disciplines, each more specialized than the next. Faced with this apparent variety, many philosophers of science have defended some form of scientific pluralism, arguing that science is characterized by diversity, as opposed to unity.

This pluralism has been characterized in multiple ways, some emphasizing the methodological pluralism, arguing that scientific research is carried out in many different ways, with no single common method (e.g. Dawkins, 1976; Lloyd, 1989, 2005; Okasha, 2006; Sober, 1990; Sterelny & Kitcher, 1988). Others have defended ontological pluralism, proposing that the entities posited by scientific theories or explanations are many, and need not concord with those posited by other theories or explanations (Cartwright, 1999; e.g. Dupré, 1993; Waters, 2017). Though these two kinds of pluralism will be briefly touched upon, this thesis concerns more specifically purported *explanatory* pluralism, or the idea that there exists a plurality of scientific explanations. This explanatory pluralism has been depicted in many different ways, and indeed, the first challenge when tackling this topic is to understand more precisely what is at stake, and what, exactly, the various authors are defending.

Explanatory pluralism has often been defended in the context of biological explanations of behaviour (see e.g. Aizawa & Gillett, 2019; Campaner, 2014; Kellert et al., 2006a; Longino, 2002, 2013; Mitchell, 2003, 2009). This branch of research is particularly interesting for explanatory pluralists, since explanations of behaviour are almost invariably complex. Any number of mechanisms or explanations could be brought to bear, and indeed, behaviours are more often than not the result of an incredible sum of causes. Behaviour has also been attributed to an enormous number of species and individuals, from humans, to dogs, to insects, to plants, and even bacteria (Levitis et al., 2009). This means that the explanations given for one species or category of organisms may not be transposable to another, revealing yet another potential source of explanatory plurality. For instance, when scientists attempt to explain aggressivity, whether in humans or in animals, they can call on all manner of causes, whether genetic, developmental, neurobiological, environmental, or any other type of cause that is found to be relevant. As such, many different disciplines or approaches can be called on, including but not limited to molecular biology, neurobiology, quantitative behavioural genetics, socio-environmental approaches, investigations into the evolutionary history, or current adaptedness of a given behaviour. Explanatory pluralists will look at this diversity of approaches and explanations, and conclude that it is the result of successful science running its course. Pluralists will contend that the diversity is here to stay, and that attempts at unifying approaches or explanations are bound to fail at some point, implying that science is and always will be characterized by pluralism. In contrast, monists will propose that there are, or will be, ways of unifying explanations or approaches, and that this will tend towards the unification of science, rather than its fragmentation. (Other pluralists propose that the diversity is temporary, and that future science will find ways to unify the explanations, but those who propose that pluralism is permanent accuse the temporary pluralists of being closet monists.) This means that for pluralists, explanations of aggressivity, for example, will always call upon multiple causes, and that there will be ways in which it will be impossible to integrate those causes to yield a unified explanation, either because the explanations are at different, unreconcilable levels, because various models approximate actual causes in different ways, or because the different approaches produce incommensurable explanations. The specific reasons for upholding pluralism differ from one author to the other, and imply different conceptions of pluralism; these distinctions will be elaborated on throughout the next chapters.

In this thesis, I will both attack explanatory pluralism, and defend of a form of explanatory monism. I too focus on biological explanations of behaviour since they are taken by many to be one of the bastions of explanatory pluralism; I am thus taking the pluralists to task on their own ground. If I can show that the forms of explanatory pluralism they defend do not hold up to scrutiny even for biological explanations of behaviour—one of the domains of inquiry which most clearly calls on contributions from a multiplicity of approaches—then that gives us good reasons to suppose that it will not hold up in other domains of inquiry.

The initial impetus for this thesis was the recognition of a confusion regarding what it is that explanatory pluralism means, and what it implies regarding what we could know about the world. While many philosophers endorse explanatory pluralism, it can be difficult to understand what they understand to be plural: is it that there can be a plurality of types of explanations? Could there be a plurality of compatible explanations for a single phenomenon? Or even a plurality of *incompatible* explanations for a given phenomenon? And can those incompatible explanations be said to all be true despite being incompatible? If explanatory pluralism does imply that there can be successful yet incompatible explanations of a given phenomenon, what does that mean regarding the role of science in our understanding of the world? The repercussions of the answers to these questions are wide-ranging, both for epistemic questions, as well as social ones. If we take for granted that successful scientific explanations tell us something truthful about the world, and if pluralism holds, then that means either that the world is disunified in some sense, or that our successful explanations are only ever approximate or incomplete. The pluralists which I tackle in this thesis deny the former conclusion, and adopt the latter, proposing that in some meaningful way, scientific explanations are limited when it comes to explaining the world as it actually is, giving enough 'room' for a permanent diversity of explanations about certain phenomena. On this account, science, by its very nature, is limited in what it can explain, recasting its ultimate purpose not as giving us complete explanations of phenomena, but only piecemeal, partial explanations. On a social level, emphasizing pluralism about science could also open the door to relativism, or the idea anyone can in some sense have their own truth or their own explanation; this conception of truth can of course have disastrous consequences, leading to disinformation and breaking down avenues for discourse. It is important to understand however that pluralists themselves are explicitly not relativists, trying instead to find a way of accounting for the plurality seen in science *without* falling into relativism. Nevertheless, even without falling into relativism, pluralism changes our understanding of what science can tell us about important social issues, opening the door—albeit slightly—to the possibility of having competing yet truthful accounts for a given phenomenon, which can lead to additional difficulties in reaching consensual solutions to our some pressing problems. These and many other questions led me to delve into pluralist writings to understand what the position is, and whether explanatory pluralism is tenable.

The conclusion of this immersion in the literature, and what will be argued in this thesis, is that explanatory pluralism is not tenable for biological explanations of behaviour. In order to arrive at this conclusion, I take a two-pronged approach: the first is to critically evaluate the two most well-known explanatory pluralist positions, namely those defended by Sandra Mitchell (2003, 2009) and Helen Longino (2002, 2013; Kellert et al., 2006b). Each defends her own particular kind of explanatory pluralism, defended by its own set of arguments, and as such are treated separately. Both however emphasize how scientific explanations are *partial* representations of reality, and as such, distort in some respects how that reality is represented, opening the door to the plurality of explanations. My aim is to show the flaws in their positions, and how their arguments do not in fact lead to a convincing defense of explanatory pluralism. The second prong of my approach is to call on explanatory integration as a remedy for explanatory pluralism. This is a novel strategy insofar as integration is typically discussed by pluralists and is more often than not understood as a strategy which avoids a commitment to any kind of unity of science (e.g. Brigandt, 2013; Cusimano & Sterner, 2019; Faucher, 2014; Mitchell, 2003, 2009). I propose instead that the proliferation of ways of carrying out explanatory integration shows that there is no limit to the types of strategies which can be invented or discovered by researchers. Every one of these new ways of doing integration adds a tool to (what I call) the integrative toolkit, multiplying the ways in which epistemic units of all types (theories, explanations, models, data, results, etc.) can be brought together to produce original explanations which allow us to go further then we had before. I thus both critically assess explanatory pluralist positions to highlight their shortcomings, and subvert the role of integration to show how it in fact is a tool for dissolving explanatory pluralism.

With these tools in hand, I then turn to a defense of tempered integrative monism, which proposes that there is no in-principle limit to explanatory integration. Pluralists see themselves as defending their conceptions against monists, but seldom if ever call on actual, contemporary, monist writings. I thus propose here to remedy this problem by explicitly defending a novel form of explanatory monism which rests on a contemporary understanding of research into the biology of behaviour. My position rests predominantly on the acceptance of scientific realism, which is to say that our best scientific explanations are successful insofar as they tell us something about the world as it actually is. This in and of itself should not be contentious within this debate, since explanatory pluralists as well are explicit about their acceptance of scientific realism. Scientific realism however implies that scientific explanations will come bundled with empirical and ontological commitments about the world as it is, which I argue will always be sufficient to create points of contact between the plurality of explanations for a given phenomenon. These points of contact highlight incompatibilities between explanations, fueling the dynamism of scientific researchers, who will actively try to explore those points of contact. This opens the door to the possibility of explanatory integration, the remedy against explanatory pluralism: because integrative explanations will always, in principle, be possible, then pluralism is only ever temporary. The possibility of explanatory integration is tempered however, but only through pragmatic constraints, such as ethical concerns and limitations of access to phenomena or funding.

What we are left with is tempered integrative monism: a picture of scientific research joined together through explanatory integration. This type of monism accepts the plurality of methods, the plurality of approaches, and recognizes the challenges that implies regarding the coordination of the resulting plurality of explanations. But it also posits that this plurality will only ever be temporary, since our successful scientific explanations will always tell us something about the world as it is, implying that no explanation can ever be independent from all other explanations. This is significant, since it changes our perspective on what successful scientific research is: while the last decades have rightly seen the diversity of approaches and perspectives in a positive light, we must not stop there. This diversity cannot and should not lead to the isolation of approaches, and to the idea that independent or incompatible explanations should be embraced; instead, successful scientific explanations need to be brought into the fold, into our understanding of the world. And we can do so by bringing them into contact with other successful scientific explanations, integrating explanations one with the other, thus painting a truer picture of the world as it actually is, in all its complexity.

The following thesis is broken down into seven chapters. The first four chapters are devoted to understanding what explanatory pluralism is about, and critically evaluating specific explanatory pluralist positions. The two following chapters then build new foundations for tempered integrative monism, which is defended in the last chapter.

Chapter 1 thus begins with a characterization of the different kinds of explanatory pluralism found in the literature.[1] I propose a new typology of explanatory pluralisms based on the target of the purported pluralism: *type pluralism*, which puts to the fore the different types of explanations that can be found in science, *fragmentation pluralism*, which denies any possible overarching unity to science, and *insular pluralism*, which accepts the co-existence of multiple incompatible explanations for any given phenomenon. I then describe some of the possible motivations for a defense of any form of pluralism, as well as some of the epistemological foundations for a defense of explanatory pluralism.

Type pluralism is quite reasonably a default stance for virtually all philosophers of science, and as such, is not discussed at length in this thesis; instead, I concentrate on fragmentation and insular pluralism. The two following chapters cover Sandra Mitchell's take on fragmentation pluralism (2003, 2009), which is broken down into two dimensions. Chapter 2 covers her vertical approach to pluralism, which calls on a

[1] Parts of chapter 1 have been previously published in Muszynski, Eric & Malaterre, Christophe (2020). Behaviour and Biology: An Introduction. In C. Malaterre & E. Muszynski (Eds.), *Biology and Behaviour: Explanatory pluralism across the Life Sciences* [Topical Collection], *Synthese*, Springer, DOI: 10.1007/s11229-020-02856-0.

very popular anti-reductionist, and emergentist position. Chapter 3 describes her horizontal approach to fragmentation pluralism, which suggests that there can exist multiple idealized models which explain a given phenomenon, and that the plurality of models is here to stay. Each of these two chapters takes a critical look at the arguments which Mitchell marshals to show how they fail to adequately defend her pluralism. Her anti-reductionist arguments misconstrue the reductionism she attacks and therefore fail to show why we ought to reject it, and her redefinition of emergence robs it of metaphysical and epistemological import. As for her many-models approach, a clear-eyed understanding of what is at stake shows that it is at best an overstatement of what explanatory pluralism could be, and at worst a trivially true description of how many causes can lead to a single outcome.

Chapter 4 takes a look at Helen Longino's insular pluralism (2002, 2013), which accepts that the different approaches to scientific research into the biology of behaviour could produce a multiplicity of correct yet incompatible explanations for any given behaviour. She proposes that once it is clear that a multiplicity of explanations do in fact target one specific phenomenon, we can come to see how it is possible for those explanations to be incompatible, yet nevertheless correct. This, she argues, is so for two reasons. The first is because of the different ways that an approach will parse the space of possible causes for the phenomenon, which entails that they do not see and measure the world in the same way. The second is because Longino understands the success of scientific explanations to be measurable only from within a given approach, meaning that cross-approach comparisons are—or at least may be—impossible. In other words, approaches can be incommensurable, leading to permanent, irreconcilable explanatory pluralism. Her particular brand of insular pluralism is then critically appraised and ultimately rejected through the remaining three chapters.

Chapter 5 surveys the broad understanding of scientific integration found in the philosophical literature. It is widely accepted that scientific explanations, theories, methods, results, or indeed any kind of epistemic unit used in scientific practice can sometimes be combined with others to yield greater insights into a given phenomenon. The chapter covers many of the different ways this has been conceptualised, proposing that each new integrative strategy adds to the 'integrative toolkit'. This toolkit is constantly expanding, and each new tool that is added is a new way of breaking down the incommensurabilities that Longino takes to be widespread in scientific research. I suggest that there is no end in sight to the ways in which integration can be carried out.

Chapter 6 then proposes a novel and concrete addition to the integrative toolkit, applicable specifically to the domain of biological explanations of behaviour, through the analysis and re-conceptualisation of the term 'behaviour'.[2] The term is ubiquitous in much biological research, but it is seldom defined, and is applied to all manner of phenomena and entities (Levitis et al., 2009). I propose that this apparent disunity in its use dissolves when we define 'behaviour' not as a binary concept, but instead along a gradient. This highlights both the similarities and differences in what biologists mean when—for instance—talking about behaviour both in humans and in plants. My definition thus opens the door to the integration of explanations by facilitating communication across various approaches, who could better understand the use of the concept of 'behaviour' in other contexts.

Finally, chapter 7 is a defense of tempered integrative monism, which puts to the fore that explanatory integration of epistemic units (such as theories, explanations, models, results, etc.) will always be in principle possible. I argue that if we accept scientific realism, or the idea that our best scientific explanations tell us something about the world as it actually is, then explanatory pluralism is untenable. Because successful explanations make commitments about how the world is, either through empirical or ontological commitments, there will always be points of contact between apparently incompatible explanations, and that incompatibility and incommensurability therefore cannot be permanent. These purported incompatibilities will instead open new avenues of inquiry through new questions, methods, models, etc. Integration will always be possible, leading to a form of monism which sees all explanations as eventually joined together through integration. This integration is tempered by pragmatic constraints such as limits imposed by ethical concerns, or simply access to phenomena, but it is not in-principle constrained.

[2] Most of this chapter was previously published as Muszynski, Eric & Malaterre, Christophe (2019), Best Behaviour: A proposal for a non-binary conceptualization of behaviour in biology, in *Studies in the History and Philosophy of Biol & Biomed Sci* vol.79, DOI: 10.1016/j.shpsc.2019.101222.

CHAPTER 1
TYPOLOGY OF PLURALISMS

The last few decades have seen a surge of publications in science and philosophy of science relating to pluralism.[3] Authors have argued that contrary to what was defended in twentieth-century philosophy, science is not a unified endeavour, and instead calls on many theories, methods, taxonomies, ontologies, and explanatory strategies. Philosophy of biology has proven to be particularly ripe for pluralist positions, for various reasons. Some argue that the complexity of the subject matter implies that no single overarching theory will ever be sufficient to explain everything within biology (Mitchell, 2003), others that the historical contingency of natural selection (Beatty, 1993), the diversity of questions being asked (Longino, 2013), or the competing scientific traditions within biology (Morange, 2015), preclude any kind of unifying schemes. These issues seem to be compounded in biological research relating more specifically to behaviour, where the complexity of the phenomena and the diversity of approaches which can contribute to explanations give rise to much interesting research (e.g. Laland & Brown, 2011; Longino, 2013; Mitchell, 2003; Plaisance & Reydon, 2012; Tabery, 2014). This makes the philosophy of behavioural biology a particularly important litmus test for pluralist positions: if pluralism does not hold in this area, then chances are that it may be difficult to defend in other areas.[4]

Scientific pluralism can mean many different things depending on what the plurality is attributed to (Kellert et al., 2006a). Ontological pluralism argues that we ought to rid ourselves of the assumption that the metaphysical foundation of the world is unified. Instead, the fundamental building blocks are recognized as being multiple (Cartwright, 1999; Dupré, 1993; Waters, 2017). This is sometimes reflected in forms of taxonomic pluralism (Dupré, 1993), though an acceptance of diverse taxonomies need not imply ontological commitments (Ereshefsky, 2001; Kitcher, 1984). Methodological pluralism has been the object of debates, for instance as regards to the units of selection (Dawkins, 1976; Lloyd, 1989, 2005; Okasha,

[3] Much of sections 1.1 and 1.2 have been previously published in Muszynski, Eric & Malaterre, Christophe (2020). Behaviour and Biology: An Introduction. In C. Malaterre & E. Muszynski (Eds.), *Biology and Behaviour: Explanatory pluralism across the Life Sciences* [Topical Collection], *Synthese*, Springer.

[4] Pluralism in general, of course, does not stand or fall with pluralism in behavioural biology, since it would remain to be demonstrated that all other scientific research has the relevant characteristics of behavioural biology. Nevertheless, insofar as it tackles general concepts in science such as complexity, emergentism, and others, as well as the fact that pluralists themselves seem to think that it is particularly ripe for defenses of pluralism, it stands to reason that it would be a hard blow for pluralists if pluralism in behavioural biology did not stand up to scrutiny.

2006; Sober, 1990; Sterelny & Kitcher, 1988). These issues dovetail with many other forms of pluralism which are concerned with epistemic issues, such as anti-reductionist positions (Fodor, 1974), as well as interactions within multi-disciplinary research domains (Kellert, 2008; Longino, 2002; Repko, 2012). Many of these positions also relate to explanatory pluralism, which highlights the variety of explanations that science produces (Braillard & Malaterre, 2015; Kellert et al., 2006a; Kendler, 2012; Longino, 2013; Mitchell, 2002, 2009; Ruphy, 2013).

Explanatory pluralism has recently become the focus of much research in the philosophy of biology and behaviour. Yet despite this increased attention, it is not always entirely clear what explanatory pluralism involves, leading to some apparent confusion in the implications of such positions. For instance Gijsbers (2016) points out that whereas Campaner (2014) speaks of explanatory pluralism as emphasizing the isolation and incommensurability of different explanatory schemes, Abney and colleagues (2014) describe it as the capacity for different approaches to offer complementary and ultimately integrable explanations. Of course, both these positions could appropriately be called pluralist by different authors, but at the very least it is clear that the same term, 'explanatory pluralism', can come to mean seemingly opposite things.

This chapter offers a roadmap which lays out the different forms of explanatory pluralism, covering three different targets that have been identified which relate to the plurality of explanations (section 1.1). I then cover the possible motivations for such positions (section 1.2), looking specifically at the literature in philosophy of biology and behaviour, as this seems to be where many of these ideas are burgeoning. I finish with a look at some of the foundations typically used to defend explanatory pluralism (1.3), which will serve to better understand the pluralisms tackled in chapters 2, 3 and 4.

1.1 Targets of explanatory pluralisms

Those defending explanatory pluralism agree that notions regarding the unity—or eventual unification—of explanations in science are at best misguided, or worse, simply wrong. But aside from this general agreement on who the adversaries ought to be, explanatory pluralisms can come with a wide variety of commitments and consequences, each with their own specific monist adversaries.

1.1.1 Existing typologies of pluralism

To make sense of this multiplicity, many different typologies of pluralisms have been proposed. For example, Longino (2013, p. 147) distinguishes between eliminable and ineliminable pluralism. While the former acknowledges a current plurality of scientific explanations, it is understood to be temporary matter, with unification as the ultimate—and realistic—goal. Ineliminable pluralism, on the other hand, proposes that the plurality is here to stay. Mitchell (2003, pp. 186–192), for her part, outlines three different kinds of pluralisms, starting with 'anything goes pluralism,' represented by approaches such as Feyerabend's epistemological anarchism (1975, 1981). The second is 'isolationist pluralism,' which understands explanations at a given level of analysis to be impervious to explanations at other levels, as some have interpreted Mayr (1961) and Tinbergen (1963) as proposing. And third, 'integrative pluralism,' defended by Mitchell herself, which advances that certain explanations—but not all—can be integrated one with the other in various ways. Van Bouwel (2014) proposes a fivefold distinction, recognizing Longino's 'eliminable pluralism' (rebranded as 'moderate pluralism'), adopting Mitchell's tripartite nomenclature, and adding another type of pluralism to this list, calling it 'interactive pluralism.' This last kind is described as a middle-of-the-road position between isolationist and integrative pluralism, which recognizes the value of interactions between explanations (or approaches) without making integration an imperative.

Though these various ways of cataloguing pluralisms have the merit of showing the nuance across positions, it can sometimes be difficult to understand precisely what it is about explanations which is plural. What does it mean for explanations to be isolated, interactive, or integrated? Are pluralists highlighting the diversity of forms, or the diversity of content within biological explanations? To shed light on these issues, I propose here to classify explanatory pluralist positions with respect to the target of pluralist claims. In other words, what it is about the explanations which is understood to be plural.

1.1.2 Type pluralism

The first kind of pluralism I call '*type pluralism*' applies to the types of explanations which are found in science. Defenders of this pluralism argue that scientific explanations do not—and need not—partake of a single explanatory model. Instead, sometimes even within a single discipline, different types of explanations are possible, such as explanations that appeal to covering laws, mechanistic explanations, statistical relevance explanations, causal explanations, or others. The opposite view would be a monism about the types of scientific explanations. Probably the most well-known example in philosophy of science is Hempel and Oppenheim's Deductive-Nomological account of explanation (Hempel & Oppenheim, 1948), which put forward that all scientific explanations, including explanations in social sciences, take the form of a deductive argument (called the *explanans*), where at least one of the premises is a law of nature. The deduction leads to the explanation of a particular event or phenomenon (also called *explanandum*) and shows why it follows nomologically from the laws of nature. Other examples are van Fraassen's (1980) contextual account of explanation, Salmon's (1984) causal mechanical model, or Kitcher's (1981) unificationist account. All these positions have in common that they propose that scientific explanations all adhere to a single, recognizable type of explanation, differentiated from non-scientific explanations. Type pluralists, on the other hand, insist that there is not one single kind of explanation in science, but that there are many. Put succinctly, this kind of pluralism can be defined in the following way:

> Type pluralism: There exist many types of explanations in science.

It is often seen as a matter of fact that there currently exists a plurality of types of explanations in science (see for instance the articles in Kellert et al., 2006b; Mitchell, 2002). Type pluralism for scientific explanations therefore seems to be the default stance for most contemporary philosophers of science, and as such is often left undefended, at least in any explicit form (though some are interested in fine-grained distinctions between types; see e.g. Aizawa & Gillett, 2019; Issad & Malaterre, 2015). Note that as Plutynski (2016) points out, even in this narrow application of type pluralism, many positions can be espoused, such as defending a diversity of representations, a diversity of questions, or a diversity of "modes" of explanation. These various positions are often advanced without explicit recognition of their differences. In all these cases, it remains that the target of this kind of pluralism is the type of explanations that are possible in science.

1.1.3 Fragmentation pluralism

The second kind of explanatory pluralism is '*fragmentation pluralism,*' and defends the idea that explanations will never merge into a grand, unified explanation or theory of everything. This pluralism is pitted against a monism which relies on ways to unify disparate explanations one with the other, most commonly through reduction. Reduction itself is a multifaceted concept that can be generally understood as the idea that theories or explanations can be reduced to more fundamental theories or explanations. Nagel (1961), for instance, proposed that intertheoretic reduction can be done through logical relations between various elements in two different theories, sometimes through the use of bridge principles, with the goal of showing how one theory (typically of a higher level) can be explained through, or logically derived from, another theory (typically of a lower level of organisation). Reduction of this sort is sometimes understood to apply across the board, and would lead to a unified explanation of everything through the operation of our most fundamental theories, presumably physics. Explanatory reduction is similar to theory reduction, but allows reduction to apply more locally, such as to parts of theories, mechanisms or explanations of individual phenomena, as opposed to entire theories (Brigandt & Love, 2017). In biology, this type of reduction often relies on a mechanistic view of explanation, whereby a phenomenon is explained insofar as the mechanism causing it is described through the decomposition of its constituent parts (Machamer et al., 2000), an approach which dovetails with reductionist intuitions. Other accounts of unification need not rely on reductionism, such as Kitcher's (1981) unificationist account of explanation, which proposes that explanations are successful insofar as they unify explanations of different phenomena under one explanatory schema. Whether it be through reduction or some other strategy, the monist intuition is that all of science could eventually be unified under one grand explanation of everything (though what exactly this means and how it could be done is a matter of debate).

Sandra Mitchell is currently a well-known philosopher of science to defend fragmentation pluralism. In her view, it is sometimes possible to combine or coordinate explanations for a particular phenomenon through specific integrative strategies, but not in every case. For instance, some explanations will resist reduction from one level to another (Mitchell, 2003, p.186; 2009, chap. 2), and others, even within a single level, will not converge because they "only describe what would happen in non-overlapping ideal worlds" (Mitchell, 2003, p. 64). Opportunities for integration arise in particular, concrete, non-idealized cases, where a specific phenomenon can be explained using diverse approaches. As Mitchell states: "However complex, and however many contributing causes participated, there is only one causal history that, in fact, has generated a phenomenon to be explained" (Mitchell, 2002, p. 65). That single particular causal history is understood to be amenable to only one complete explanation, which may call on the integration of various models. However, though there may be opportunities for integration in particular cases and therefore little chance for pluralism at that level, the abstracted, generalized, broad, unified explanations will forever be out of reach (Mitchell & Dietrich, 2006). Thus, what is plural for this kind of pluralism is the general, abstract, or theoretical level of explanations, the contention being that no explanation will ever be large enough to contain all the particular phenomena to be explained by science; as such, science will remain fragmented.

The general definition of fragmentation pluralism can be phrased as follows:

> Fragmentation pluralism: Scientific explanations can be integrated for specific concrete phenomena, but the unification of science under the aegis of a single theory or explanation is impossible.

1.1.4 Insular pluralism

The third and final form of pluralism is '*insular pluralism*' and is in many respects the most forceful application of explanatory pluralist ideas. This is the view that two (or more) different scientific approaches may explain a given phenomenon in ways which are not only different, but impossible to integrate or even compare. This is furthermore not seen as a problem in need of resolution but instead as the normal result of science running its course (Kellert et al., 2006b). "Scientific approaches" is a broad term which "includes at a minimum, characteristic questions, characteristic methods for addressing those questions, and a commitment to the importance of the questions and the answers generated by the available methods." (Longino 2013, p.15; see also Kellert et al. 2006b). The explanations put forward by different approaches can be seen as so many islands, forever isolated from other explanations. In a nutshell, the kind of pluralism is defined as:

> Insular pluralism: multiple scientific explanations can target a single phenomenon and be permanently incompatible, and this is a positive aspect of science.

Longino (2013) defends this form of pluralism in behavioural biology, arguing that different approaches will parse the causal space implicated in a given behaviour in different ways, sometimes leading to incompatible explanations. Furthermore, the resulting explanations may not be in a position to invalidate the others (see also Waters, 2005). Longino argues that this is the case because different approaches will have differing methods, scopes, and assumptions, which will lead them to parse the space of possible causes in different and irreconcilable ways. She furthermore proposes that the success of a scientific explanation is not evaluated in terms of truth or falsity, but rather "conformation" (a notion introduced in Longino, 2002). Conformation is a non-binary way of evaluating the acceptability of a scientific explanation, which highlights that an explanation can conform, sometimes more, sometimes less, to the phenomenon to be explained, without needing to state that it is simply true or not true (Longino, 2013, pp. 147–148). By removing the binary of truth and falsity, explanations are no longer understood to be the whole story about a given phenomenon, but only a more or less successful representation thereof, making room for a plurality or partial representations. Moreover, because the measure of conformation itself is done only within the context of each approach, cross-approach comparisons of levels of conformation are, at least sometimes, impossible. Explanations can therefore be incommensurable, since the different approaches have no common measure, and consequently no way of evaluating or comparing their differing explanations, isolating each approach within their own methods, scopes, and assumptions. This precludes the integration of explanations for a given phenomenon, meaning that there can exist a plurality of explanations which is here to stay. In sum, insular pluralism targets explanations of particular phenomena, positing that there can exist a plurality of explanations which, even if they are incompatible, is not problematic, and will not be reconcilable.

The opposing monism would stress that despite the fact that multiple approaches have different ways of examining and explaining a given phenomenon, incompatibility and incommensurability are ultimately impossible. This is the type of monism I defend in chapter 7.

1.1.5 What is captured by this typology

This typology of explanatory pluralisms, distinguishing type pluralism, fragmentation pluralism and insular pluralism, captures divisions which were not covered by those previously proposed. Consider Longino's (2013) two-way distinction between eliminable and ineliminable pluralisms: each of the kinds of pluralism we here laid out can be of either sort. For instance, one could either defend that the variety of types of explanations could one day be subsumed under a single type, hence be eliminable, or that type pluralism is permanent, hence ineliminable. Insular pluralism seems constrained to be of the ineliminable sort, though one could imagine a future state of science where extensive interdisciplinary work could break down at least some of the incommensurability between approaches. Nevertheless, one may still argue that this would never eliminate all explanatory pluralism. As for fragmentation pluralism, it explicitly endorses ineliminable pluralism at the more abstract level but endorses some form of eliminable pluralism more locally. Considering the distinctions proposed by Mitchell and extended by Van Bouwel—anything goes pluralism, isolationist pluralism, integrative pluralism, as well as moderate pluralism and interactive pluralism—some could be interpreted as applying to explanations at a broad, theoretical scale (hence a form of fragmentation pluralism), or explanations applied to specific, concrete phenomena (and in that case related to insular pluralism). It is also interesting to note that the concept of levels of analysis present in Mitchell's 'isolationist pluralism' could be put forward to argue for forms of fragmentation pluralism or forms of insular pluralism. Finally, none of the previously proposed pluralisms easily apply to types of explanations, reinforcing the idea that type pluralism is the default stance for philosophers of science.

1.1.6 Explanatory pluralism is not methodological pluralism

Explanatory pluralism and methodological pluralism are often conflated, and as such it is important to emphasize their differences and the possible links between the two. While explanatory pluralist positions do tend to favour methodological pluralism, monists as well can favour a diversity of methodologies.

Type pluralism is the only explanatory pluralism that seems to necessarily go hand in hand with methodological pluralism. This is because it is often thought that it is precisely the diversity of methodologies which leads to the diversity of types of explanations; in contrast, it is not clear how a single methodology could yield a variety of types of explanations. For their part, fragmentation pluralism and insular pluralism naturally lead to methodological pluralism since they emphasize differing approaches to a given phenomenon, but it is not a necessary entailment. Just as is the case with type pluralism, the different methodologies can be seen as the source of incompatibilities and incommensurabilities among explanations. As will be shown in chapter 4, this most notably the case with arguments pertaining to insular pluralism: different approaches use different methods, limiting what they can explain about a given phenomenon, and more importantly, constraining what they can say about the success or failure of rival methodologies. However, as will be seen in chapters 2 and 3, Mitchell's brand of fragmentation pluralism relies not on the variety of methodologies, but instead on anti-reductionism and levels of abstraction. As such, at least for some brands of explanatory pluralism, methodological pluralism is not necessary.

While the link between explanatory and methodological pluralism is strong, what is important to recognize is that methodological pluralism is compatible with different forms of explanatory monism (namely those mentioned as antithetical to each kind of pluralism laid out above). This is because the different methodologies can be understood either as a temporary step in the search for the "right" methodology, or because the different methodologies are seen as contributing to the goal of a unified science. For instance, Hempel and Oppenheim's (1948) contention was that regardless of the methods used, all scientific explanations would eventually take the form of a deductive-nomological derivation, meaning that they could presumably accept methodological pluralism all the while denying type pluralism. The monisms opposed to fragmentation and insular pluralism, for their part, could accept that multiple methodologies each contribute valuable explanations, which will eventually be unified, for instance through reduction, unification, or integration. In other words—and this is the important point—monism can be favourable to methodological pluralism all the while defending some form of unity of explanations in science, meaning that there is an important difference between explanatory pluralism and methodological pluralism.

1.2 Motivations for pluralism

Many different arguments are marshalled in favour of pluralist positions in philosophy of science and can be classified into three broad categories: those which are motivated by pragmatic considerations, epistemic considerations, and finally ontological considerations.

1.2.1 Pragmatic motivations

The pragmatic motivations generally concern resource allocation for research and the promotion of a diversity of perspectives (e.g. Dupré, 2002; Kitcher, 1990; Mitchell, 2009). Pluralism is here seen as a way to hedge bets against monist stances which are seen as encouraging only one, or a select few, disciplines or research agendas considered to be the 'right' way of going about making scientific discoveries (Kellert et al., 2006b, p. xxi). Instead, it is argued, we ought to encourage a wide variety of approaches, and ideally finance many different research groups, each with their own specific goals or approaches, in order to maximize the number of perspectives as relates to a given research question. As just mentioned however, because explanatory pluralism and methodological pluralism are not identical, the charge laid against monists of privileging a single approach are—at least in certain cases—exaggerated. Nevertheless, it is certainly the case that by default, pluralist positions will see the diversity of approaches in a positive light, meaning that for many, these pragmatic considerations lead to pluralist positions.

1.2.2 Epistemic motivations

The pragmatic motivations obviously dovetail with some of the epistemic considerations in favour of explanatory pluralism. This is the case, for instance, with new trends in epistemology and philosophy of science, as seen in the 'values in science' debate (e.g., Elliott & Steel, 2017), which emphasize the fact that science is practiced within a social context. This implies that both the production and the evaluation of scientific knowledge are inextricably linked to values other than those traditionally attributed to objective knowledge about the world (Douglas, 2007, 2009; Dupré, 2007; Longino, 1990, 2002). These arguments are frequently informed by feminist approaches which are critical of mainstream science (Crasnow, 2013; Wylie, 2003), and often highlight the potentially damaging aspects of the pursuit of 'objective truth', such as discrimination or the silencing of minorities. In sum, since the pursuit of traditional 'objective truth' is regarded as an impossible and sometimes harmful objective, it is best to recognize that contributions to scientific knowledge can come from many different sources, opening the door to different forms of pluralism.

Another epistemic motivation for pluralism comes from the exploration of the limitations of any given scientific explanation. By virtue of being representations, explanations of any given phenomenon will by definition be partial, implying that no single explanation will suffice for a complete explanation (Longino, 2013, p. 147; Mitchell, 2009 chap. 2). Giere (2006) compares this situation with that of colour vision: visual systems can differ in humans and other species, with no way of determining which is the 'correct' way of perceiving colour—though some can be richer than others insofar as they allow for the perception of distinctions that others may not. In much the same way, scientific explanations only ever give us a part of the complete explanation. Horst's (2016) cognitive pluralism extends this approach by showing the various ways cognitive mechanisms shape our parsing of the world. Explanatory pluralism can thus be defended on epistemic grounds without appeal to social factors, by instead emphasizing the limitations inherent to our ways of apprehending the world, and the resulting representations of that understanding through scientific explanations.

1.2.3 Ontological motivations

Finally, some will tie in motivations relating to ontological pluralism. This is the case for instance with Dupré's (1993) "promiscuous realism" which advances that the multiple, even conflicting, classificatory practices used to further human interests are all real and referential. This multiplicity is in turn reflected in the resulting scientific explanations, which entails that the metaphysical foundations of the world are similarly disunified. Nancy Cartwright (1999) has also defended a metaphysical view of a "dappled world", which is composed of different realms, restricting the scope of explanations, thus precluding unification. In a similar vein, Waters (2017) argues that current biological practice reflects the fact that there is "no general structure" in the world's underlying framework; instead, its "messiness" is reflected in the piecemeal descriptions and explanations put forth by (for instance) geneticists. Ontological pluralism can thus be seen as the reason for different kinds of explanatory pluralisms and is therefore often revealed through the explanatory practices of practicing scientists.

Of course, within these three broad categories of motivations for explanatory pluralism are countless arguments and strategies. Some will begin with specific case studies, others with broad considerations regarding science in general, others still with a look at science in practice. There is thus a wide range of motivations, arguments and nuances which can be brought forth in favour of explanatory pluralisms.

1.3 Pluralist foundations: complexity

Before taking a deeper dive into Mitchell's fragmentation pluralism (chapters 2 & 3) and Longino's insular pluralism (chapter 4) as relates to biology, I will here lay out some of the foundations used in many pluralist arguments. Though it is important to stress that Mitchell and Longino do not defend the same kind of pluralism, and that their arguments are therefore different—sometimes slightly, and sometimes drastically—there is still a common thread to some of the premises of the arguments, namely in the emphasis on the complexity of the subject matter, as well as the fact that scientific explanations are best understood as representations.

The first of two major foundations for those forms of pluralism relates to the complexity of the biological world. The claim is that many of the explanandum phenomena of the life sciences are such that they preclude traditional, reductive approaches to scientific investigation. Thus, while physics, for instance, may be amenable to research which attempts to explain a phenomenon by taking it apart and looking at its constituent, simpler parts, "complex behaviors in biological and social sciences seem not to yield as well to a reductive approach" (Mitchell, 2009, p. 2). In a similar vein, Longino proposes that "the complexity of natural entities and processes (either all such or just organic entities and processes) eludes complete representation by any single theoretical or investigative approach." (Longino, 2002, p.93; see also Kellert et al., 2006). This complexity leads to a "deep" uncertainty (Mitchell, 2009, p.3): in certain cases, biological research will be unable to completely and adequately explain a phenomenon, leading to an ineliminable plurality of explanations.

Mitchell takes the time to attempt to characterize what complexity is, detailing the different roles it plays in pluralism. She recognizes that "complexity" is a challenge to define adequately. Rather than attempt to address all of the 30 to 45 available definitions (Mitchell attributes this census to Horgan, 1997), she focuses on the three kinds of complexity which are most prominent in biology: constitutive complexity, dynamic complexity, and evolved complexity. Though Longino does not adopt this nomenclature, it will allow us to better understand the kinds of complexity which are at play even in her defense of insular pluralism.

1.3.1 Constitutive complexity

The first of these kinds of complexities refers to the fact that organisms or other biological phenomena are composed of multiple interacting parts at various levels of organisation. Alternately called "constitutive complexity," "compositional complexity," or more broadly "multilevel organisation," this kind of complexity refers to the fact that biological organisms and phenomena typically involve numerous interacting parts. As Mitchell puts it, these "complex systems" have "multiple parts that stand in non-simple relations" (Mitchell, 2003, p. 5). This applies to organisms, which contain many organs, cell types, fluids, etc., as well as broader systems such as ecosystems, which contain many interacting biotic and abiotic elements. The relations among the parts are considered non-simple insofar as there are intricate pathways of cause and effect, including nonlinear interaction, feedback loops, and the integration of a multiplicity of causes, as opposed to simpler relations such as additivity, aggregativity, or straightforward cause-to-effect relations.

Constitutive complexity impacts the way phenomena are explained because of the difficulties involved in accounting both for the causal pathways at a given level, and the appropriate links between the levels. Though Mitchell never defines the notion of "level," she does name them, for instance when describing major depressive disorder: "The behaviour is associated with multiple levels of organization, from gene, to cell, to region of the brain, to hormonal systems, to affect and behavior" (2009, p. 10). Without going into detail, she punctuates this sentence by calling on Craver's (2007) notion of levels. In his chapter dedicated to levels, Craver remarks that the term is widespread, and "multiply ambiguous" (p.163). Indeed, it seems the only necessary condition for talk of levels is for items to be ordered hierarchically, typically such that lower levels are nested in higher levels.

Craver proposes that the notion is best understood through mechanistic explanations. These types of explanations explain in virtue of uncovering the underlying mechanisms which reliably produce the phenomenon that is in need of explanation (Bechtel & Richardson, 2010; Craver, 2007; Glennan, 2002; Machamer et al., 2000). They point to a number of specific entities—organized in a particular way and which carry out a number of specific activities—which, when affected by certain input conditions, are susceptible to produce a certain phenomenon. These mechanisms are nested hierarchically one within the other, as is the case for instance with neurons inducing long-term potentiation, which are components of the hippocampus, which in turn constitutes the cerebral mechanism which accounts for a mouse navigating a maze (Craver, 2007, p. 166). Craver's notion of levels is dependent on this hierarchy of mechanistic explanations: “Lower levels in this hierarchy are the components in mechanisms for the phenomena at higher levels. Components at lower levels are organized to make up the behaviors at higher levels” (2007, p. 170). A level is therefore understood as those constitutive entities needed to explain a phenomenon, regardless of other characteristics sometimes attributed to levels, such as size (P. S. Churchland & Sejnowski, 2000; Wimsatt, 1976a), causation (Campbell, 1974), or being governed by the same laws (Gould, 1980).

Mitchell’s claim regarding constitutive complexity can therefore be summarized as follows:

> CC (Constitutive Complexity): Biological phenomena are considered constitutively complex when they are explained through mechanisms spanning multiple levels, which stand in non-simple relations to one another, both within a level and between levels.

These phenomena will therefore call on explanations at various levels which, according to Mitchell, could be difficult to relate one to the other due to a lack of tractability of the causal pathways. Among other uses, the fact that CC is found in certain phenomena is used as a basis for arguing that reductionism will fail in certain cases (see chapter 2).

1.3.2 Dynamic complexity

The second kind of complexity Mitchell attributes to biological systems is dynamic complexity. This refers to the general idea that the behaviour of complex systems will at times be surprising, and sometimes unpredictable, even when the initial conditions and the rules governing the behaviour of the components are known. The most well-known example of this is chaotic systems, where the output is always unpredictable. What Mitchell has in mind however are biological phenomena whose behaviour cannot be modeled through linear equations, or as the sum of multiple independent causes (2009, pp. 34–35). Phenomena which Mitchell classifies under this heading are “self-organising and recursive patterning (e.g. thermal convection patterns), and negative and positive feedback regimes (amplification and damping)” (2003, p. 6). Many of these properties will interact with one another, as in the case Mitchell presents: the division of labour in social insects.

This can be summarized as follows:

> DC (Dynamic Complexity): Biological phenomena are considered dynamically complex when they can only be modeled through nonlinear equations.

Page and Mitchell (1998) propose models to account for the division of labour in honey bee hives, which rely on self-organisation and feedback regimens. Despite the fact that none of the individual bees contain a blueprint or plan for the general order found in the hive, the bees will divide labour amongst themselves, some favouring cleaning, others foraging, others reproducing, and others still guarding the entrance to the hive. Page and Mitchell elaborate models where individual insects will behave differently depending on the stimuli coming from the environment, and where the behaviour of individual insects will change the environment, leading to a change in stimulus for themselves and for the other insects in the vicinity. This means that there is a feedback loop between the behaviour of individual insects and their environment, since both are liable to change the other. Page and Mitchell show that by assigning random thresholds for responses to stimuli to the individual insects, the insect colony will spontaneously create a seemingly ordered division of labour through their individual choices. Their research supports the view that the division of labour is a result not only of phylogeny, but also of ontogeny: though natural selection could have played a role in some features of social insects, “at least some aspects of those features might well be the result of complex dynamics of the development of insect colonies” (Mitchell, 2003, p. 39). In this respect, for Mitchell, a beehive is dynamically complex, since the complex structure will be dependent on the initial conditions (i.e. the behaviour thresholds and the initial environmental stimuli), and the resulting feedback loops and self-organisation of the individual members.

As will be shown in chapter 2, dynamic complexity is associated with unpredictability, which Mitchell will draw on in order to argue that modern science provides examples of emergent phenomena.

1.3.3 Evolved complexity

Finally, evolved complexity relates to the diversity of biological organisms, and the contingent nature of their evolutionary history. The traits of any given species, or any given organism, could in principle be the result of many different evolutionary histories, and any given environment can lead to differing adaptations (Mitchell, 2009, p. 45). This implies that the organisms which currently exist could very well *not* have existed, had evolutionary history played out even slightly differently. This type of complexity is somewhat different from the other two insofar as it relates to the diversity of entities which interest biologists, rather than the inherent complexity of those entities. It points out the sheer multitude of “ways to ‘solve’ the problems of survival and reproduction” (Mitchell, 2003, p. 7), and how those adaptations have downstream effects on the possible future adaptations, which combine to give us the multitude of organisms and entities in the life sciences.

This premise could be succinctly formulated as:

> EC (Evolved Complexity): Biological phenomena are considered complex because they present a great variety of phenomena which are the result of the contingent history of evolution.

This implies that the generalisations that can be made about biological phenomena do not have the same status as those from, say, fundamental physics, since they are applied to phenomena which could have been otherwise. Mitchell ties this to John Beatty's "evolutionary contingency thesis" (Beatty, 1993) to argue that within biology, "the requirements for lawfulness fail to reflect the reality of scientific practice" (Mitchell, 2009, p. 53). In other words, because the multitude of species and biological systems could have turned out another way, the generalisations which are possible in biology cannot be considered universal and immutable laws of nature in the same way that are the laws of physics. Evolved Complexity therefore implies a complexity not so much in the explanations that will be proffered (at least not necessarily), but in the diversity of phenomena which the explanations will attempt to cover, which could translate to a diversity of explanations under certain circumstances.

1.3.4 Complexity, predictability, and explanations

Though Mitchell uses these kinds of complexity to highlight difficulties in attempting to unify the biological sciences, she nevertheless maintains that the complications that arise do not entail that biological phenomena are unintelligible. Indeed, she remarks that the complexities she has identified do not result in chaos, or a "blooming, buzzing confusion" (James, 1890). Instead, they are to be understood as "tractable, understandable, evolved, and dynamic" (Mitchell, 2009, p. 11). Tying together issues related to complexity, determinism, predictability, and the possibility of explanations, her claim is that understanding is within reach, though we should expect the resulting explanations to reflect the complexities of the world they attempt to capture.

Mitchell proposes that the complexity of biological phenomena may entail unpredictability in some cases, but it does not entail unexplainability. Indeed, though DC and CC can imply (by definition) unpredictability, it need not imply that we are unable to make sense of the phenomena. Mitchell puts forward a few illustrations of these phenomena which will be covered in detail in the following chapters, but for now, we can focus on an example which elegantly captures the difference between predictability and explainability: research into the flight patterns of *Drosophila* has shown that the behaviour of any individual fly at any given time is unpredictable, but nevertheless explainable. Maye and colleagues (2007) have proposed that the flight patterns do not reflect random noise, but instead a “fractal order,” which they argue is the result of a nonlinear system. This system is furthermore adaptive, since it allows the fruit flies to tread a fine line between predictable and random behaviour, making them more difficult for predators to catch. We can see through this example that though the phenomenon is unpredictable, it is nevertheless tractable and understandable, and indeed, Maye et al. (2007) explain how the neurological systems work to create the observed behaviour. We can see that the unpredictability of the phenomenon stems not from an epistemological problem, nor from any spooky business at the ontological level; instead, it is the result of the very nature of such a complex system which precludes deducing the later states of the system from the initial state, or the higher levels from the lower ones.

1.4 Pluralist foundations: explanations as representations

The second pillar of the pluralist foundations is the emphasis on scientific explanations as representations. Mitchell and Longino both argue that pluralism is a result of the fact that our scientific explanations of the world are representations, as opposed to perfectly accurate mappings of the world, or propositions which can only be true or false. This implies that even if the world itself were amenable to some sort of unity of order, that unity would not translate to a unity of explanations. This epistemic notion of explanation also contrasts with the ontic conception of scientific explanations, which proposes that it is the things in the world which in fact explain the phenomenon, and not mere representations thereof (Craver, 2014).

1.4.1 Representations of reality

Pluralists are generally explicitly physicalists, meaning that they take for granted that the world is all and only physical or material (Cartwright, 1999; Dupré, 1993; Giere, 2006; Longino, 2013; Mitchell, 2009). They are also typically scientific realists, meaning that they both deny the possibility that there are multiple worlds, as well as taking for granted that scientific explanations of our world attempt to map onto what is actually happening, to “capture the relations and causal structures of one world” (Mitchell, 2009, p. 23, see also p. 13). When applied to the targets of explanations, this implies that “however complex, and however many contributing causes participated, there is only one causal history that, in fact, has generated a phenomenon to be explained” (Mitchell, 2002, p. 65). Scientific explanations will be approximations of that reality, validated by “a combination of measures of predictive use, consistency, robustness, and relevance” (Mitchell, 2009, p. 14). Though some of these standards also call on pragmatic aims (Longino, 2013, p. 149), they nevertheless call on a (more or less) strong link with the actual causal history of a phenomenon (Longino, 2002, p. 116), allowing for adequate predictions, and replicability of experiments.

In sum, as physicalists, pluralists will typically grant that there is only one world to be explained, that any given phenomenon is the result of a single set of causes, and—as realists—that this causal history is what scientific explanations (at least generally) attempt to uncover.

And yet—as pluralists point out—there can be multiple explanations for a single phenomenon; how can that be if there is only a single causal history? As Fodor remarks, "as of now, the hardest part [of philosophy] is to reconcile a physicalist ontology with the apparent ineliminable multiplicity of discourses that we require when we try to say how things are" (1998). There is thus a tension between the unity of the world itself, and the apparent pluralism of the explanations used to describe it. For Mitchell and Longino, this tension is resolved (at least in part) by pointing out that explanations do not map directly onto the world itself, but instead are mere representations of that world, and as such offer a partial account of the phenomenon to be explained.

1.4.2 Partiality of representations

The partiality of representations is due to fundamental assumptions about our relation to the world, as well as notions regarding the nature of explanations. Fundamentally, none of our representations can have an unmediated relationship to the world itself: "To think that our language (or any human artifact intended for representation, including mathematics and simulation) captures the material world exactly is something that most post-Kantian philosophers have rejected as simply misconceived" (Mitchell, 2009, p. 33). In virtue of being a *re*-presentation, these languages of representation are not the object itself, not the referent itself, implying that there is a certain disconnect between the representation and the object it is meant to represent. More specifically, when it comes to scientific explanations, any explanation qua representation will deliberately omit certain details of the explanandum phenomenon.

This partiality of scientific explanations is both a weakness and a strength. It is a weakness since it implies that any given explanation will not, and in fact cannot, capture all of the contributing causes to any particular, concrete phenomenon. As previously mentioned, according to Mitchell, the objective of any explanation is to describe the causal history that led to the phenomenon under investigation. However, for an explanation "to be useful, it cannot include every feature in all the glorious detail of the original, or it is just another full-blown instance of the item it represents" (Mitchell, 2009, p.31; see also Longino, 2002, p.116). This deliberate omission of details is understood to render the explanations "usable", which is to say that they are simple enough to be understandable, and tractable when applied in various contexts

(Mitchell, 2009, pp. 13–14, 33). Mitchell (2009, pp. 116–117) evokes the problem using Jorge Luis Borges' short story "On exactitude in Science," (Borges, 1975) where an overenthusiastic cartographer's guild creates a perfect map of their Empire, but by virtue of it representing every detail of the Empire, it is the same size as the Empire, and therefore completely useless as a map (though see chapter 7 to see why this is an erroneous way of understanding the relation between representations and reality). Therefore an adequate scientific explanation will represent only the causes and facets of a phenomenon which correspond to our interests and abilities in that context (Mitchell, 2009, p. 13).

On the other hand, the partiality of representations is a strength because it is what allows for different causes to be tractable, or useful. By eliminating the details which are not relevant to our interests, the resulting explanations are able to be grasped by the limited beings that we are. As will become clearer in chapter 3, this also allows for generalizations and abstractions of phenomena, rendering explanations applicable to a broader range of sometimes seemingly unconnected phenomena. Explanations thus rely on representations, in which the causal history of the explanandum phenomenon will be incompletely described, for better or for worse.

As will be shown in the next two chapters, the complexities of biological phenomena and the nature of explanations as representations are used by Mitchell as foundations for two somewhat independent defences of fragmentation pluralism. The first is a vertical approach, which argues against reduction and for emergence, and the second is a horizontal approach, which relies on the multiple explanations used to account for a given general phenomenon. These two lines of argumentation are tackled one after the other in the next two chapters. Chapter 4 then takes on Longino's insular pluralism.

CHAPTER 2

VERTICAL APPROACH TO FRAGMENTATION PLURALISM: ANTI-REDUCTIONISM

Fragmentation pluralism is the idea that there are respects in which scientific explanations preclude unification. Contrary to insular pluralism, proponents of fragmentation pluralism do not tolerate the existence of multiple incompatible explanations for a given phenomenon, but they do argue that science will come in fragments, with certain explanations being forever separate from others, with no possibility of unifying them. As can be expected, this broad position can come in many different versions, and be defended in myriad ways.

One of the most well-known contemporary defenses of fragmentation pluralism is Sandra Mitchell's "integrative pluralism", elaborated most notably in her *Biological Complexity and Integrative Pluralism* (2003), and revisited in *Unsimple Truths: Science, Complexity and Policy* (2009). In these books and a selection of articles, she proposes that the biological sciences are such that the unification of explanations into a single unified theory is impossible, though integration of explanations at the local level is sometimes possible, and even valuable. She therefore advocates for "expanded understandings of both the world and our representations of it as a rich, variegated, interdependent fabric of many levels and kinds of explanations that are integrated into one another to ground effective prediction and action" (2009, p. 19). She furthermore believes that these lessons learned through the practice of biology "are applicable in all sciences of the complex" (2009, p. 12).

Though forms of fragmentation pluralism are held explicitly (Cartwright, 1999; Dupré, 1993; Faucher, 2012; Gijsbers, 2016) or tacitly (the following could be interpreted as defending forms of fragmentation pluralism: Craver, 2007; Hochstein, 2017; Lloyd et al., 2005; see also Galison & Stump, 1996) by apparently many philosophers of science, Mitchell's account is particularly interesting with respect to the present research because of the ways it explicitly addresses and relies on contemporary biological research into behaviour. Because of this, it appears to tap into many of the anti-reductionist and pluralist intuitions of scientists themselves. She also considers her approach as a reasonable, middle-of-the-road position between the extremes of scientific unity and disunity (2003, p. 186), making it *prima facia* appealing.

Mitchell's arguments for her explanatory pluralism are based on two general foundations, which have been covered in chapter 1. The first is the various kinds of complexity found in the phenomena of interest in the biological sciences, which she proposes entail consequences for the kind of knowledge that can be created through their study. Though Mitchell grants that explanations are still within reach, she maintains that the varieties of complexity presented earlier preclude the unification of explanations as put forward in nineteenth century philosophy, which often looked to Newtonian physics as the epitome of science. Mitchell describes the writings of Herschel (1830), Whewell (1840) and John Stuart Mill (1843) as exhorting scientists to emulate Newton's discovery of simple laws governing apparently complex phenomena. Simplicity, unity, and universalism were the hallmarks of any good scientific theory. Mitchell, on the other hand, argues that the various forms of complexity in biology lead to multiple explanations, accounting for multiple aspects of phenomena, which will not lead to a single, simple, unified and universal explanation of biological phenomena. In sum, Mitchell's claim is that faced with these complexities, normal scientific advancement will result in an increase in the quantity and diversity of explanations, rather than convergence on to a single explanation. In other words, explanatory pluralism is to be expected, as it is a reflection of the complexity of the subject matter (Mitchell, 2003, p. 3). As she succinctly summarizes: "Life is not simple, and our representations of life, our explanations of life, our theories of how life works, will not be simple either" (Mitchell, 2009, p. 13). The second general foundation is the conception of scientific explanations as representations, which implies a necessary partialness in the descriptions they make of the world.

These two premises underpin her two broad argumentative strategies for pluralism. The first of these approaches, covered in the present chapter, is a vertical approach, characterized by anti-reductionism and pro-emergentism, with the objective of dismantling the leading contender for the unification of science, namely reductionism. The second approach, covered in chapter 3, is horizontal insofar as it proposes that though explanations of concrete phenomena at the local level may only admit of a single explanation—implying a lack of pluralism—, those at the broad, abstract, level can be amenable to multiple explanations. Thus, according to both the vertical and horizontal approaches, *fragmentation* pluralism obtains, since unification is precluded due to the failures of reductionism, as well as the impossibility of accounting for broad (biological) phenomena with a single explanation.

In this chapter and the next, I take a close look at each of Mitchell's arguments to show that they ultimately fail as an adequate defense of fragmentation pluralism. I will argue that though Mitchell's premises are sound, the conclusions she draws from them are either not supported by the evidence when interpreted in their strong form, or uncontroversial when interpreted in a more tempered way. A recurring issue is that Mitchell tends to inflate the concepts she puts forward, such as emergence, downward causation, and even pluralism, to the point where they can be acceptable even to purported detractors of her positions. The result is an uncontroversial form of pluralism which is at worst unfalsifiable, and at best a form of pluralism which ends up looking a lot like monism.

This chapter begins by describing different forms of reductionism, focusing on epistemological reductionism (section 3.1.1). I then describe Jaegwon Kim's functional reductionism, which is the particular form which Mitchell criticizes (section 3.1.2). In section 3.2, I tackle the two anti-reductionist arguments put forward by Mitchell, and show how they fail to properly take into account Kim's claims and assumptions. Section 3.3 covers Mitchell's arguments in favour of emergentism. I show how her claims that contemporary biological research describes emergent phenomena through novelty and unpredictability (section 3.3.2) and downward causation (3.3.3) do not run counter to reductionist claims. I conclude with section 3.4, arguing that Mitchell has not shown that reductionism is impossible in principle, and that her form of anti-reductionism seems to be unfalsifiable.

2.1 Forms of reductionism

It comes as no surprise that Mitchell—and indeed virtually all pluralists—adopts an anti-reductionist position, since reductionism is the foremost way that explanations are understood to be unifiable. Reduction itself is a multifaceted concept that can be broadly understood as the idea that theories or explanations can be made to correspond to more fundamental theories or explanations.[5] In this view, biological phenomena—and presumably all phenomena—are understood to be found within a hierarchy of levels of organisation, going from ecosystems, to organisms, to neurological systems, to cellular mechanisms, and eventually all the way down to fundamental physical matter and energy. Reductionism proposes that phenomena or properties at higher levels are explainable through the phenomena or properties at lower levels. So, for instance, the behaviour of an animal can be explained through its neurological system, the neurological activity can be reduced to firing of the neurons themselves, which in turn can be reduced to chemical reactions, and so on and so on. If reduction works, and is applicable to any and all cases, then eventually all explanations could be reduced to the fundamental level—presumably

[5] There is another, colloquial sense to "reduction" which is used to denote explanations which are overly simplistic, or crude, as when talking about a "reductive explanation." For the purposes of this thesis, I will be looking exclusively at the more technical use, as found in writings pertaining to philosophy of science, and more specifically at 'functional reduction' as described in this section 3.1.

fundamental physics—and scientific explanations would be unified. This makes the repudiation of reductionist positions a practical necessity for fragmentation pluralists, in order to discount one of the most well-known approaches to the unification of science.

Though this thesis concerns explanatory pluralism and therefore explanatory reduction, a very brief overview of all forms of reduction is presented, to situate the present research within the possible reductionisms. The rest of this chapter pits Mitchell's anti-reductionism and emergentism against Jaegwon Kim's (1999) defence of reduction, showing how Mitchell's approach ultimately fails to properly address Kim's arguments.

2.1.1 Ontological & epistemological reductionism

Many forms of reductionism have been defended in philosophy of biology, including ontological, epistemological, and methodological approaches (Ayala, 1974). Ontological reduction, otherwise known as physicalism or materialism, is the (by now) uncontroversial idea that every biological entity is composed entirely of physical stuff and nothing more. This implies that for each entity, property, or process, there is a corresponding lower-level physico-chemical entity, property or process (or set thereof). Though most everyone agrees that there is nothing over and above the physical, certain details remain to be clarified (Dowell, 2006). Nevertheless, this intuitive idea is often the motivation for the other forms of reductionism, though they are not entailed by materialism (as Mitchell's, and others', defence of non-reductive physicalism testifies). Methodological reduction is the position that the best way to study a biological phenomenon is to look at its underlying structure and components (Andersen, 2017; Wimsatt, 2006), as can be the case with mechanical explanations (Bechtel & Richardson, 2010). Epistemic reduction is the type of reduction which is most debated within the philosophy of biology, and is the proposal that certain bodies of scientific knowledge can be reduced to more fundamental bodies of scientific knowledge (Brigandt & Love, 2017). This last form of reductionism is the kind that Mitchell is most concerned with when defending integrative pluralism.

Yet even within epistemic reductionism many positions exist: just as there are many ways of characterizing bodies of scientific knowledge, from theories, to laws, to models, or explanations, and there are even more ways of describing the possible reductions between them. The most well-known form of epistemic reduction is Nagel's (1961) proposal that intertheoretic reduction can be done through logical relations between the various elements in either theory, sometimes through the use of bridge principles, with the goal of showing how a theory of a higher level can be deduced or derived from a theory of a lower level of organisation. This kind of interlevel theory reduction sparked much discussion in the second half of the twentieth century, with Kenneth Schaffner applying it specifically to the biological contexts, in an attempt to reduce classical genetics to biochemistry (Schaffner, 1967, 1974, 1993, see also 2012, 2016). In philosophy of biology, intertheoretic reduction has mostly fallen out of favour in the twenty-first century, due in part to problems internal to Nagel's and Schaffner's accounts (van Riel, 2011), the abandonment of the syntactic view of theories, on which this kind of reduction is premised (Culp & Kitcher, 1989; Sarkar, 1998), and the ascertainment that biology has not in practice proceeded in the way predicted by Nagel and Schaffner.

More recent accounts of reduction in biology tend to focus instead on explanatory reduction, which allows reduction to apply more locally, to parts of theories, mechanisms, or individual explanations. This typically revolves around finding the underlying causes or constituents of a given phenomenon as follows: identify a higher-level phenomenon or explanation, and attempt to relate it in various ways to a lower level. This can be done for instance by finding the most relevant causal or constitutive components in the lower level (Kauffman, 1971; Wimsatt, 1976b). Kenneth Waters further suggested that this can be done by identifying and relating the difference-making principle found both at the higher and lower level. For instance, the classical mendelian version of genetics posited the existence of genes (the higher level), which molecular biologists showed to be reducible to particular bits of DNA (the lower level), since those parts of DNA were responsible for the observed difference in phenotypes attributed to genes (Waters, 1990). Because their causal roles correspond (in Waters' parlance: their difference-making), genes can be reduced to DNA. Mechanistic explanations could also be considered as a form of explanatory reduction, explaining by decomposing the higher-level phenomenon into its interacting parts (Bechtel & Richardson, 2010; Machamer et al., 2000). Reductionism could also be applied to ultimate explanations, including reducing explanations of phenomena in terms of natural selection to explanations invoking Mendelian genetics or molecular genetics (see for instance Kaiser, 2015; Rosenberg, 2006; Weber, 2005). Of course, there are practical limitations to the possibility of reducing ultimate explanations, such as the fact that much of the nitty-gritty details regarding what specifically led to certain selection pressures is lost to time, and the fact that ultimate explanations can synthesize a vast number of causes for a phenotype into a single explanation, yielding a potentially unwieldy reduction. But it is important to remember that even according to Mitchell, it is only the in-principle possibility of reduction that is necessary to make the case for reductionism, meaning that practical limitations are irrelevant. In sum, explanatory reduction is limited in scope: rather than attempting the difficult task of delimiting both a higher- and lower-level theory and relating them in predetermined ways, it puts forward a more modest approach, by emphasizing piecemeal reductions of smaller epistemic units of scientific explanations, such as explanations of particular instances, specific entities, or specific processes, as opposed to entire theories or domains.

Mitchell takes aim at epistemological reductionism, but she does not specify which kind of reduction she is attacking, claiming instead that they all share the conception that what is explanatorily important is to be found at the lowest levels of organisation. According to Mitchell, all forms of reductionism[6] posit an asymmetry of explanatory power in the levels of organisation: "they all share the view that explanation flows "upwards" from the behaviour of fundamental components to the behaviour of the containing system" (Mitchell, 2009, p. 22). The idea here is that regardless of the specifics of the reductionism adopted, all approaches concur that the most explanatorily relevant parts of the phenomenon are to be found in the lowest levels of organisation. One thing to note however is that reductionism need not (and for many, does not) imply eliminativism (the best known example of eliminativism is P. M. Churchland, 1981), generally understood as the idea that once a higher-level phenomenon is reduced, it should then be discarded, with only the lower-level explanation remaining as a valid scientific explanation. To be sure, this tempers Mitchell's claim that reductionism maintains that all explanatory force comes from the lower level, since higher-level explanations are not discarded, merely complimented with corresponding lower-level ones.

It is important to note that Mitchell's claim is not that reductionism fails in all cases, only that it fails in certain cases—described shortly—, and as such, it is only one of many legitimate explanatory strategies in biological sciences (Mitchell, 2009, pp. 22-23,44). In other words, reductionism has some known successes, but it will need to be complemented by other explanatory strategies in the contexts where it fails.

Her approach is furthermore to show that reduction is impossible *in principle* in certain cases, not merely in practice. Others, such as Dupré (1993) have attempted to show how reduction fails by showing that it is unlikely to succeed in practice in certain specific cases. Mitchell, however, considers Dupré's approach to be insufficient:

[6] From here on in, unqualified references to "reductionism" refer exclusively to epistemological reductionism, just as Mitchell uses the term.

> all this shows is that reduction is unavailable in fact in these cases. It does not show that it is impossible in principle. It is the second, stronger claim that is needed to overturn causal completeness [i.e. the reductionist argument which Dupré attacks], for it is in principle reduction that figures as the conclusion of the reductionist argument." (Mitchell, 2003, p. 184)

She has therefore committed to showing that reductionism is in principle impossible, and not only that it could be difficult in practice. This is an important detail for three reasons. First, it is a better characterization of the reductionist position, since reductionists know very well that certain phenomena are not (yet) reduced, and their claims are about the possibilities of an ideal science, not the state of current science. Second, it is a far more difficult position to criticize as an anti-reductionist, since it will be necessary to show that certain phenomena, and the way science is able to characterize them, are—and always will be—irreducible to their component parts. And third, I will be defending the in-principle possibility of integration of all explanations in chapter 7, and the in-principle functional reduction described in the present chapter is one of the components of such an approach, meaning that Mitchell's target is precisely the one I defend. In other words, my position stands and falls along with the form of in-principle functional reduction defended in this chapter.

Mitchell criticizes reductionism using two interrelated strategies, which will be explained in detail throughout the next pages. The first is a series of anti-reductionist arguments, which attempt to show how and where reductionist arguments fail. The second is to argue that emergent phenomena exist in contemporary science. This idea will be explained in detail, but briefly, emergent phenomena are understood to be novel and unpredictable with respect to their constituent base, meaning that they are not reducible to the underlying microstructure. Thus if Mitchell can show that emergent phenomena exist, then reductionism has been shown to fail at least in those cases (Mitchell, 2009, p. 24).

2.1.2 Kim's reductionist argument

To make a case for emergent phenomena and to ground her argumentation in the philosophical debate, Mitchell reviews and criticizes one of the most well-known anti-emergentist positions: Jaegwon Kim's (1999) "Making sense of emergence." In this paper, Kim sets out to describe in detail what would be necessary in order to reconcile materialism with anti-reductionism through the possibility of emergent phenomena. This is precisely the position advocated by Mitchell, who defends a 'nonreductive materialist' position, affirming both that phenomena are only ever made of physical stuff, but that our best scientific explanations sometimes will not be reducible. Emergent phenomena are understood to be antithetical to reductionism, since an emergent property, by definition, is not reducible to its constituent parts, nor predictable from them: “to be emergent is to be nonreducible” (Mitchell, 2009, p. 27). Kim describes the necessary conditions for nonreductive materialism to even be possible, and in so doing, shows how emergent phenomena seem to be an impossibility if certain materialist assumptions are accepted. We will begin by looking at how Mitchell characterizes Kim’s thoughts, followed by the many ways Mitchell criticizes Kim, and how those criticisms fail to properly tackle Kim’s arguments.

Kim begins by describing a particular form of reductionism called functional reduction and the necessary conditions for that reduction to go through. This allows him to clarify the reductionist position, but more importantly for his purposes, to later identify what it is that a phenomenon and its explanation need to do differently if they are to be labelled as emergent. To reduce a phenomenon, one begins with a functional description of the higher-level phenomenon: “a functional description of a property is one in terms of causes and consequences instead of structural components” (Mitchell, 2009, p. 27). Mitchell gives the example of a chair, which functions as something on which to sit, which is (at least to some extent) divorced from the actual material components and structure; a chair can be made of various different materials, and can come in a variety of shapes while still functioning as something to sit on. One can apply the same logic to any higher-level phenomenon by identifying it as ‘what has been caused by X, and what in turn causes Y.’ Now, because compositional materialism is taken for granted, it follows that there is a lower level of material substrate which corresponds to that which is caused by X and which entails Y as consequence. So for instance, using the classic example from cognitive science, one can pick out the higher-level phenomenon ‘pain’, functionally identify it as “that which is caused by tissue damage and which in turn causes winces and groans” (Mitchell, 2009, p. 28), then point to the lower-level neurological realization as demonstration of its reducibility. Of course, it is not enough to point out a correlation between the higher and lower levels: the higher-level phenomenon must be explained by the lower level (Kim, 1992, p. 126). That explanation could take many forms, including a causal and/or constitutive relation; as Kim puts it, if we could potentially design a microstructure that realizes the higher-level phenomenon, then we could say with confidence that the phenomenon has been reduced (1999, p. 9). Mitchell summarizes Kim’s conclusion thus: if a higher-level phenomenon can be functionally described, “then there will always be some configuration of material components at a lower level that can be identified as realizing the functional property of being caused by X and in turn causing Y” (Mitchell, 2009, p. 28). This reductionist conclusion is borne out by the materialist assumptions since for any causal phenomenon, it is taken for granted that there indeed exists a material substrate which accounts for the higher level.

Of note is that functional reductionism of this sort is fairly narrow in scope. Being a form of explanatory reduction, it makes no claims about reductions of theories or domains, and focuses only on functionalizable phenomena. As such, it allows merely for very local reductions (as opposed to global reductions), reducing explanations applied only to particular phenomena (Faucher & Poirier, 2001; Kim, 1992). By requiring only that the cause and effect of the higher-level phenomenon be explained through the lower-level cause and effect, the reductionist claim is far more modest than the classical Nagelian intertheoretic reductions. One might (legitimately) be tempted to point out that the goalposts have changed, and that reductionism of this sort is a far cry from the original promises of reducing (say) all of biology to chemistry. Nevertheless, this is the reductionism that is targeted by Mitchell, which she attempts to show as being faulty. It is also a far more reasonable understanding of reduction, in line with contemporary scientific research.

Functional reduction is also quite flexible in its application. Among other virtues, it can target any higher-level phenomenon, as long as it has a cause and an effect. This implies that diffuse or complex higher-level phenomena, even those which seem to make sense only at a higher level, are amenable to this type of analysis. For instance, it may be difficult to see how 'information' from the environment, as used by organisms to influence their behaviour, could be reduced (Réale, personal communication). Indeed, being the fruit of an interaction between the organism and the environment and therefore not exactly an entity, it may seem as though 'information' is too diffuse to be reduced, or somehow ephemeral. Nevertheless, insofar as 'information' has a cause and an effect, it is functionalizable, and we can look at the underlying components which constitute or cause it. In a process which Wimsatt characterizes as "finding the larger embedding system" (2000, p. 270), we can simply expand the search for components to the organism *and* the environment, and find 'information' to be reduced to specific interactions between organisms and environments. Another potential difficulty is conserving the very concept of 'information' once it has been reduced. However, though it is true that the notion of 'information' is not to be found at the lower level (at least not in the same way), it is not a concern since no elimination of the higher-level concepts is necessary, only a link between the two. Of course, another possible (and well-known) argument concerns multiple realizability, which will be covered in more detail in the next section (3.2.1).

Now, to return to Mitchell's understanding of Kim's argument: according to Mitchell, in order to arrive at the strong reductionist conclusion that all functionalizable phenomena are reducible, an additional assumption is needed. In order to relate the higher and lower levels, it is necessary for "every material object [to have] a unique complete microstructural description" (Kim, 1999, p.6, quoted in Mitchell, 2009, p.28). The reason Mitchell highlights this requirement is because the reduction is understood to entirely account for the higher-level phenomenon. In other words, the higher-level phenomenon is *nothing more* than the result of the lower-level entities and activities. Mitchell claims this additional assumption of a complete description is necessary because it ensures that the materialist assumption that all higher-level phenomena are nothing more than the sum of lower-level phenomena is reflected in the descriptions we can make of them. In other words, she takes this to be the claim that it must be possible to completely describe a phenomenon at the lower level, without which it would be impossible to ensure that that which is functionally identified at the higher level is indeed that which is picked out at the lower level.

With this characterization of reduction in hand, we can turn to emergent phenomena: if a phenomenon or property is to be emergent, it must somehow circumvent the conditions for reductionism laid out by Kim. One possibility is for it to not be functionally describable, which implies that it would have no causal efficacy. This possibility of epiphenomenalism is rejected outright by emergentists, since an entity with no causal powers whatsoever has no apparent impact in the world, and therefore plays no role in either explanations or the causation of anything (Kim, 1999, p. 22). There is no scientific interest in positing the existence of an emergent phenomenon if it has no effect on the world (Mitchell, 2009, p. 29). Another possibility, described in the next subsection, is that the identity relation between the higher level and the lower level is not so simple, and that it is for some reason impossible to establish the correspondence between levels. If this can be demonstrated, then it could preclude the possibility of reduction. This is the strategy that Mitchell employs, through a variety of arguments, outlined in what follows.

2.2 Anti-reductionist arguments

2.2.1 Anti-reductionism through representation

Mitchell takes aim at Kim's position starting with anti-reductionist arguments criticizing some of Kim's purported assumptions. The first of these assumptions is that "there is always a *unique* and *complete* description of the higher-level phenomena in terms of the lower level" (Mitchell, 2009, p. 30, my emphasis). Mitchell attempts to invalidate it through two arguments:

(1) some higher-level phenomena are multiply realizable, implying that the uniqueness condition is not satisfied;

(2) explanations are representations (see chapter 1), and as such are always at best a partial description of a given phenomenon, implying that the completeness condition is not satisfied.

I will describe each of these criticisms, and how they can be shown to fail when Kim's full paper is taken into account.

Let us start with (1) the argument that multiple realizability is a blow to the uniqueness condition. Multiple realizability is a notion made famous through debates in philosophy of mind and cognitive science (Fodor, 1974; Putnam, 1965), and is the idea that a given higher-level phenomenon—for instance, a mind-state—can be instantiated in multiple different ways at a lower level—for instance the brain-states. This is part and parcel of a functionalist approach to higher-level phenomena, since, by definition, and as mentioned earlier, a given function can be carried out through many different means. Moreover, higher-level phenomena can be understood as either tokens or types; whereas a token is a particular, concrete instance of a phenomenon (e.g. the pain felt at time *t* by person *x*), a type is the set of all functionally similar phenomena (e.g. the pain felt at all times by all people). According to Mitchell, once "we are concerned with *types* of higher level phenomena (rather than particular instances), then [Kim's] uniqueness claim is not satisfied" (2009, p. 30). So, for instance, if the type 'pain' is multiply realized through various neurological states, it stands to reason that it is not reducible to any particular *one* of those lower-level states, since it is reducible to any of a disjunctive set of neurological states (state N_1, or N_2, or N_3, … or N_x). Since types come in a variety of tokens, this implies that for each type there is a disjunction of tokens which can instantiate it, precluding the uniqueness condition for correspondence between levels.

The first thing to note about this argument is that while Mitchell is very careful to differentiate ontological and epistemic reduction in her writings, Kim can at times be ambiguous. So, while he does indeed lay out the assumption that "every material object has a unique and complete description" (1999, p.6), this could be interpreted in two ways. One is an ontological reading, since he is referring to a "material object": under this interpretation, he is simply stating that material objects have a unique material substrate, a statement that Mitchell readily endorses. The second possible interpretation is epistemic: because Kim talks about a "description," it seems he is referring to the knowledge we have of the lower level. But on this gloss, Mitchell's critique misses the mark, since in this passage Kim is not talking about a *type* of material object, but a *token*: a single material object; after all, he is proposing functional reductions, which are local. Mitchell, on the other hand, is explicitly talking about the multiple realizability of *types* of phenomena; as such, it is not clear that the assumption of a unique microstructural description of tokens can or needs to be extended to types.

But of course, types of phenomena *do* exist, and Kim's reduction needs to address this possibility, since multiple realizability does preclude the uniqueness assumption. However, according to Kim, this is clearly not an issue. Indeed, he seems to have no problem accepting that a higher-level (type of) phenomenon could be reduced to multiple realizers in its base domain. Whereas he talks of "material objects" having a unique and complete description, the same does not hold for the functionalized properties at the higher level: "Clearly, multiple realizers for *E* [the functionalized higher-level property or phenomenon] are allowed on this account; so multiply realized properties fall within the scope of the present model of reduction." (1999, p.11) Kim's logic seems to be that so long as each token of a higher-level phenomenon can be reduced to its base domain, then the functionalized property it picks out can also be said to be reduced to its base domain. The base domain for a type of phenomenon then includes all the tokens necessary to account for all the instances of the phenomenon. For Kim, the functionalized higher-level phenomenon *is nothing more than* having any one of the set of functionalized lower-level entities and relations, and is the basis for reduction. In that respect, it is irrelevant whether there is a single lower-level description or multiple. As Faucher and Poirier (2001) point out, Kim is only one of many reductionists who answer the multiple realizability argument by replying that they have never seen it as a problem (e.g. Bechtel & Mundale, 1999; Bickle, 1998; P. S. Churchland, 1986; Enc, 1983; Hardcastle, 1992; Poirier, 2000; Sober, 1999).

Of course, explanations which marshal types of phenomena will typically have a greater level of generality than do those which focus on tokens, since they encompass many instances. Nevertheless, as long as every token which is a member of the type can be locally reduced, one can still claim that the type has been reduced. After all, since functional reduction is local (unlike intertheoretic reduction), there is no requirement that every instance of a type be reduced in the same way. The very fact of the generality of an explanation based on types practically guarantees that it will be realized in slightly different ways depending on the context, which does not undermine the fact that every instance is still is realized by its lower-level components. Here again the same point holds: as long as every token can be reduced, the anti-reductionist claim falls through. In sum, Mitchell seems to have taken the assumption of a unique lower level for tokens as a *sine qua non* condition for reduction, whereas Kim explicitly denies that this is a necessity.

Let us now turn to the second of Mitchell's critiques, reproduced here for convenience:

> (2) explanations are representations (see Chapter 1), and as such are always at best a partial description of a given phenomenon, implying that the completeness condition is not satisfied.

This approach relies on the difference between the actual world and the representations explanations make of it, arguing that representations are never "complete." As described in chapter 1, though Mitchell takes for granted that the world is such that only one causal history accounts for any phenomenon, no matter the levels involved, scientific explanations qua representations are condemned to remain a partial description of that world.

According to her, this is another reason that ontological reductionism does not translate to its epistemic counterpart: "there may well be a complete causal process engaged in by *physical* entities: what else could there be? But at the same time there will not be a representation that completely captures this process in terms of *physics* entities" (2009, p. 33, emphasis in original). Mitchell's example is of a window being shattered by a rock; though all the interactions are between physical entities, any given physics theory will not be able to capture all the facets of the event in all its glorious detail. This claim, restated to remain within biology, is no doubt accepted by essentially all researchers: all biological phenomena are the result of a causal process engaged in by biological entities, and any explanation we make of that process will omit certain details[7]. Denying this partiality of explanations is a practical impossibility for at least two reasons: first, particular instances of biological phenomena are the result of an infinite causal chain, meaning that a (finite) complete description is forever out of reach. For instance, the phylogeny of a given trait in a particular organism is exclusively the result of interactions within and among biological entities and their environment, yet a 'complete' explanation is impossible due to the infinite and contingent intricacies of evolved complexity (EC, as described in chapter 1). Second, there are myriad accidental circumstances for any particular phenomenon which are irrelevant to its explanation (e.g. the colour of a billiard ball is irrelevant to its motion), and other details which are omitted because they do not result in any measurable changes (e.g. small irregularities to the sphericity of a billiard ball)[8]. Even examples of biological phenomena which are constrained within a very limited timeframe, as Mitchell's broken window

[7] On this view, an explanation *need not* be complete to nevertheless be an explanation; see Chapter 5 section 6, and Chapter 7 for further discussion.

[8] Many thanks to Tudor Baetu for pointing this out and providing examples.

is, are the result of a seemingly infinite number of interacting parts, if one wants to dig down deep enough, or consider the local, contingent components. As such, explanations of a given phenomenon will be partial, even if the event itself is the result of the interactions within the microstructure. We will return to this issue in more detail in chapter 7, arguing that while explanations are partial, the *objective* of scientific research remains to uncover all the relevant causes; for the moment however, let us take Mitchell's view for granted: explanations are always partial representations.

So, while for Mitchell ontological reduction seems to be a necessity, the partiality of the representations could preclude epistemic reduction: "the partiality of any representation leaves open the possibility that the two representations will simplify the phenomenon in incompatible ways" (Mitchell, 2009, p. 31, see also 2003, p. 185). It is thus possible that each level leaves out elements of the phenomenon described, and that whatever is left out is precisely what is needed for the interlevel functional equivalence to work. Up to here, the argument would be hard to reject: as soon as one accepts the idea that explanations are always partial—which seems uncontroversial—then it is clear that the partiality could result in incommensurability between levels of explanations. However, as will be explained in the next paragraph, this is not sufficient to invalidate reductionism, and could merely be a temporary hurdle to overcome. This may be why Mitchell goes further, claiming that not only is the correspondence between explanations at each level not guaranteed, it is in fact impossible: "without the requirement of a unique, complete description of the lower level, the functional identification *will not go through*" (Mitchell, 2009, p. 33, my emphasis). Mitchell's claim is that any missing part of the physical processes in the higher- or lower-level explanations will be enough to derail all possibility of reduction. She takes Kim's assumption that every physical phenomenon has a unique and complete lower-level description as the foundation of functional reduction; thus, if explanations are partial, then functional reductionism is an impossibility.

Yet here again, it seems Kim anticipated this potential problem, explicitly stating that parts of the description of the phenomena will be left out. In the example used earlier, 'pain' is functionally identified as that which is caused by tissue damage and causes winces and groans, and is related to a certain neurological state. This relation is discovered through empirical means, which explains in part why the quantity of information used regarding each level's phenomenon will be variable: "An important part of this procedure is to decide how much of what we know (or believe) about *E*'s [pain's] nomic/causal involvement should be taken as *defining*, or *constitutive of*, *E* and how much will be left out" (p.11). What is left out will depend on empirical knowledge, context, and theoretical desiderata of various sorts. To give a more contemporary example, research into spatial memory in rats is a multi-level research program, spanning mice navigating mazes, to identification of activated brain areas, all the way down to NMDA receptors in the neurons (see Craver, 2007, Chapter 5). Research progresses through the accumulation of information and explanations, creating mechanism sketches which will 'black-box' parts of processes which are as yet poorly understood or unknown (Craver, 2007, pp. 113–114). Because of this, the work of reducing the function of 'spatial memory' to a neurological substrate will take many years of research, and—importantly—will not be the result of having a complete description of either 'spatial memory' nor the neurological state. Instead, the link between either level will first be hypothesized, then the descriptions will be added to over time, leading to a more precise functional relation between the levels. The empirical research will therefore progress despite the fact we will not have a complete description of the microstructure; as long as there is a sufficient amount of information to claim that a certain neurological substrate plays the functional role of the higher-level phenomenon, then reduction is a possibility. In other words, reduction is not an all-or-nothing explanation, but is a process which progresses along with discoveries in science, and therefore does not need a complete description neither of the higher- nor lower-level phenomena.

There remains a question regarding to what extent it is desirable to be tolerant with respect to the paucity of information at either level. At one extreme, one would not be inclined to say that 'pain' has been reduced if the description of the microstructure is severely limited. The claim that 'pain' is reduced merely to 'physiological functions somewhere in the body' would clearly be overreaching, with the description of the microstructure being far too coarse to warrant the label 'reduction.' It seems there must be some explanatory or predictive power to the relation for it to be considered reduced. Yet the other extreme, which Mitchell apparently endorses, is too stringent, requiring a complete description of the microstructure. As just seen, Kim denies that a complete description is necessary, pointing out the necessity of partial descriptions for research into reduction. And indeed, partial descriptions can certainly, under certain circumstances, yield sufficient explanatory and/or predictive power to be considered explanations, despite the fact that they are not complete. The important point for reduction is not the completeness of the explanation, but that it be sufficient for the explanatory or predictive purpose.

Curiously, there is also good reason to believe that Mitchell herself would need to admit that the requirement for completeness is too severe. The reason for this is that Mitchell advances three different claims, which taken together are incompatible. The first is that Mitchell admits that

(a) reduction is a possibility, at least in certain cases (2009, pp. 22–23).

She readily acknowledges that reductionism "is not a wrong-headed strategy; it is an incomplete one. It should be part of a more full-textured epistemology, not the only game in town" (2009, pp. 22–23). This recognition of reduction as a valid explanatory tool, despite being limited in application, is inconsistent with two other elements of her argumentation, seen in the previous paragraphs:

(b) reduction requires complete descriptions of the microstructure, without which "functional identification will not go through" (2009, p. 33);
(c) scientific explanations, being representations, are always partial (2009, p. 23)

On the one hand, she emphasises the supposed reductionist requirement for a complete description of the microstructure, yet on the other hand, she claims that scientific explanations, insofar as they are representations, can in fact never provide a complete description of anything, due to "the partial character of any and all representations" (2009, p. 23). Defending both (b) and (c) implies that reductionism is in fact impossible in *all* cases: if, given (c), any given explanation is always partial, and, given (b), the partiality of explanations implies the impossibility of reduction, then reduction is impossible for all scientific explanations, which is in contradiction with (a). At least one of the three claims must be abandoned in order to remain consistent: either reductionism is impossible (abandoning (a)), a complete description is not necessary for reductionism (abandoning (b)), or representations can be complete (abandoning (c)). Both Kim and Mitchell agree that (a) reductionism *is* possible, and both agree that (c) representations will be incomplete. It seems then that the simplest solution is to abandon (b), or as described earlier, the idea that "reduction requires complete descriptions of the microstructure, without which functional identification will not go through". Kim explicitly does not require it, and Mitchell offers no real justification for why it would be necessary, merely stating that it precludes reduction.

One could be tempted to point out that the claim that (a) reduction is a possibility is couched in terms of an explanatory "strategy", rather than a claim about the actual possibility of reduction. This reading could imply that Mitchell endorses reductionism as an investigative strategy, which could lead to partial reductions, and is only excluding the possibility of complete reductions (claim (b)). Yet even this gloss does not seem to be weak enough to discount Kim's position. After all, Kim does not require complete descriptions, and explicitly admits that reduction will be a process which is more or less complete depending on the state of the empirical findings. As previously mentioned, reduction is not an all-or-nothing affair, and to characterize it thus is to make a straw man out of the position. Thus, even with partial reductions, Kim's position is such that the functional identifications *do* go through (contradicting (b)), meaning that the requirements Mitchell lays out are too stringent.

In sum, though the descriptions offered by explanations are clearly partial, it is enough to establish that what satisfies the function of 'pain' could be identified with 'such and such a neurological substrate' at the lower level, without needing a complete representation of the neurological state. The partiality of explanations need not be a major roadblock to reduction. In fact, it is part of the normal process of scientific discoveries: as more and more empirical data is collected, hypotheses tested, and experiments carried out, explanations which start out being tentative and very partial will become more fleshed out. As will be addressed in later chapters, any temporary incommensurability caused by partiality will be overcome through integrative strategies (chapters 5 & 6), ontological commitments, and the drive to provide more complete explanations (chapter 7).

2.2.2 Anti-reductionism through features of phenomena

Setting aside the issue of explanations as representations, Mitchell also formulates anti-reductionist critiques which rely on features of the phenomenon to be reduced, or features of the underlying microstructure. This is based on her understanding of biological phenomena as constitutively and dynamically complex (CC and DC), as laid out in section 3.1.

Focusing first on CC, her approach is to claim that the lower, physical, level alone cannot account for the causality of the higher-level phenomenon because it lacks structure. As previously mentioned, emergent phenomena must have causal efficacy, but that causality cannot be reduced to the causality of the lower level, otherwise the phenomenon would not be properly emergent. Mitchell proposes that what can account for the new causality at the higher level is the structure of the substrate: "if the physical level is construed only materially, then structure is a level up and causally significant" (2003, p. 185). On her reading then, reductionists attempt to reduce everything to the physical level, which includes physical entities, but does not include the structure of all these physical entities together. If that is the case, then structure is a higher-level property, and that structure is causally efficient at a higher level than the lowest, physical level. Hence emergent causality obtains, since it is not reducible to fundamental physical entities.

There are a few problems with this objection. For starters, as Plutynski (2004) points out, it is unclear what Mitchell has in mind when talking about the causal powers of structural features of the world, and how to make sure that no micro-level explanation could account for that causality. But more to the point, once again, Mitchell's reading of reductionist claims seems to be in contradiction with claims by reductionists themselves. The very term "microstructure," used by Kim (1999), implies that the lower level is not composed only of the entities, but also how they hang together. In fact, Kim describes the three elements needed for a complete description of the microstructure: "(i) the basic particles that constitute it […]; (ii) all the intrinsic properties of these particles; and (iii) the relations that configure these particles into a structure" (1999, p. 6). It is thus quite clear that the structure need not be understood as a macro-level property, but instead included in the microstructure. To deny this would lead to overblown claims: it could be argued that all objects, properties and phenomena that are not single entities at the level of fundamental physics rely in some sense on the structure among fundamental particles, implying that every single thing except fundamental particles is an emergent phenomenon. Even the conical shape of a pile of sand would be emergent, since on this reading the structure of the aggregate pieces of sand would be a higher-level property. Surely this is not a conclusion that Mitchell (or any anti-reductionist) would endorse and indeed, Mitchell does not repeat this line of attack in her subsequent book (viz. 2009).

Another of Mitchell's critiques relies on the dynamic complexity (DC) of biological phenomena (see Chapter 1 for a characterization of DC). She argues that reductionists are unable to consider the dynamic aspect of certain lower-level phenomena which lead to the emergence of the higher-level phenomenon. According to Mitchell, reductionists rely on "a static snapshot of the higher and lower levels" of the phenomenon; but with such a static view, "the dynamics of *how* the higher level is constituted is lost" (Mitchell, 2009, p. 32). For instance, reductionism would be unable to account for the flocking behaviour of birds, or the division of labour in social insects, since these rely on feedback loops, and relations between entities that change over time (see sections 3.3.2 and 3.3.3 for discussions of these two examples). Looking only at the microstructure at a given point in time makes it impossible to see how the higher-level phenomenon could be the result of interactions at the lower level, rather than mere static microstructure. What this occlusion of the dynamics misses

> is a question at the center of much scientific concern with emergence, namely, how is the property at the higher level produced, and what are the differences among the many kinds of relationship between higher- and lower-level properties that occur in nature? (p.32).

According to Mitchell, reductionists deliberately ignore the dynamics of the microstructures, leading them to be blind to how some higher-level phenomena are produced.

The first thing to point out is that it seems at best a misunderstanding to claim that reductionists would not be interested in knowing how a higher-level property is produced by the lower-level entities and properties. After all, reductionism is precisely concerned with the relationship between higher- and lower-level properties, proposing that the higher level is explainable by the lower level. The real question is thus whether the dynamic aspect of the lower levels can be taken into account by reductionists.

Different versions of reductionism may have different answers to this question. However, the functional reduction proposed by Kim seems to be well suited to consider the dynamics of the microstructure. The higher-level phenomenon is described through its position in a chain of cause and effect, which in turn is related to the microstructure which realizes it. Even in cases where the higher level is described as a single property or phenomenon, it is very likely that the microstructure is composed of multiple entities and the causal relations between them. To return to the example used above, even though 'pain' is taken to be a single phenomenon, the neurological substrate is understood to be a chain of cause and effect, not a static snapshot of a brain-state. After all, the activation of certain brain regions and the interactions between synapses are dynamic processes that happen over time as a chain of cause and effect. One can imagine countless other examples that reductionists would likely endorse where the higher level is accounted for by a dynamic process at the lower level, some of which will be examined in more detail in the following section. Figure 2.1 starkly illustrates the possibility that the microstructure could contain its own internal chain of cause and effect, despite the fact that the higher-level phenomenon is understood as a single entity[9]. Contrary to what Mitchell contends, the microstructure need not be understood as a static snapshot, and instead can also be conceptualized as a chain of cause and effect, including feedback loops, happening at the level of the microstructure, which accounts for the cause and effect observed at the higher level.

[9] Though the figure is taken from an account of mechanistic explanations, it nevertheless applies to functional reduction of the kind endorsed by Kim (1999) since the higher level is presented as the result of a certain cause and entailing a certain consequence, and the lower level is presented as the microstructure accounting for the higher level.

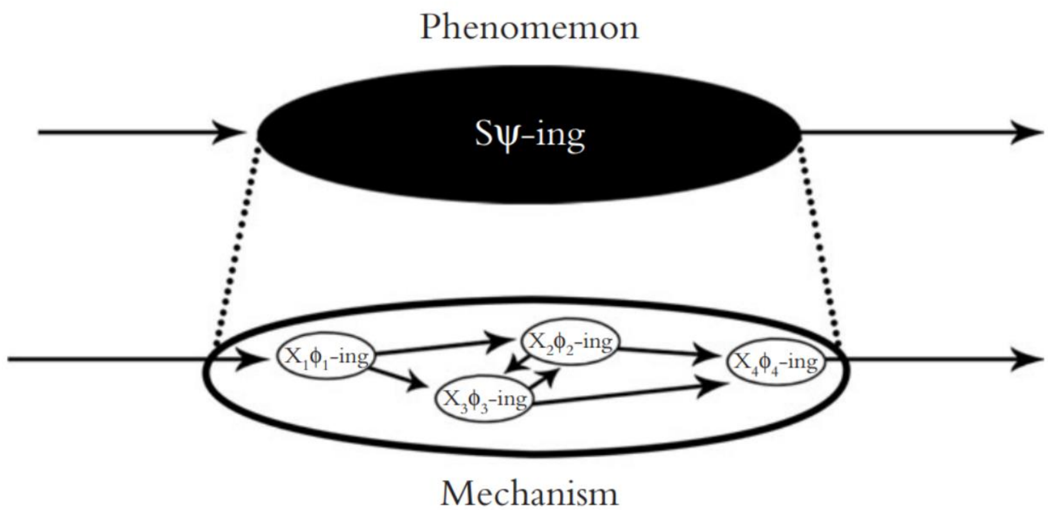

Figure 2.1 - An abstract representation of a phenomenon and its underlying structure. The circles represent entities, and the arrows represent activities. The microstructure here contains many activities, implying a dynamic aspect. Taken from Craver (2007, p.7).

Another interpretation of Mitchell's claim, supported by her example of the flocking of sparrows, is that the DC which reductionists are unable to account for is not at the lower level but at the higher level. The higher-level phenomenon is dynamic, including changes over time, and perhaps feedback loops and othercomplications of the sort which, according to anti-reductionsists, could not be captured by a reductive explanation. However this too is a misunderstanding of the possibilities of functional reduction. Figure 2.1 shows the dynamics of the lower-level entities, but there is no need to stop there: for each X which is Φ-ing, an additional lower level can be added, with its own dynamic constitution described. As such, the dynamic level of interacting Xs as well can be related to a lower level, as can any level and any interaction of entites, showing abstractly how even dynamic phenomena can be reduced.

In conclusion, Mitchell's attempts at marshalling CC and DC against reductionism fail, having attacked a straw man. Contrary to her claim, reductionists do take into account—and are interested in—elucidating the structure and dynamics of the lower levels. One could even go so far as to say that it is, in fact, the raison d'être of reductionism.

2.3 Emergence

2.3.1 Emergent phenomena

Mitchell puts forward another strategy to counter reductionist claims, arguing that emergent phenomena exist, and are evoked by contemporary scientists. The existence of emergent phenomena, should it turn out to be true, would imply that reductionism is not applicable across the board. Though it was mentioned earlier that emergent phenomena are those phenomena which cannot be reduced, more details are warranted before moving on to the examples provided by Mitchell.

Emergence, just like reduction, is a multifaceted concept which can and has been defined in myriad ways, but generally understood as the idea that certain properties at higher levels of organisation are not related in a straightforward way to their lower-level constituents. Higher-level entities or properties are generally understood to supervene on their lower-level counterparts, but cannot be reduced to them. Supervenience is broadly understood as the idea that there can be no change in the higher-level entity or property without a change in the lower-level property; however there can be changes in the lower level without changes in the higher level, thus leaving room for multiple realizability (see Kim, 1984). Emergentists, including Mitchell, are committed to materialism, and to the idea that higher-level phenomena are realized through their supervenience base.

Yet despite the supervenience relation, emergent phenomena or properties are understood to be novel, unpredictable, and have causal efficacy (Mitchell, 2009, p. 26). Novelty and unpredictability are sometimes hard to differentiate when it comes to emergence (Kim, 1999, p. 8), both implying that the higher-level phenomenon is not the clear result of the entities and activities of the lower-level. In other words, despite full knowledge of the lower-level entities and activities, the higher level is neither explainable nor predictable. The easiest way to illustrate this is through a counterexample: the classic example of a *non*-emergent property is the weight of a pile of sand, which can be explained and predicted by the aggregation of the weight of each sand particle in the pile (see also Mitchell, 2009, p. 34). This is dubbed a "resultant" property by early emergentists, meant to pick out those properties which are additive and subtractive, and therefore causally explained, and predictable (Morgan, 1923; cited in Kim, 1999, p. 7). On the other hand, for early emergentists, the fluidity or wetness of water could not be explained or predicted through its constituent molecules of hydrogen or oxygen, and as such, was believed to be emergent (Mill, 1843; cited in Mitchell, 2009, pp. 24–25). Emergent properties are such that no matter how much is known about the lower levels, the emergent property will remain in some sense explanatorily independent.

Emergent phenomena also have causal efficacy, implying that the higher level is capable of causation without relying on the causation of the lower levels. This is related to what Kim has dubbed "Alexander's dictum," which points out that if an entity is devoid of causal powers, "it supposes something to exist in nature which has nothing to do, no purpose to serve, a species of *noblesse* which depends on the works of its inferiors, but is kept for show and might as well, as undoubtedly would in time, be abolished" (S. Alexander, 1920; cited in Kim, 1992, p. 134). The possibility of epiphenomenalism is thus to be discarded, since it would imply that the emergent phenomenon would have no impact whatsoever on the world, which makes it at best only marginally scientifically interesting, if not completely uninteresting. This can be further be related to the novelty characteristic of emergent phenomena: if the higher level is to have genuinely novel properties, then those properties cannot be causally inert, otherwise they will be irrelevant with respect to the explanations, or even the world in general. A property which has no causal consequences whatsoever is a mere epiphenomenon of no particular scientific interest.

More than mere causation, the type of causation that higher-level properties or phenomena require in order to be properly causally efficient is what has come to be called 'downward causation' (Mitchell, 2009, p. 26). It is downward insofar as it needs to cause something to its own lower level. To take Kim's (1992) example (see figure 2.2 below), suppose that there is a mental property M, realized by brain property P at time t_1. If M is to have causal powers, as per Alexander's dictum, it must cause the following mental property, which we can call M*, itself realized by brain property P* at time t_2. Now, because emergentists agree that M* supervenes on P*, "this means that M* is instantiated on a given occasion only because a certain physical property P*, its emergent base, is instantiated on that occasion" (p.136). This implies that for M to cause M*, it in fact needs to cause P*, since it is P* from which M* emerges, and not the other way around. In a more colloquial language, this means that the mental-state needs to change its own brain-state in order to produce the following mental-state. There is thus downward causation because the higher-level property M causes a change in its own supervenience base, changing it from P to P*.

The problem with this idea of downward causation is what has been called the "causal exclusion argument" (Kim, 1989). It relies on the assumption that there exists a causal closure of the physical, implying that any physical event is caused entirely and only by other physical events. On this picture, it seems that P is sufficient for the causation of P*, excluding the possibility of M causing P*. Of course, it is not impossible that P* is overdetermined, which is to say that it is caused by two simultaneous sufficient causes, but it

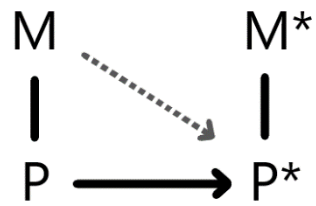

Figure 2.2 - Illustration of Kim's (1992) argument related to downward causation. The dark lines indicate a supervenience relation. The black arrow represents causation. The dotted grey arrow is the supposed downward causation.

seems very unlikely that this would be a regular occurrence. In other words, if it is M which causes P*, that implies that P has no causal link to P*, barring overdetermination. That would imply that the physicochemical laws governing the behaviour at the brain level must be interrupted by M, meaning that the expected physicochemical causal link between P and P* does not happen. If, on the other hand, one wants to claim that P* is in fact caused by P, then there seems to be no room for genuinely novel causal powers for M, since they are reduced to P's causal powers. This type of analysis is not restricted to mental properties, and can be applied to any emergent phenomenon, since they all require downward causation in order to be properly emergent (at least in Kim's and Mitchell's view).

There are, of course, many other ways of interpreting downward causation and emergentism. However, since Mitchell is responding to Kim's take on the issue, I have tried to do it justice in these pages. In the next subsections, I show how Mitchell argues that emergentism exists in contemporary scientific literature, and therefore that Kim's arguments fail. However, I will also show how her evidence for "scientific emergence" relies on a fairly different understanding both of emergence generally and downward causation more specifically, undermining the force of her arguments.

2.3.2 Mitchell's "scientific emergence": novelty and unpredictability

Are there any examples of emergent phenomena in contemporary science? If a biological property or phenomenon could be shown to be novel and unpredictable with respect to its base, as well as effect downward causation, then that would be definitive evidence that reductionism is indeed limited in its application. Mitchell contends that there are examples of such phenomena available, though perhaps only with a more liberal understanding of emergence, calling as she does on "a broader notion of emergent properties that makes sense of scientific use and is part of a more complete epistemology" (2009, p. 30). Kim's version of reductionism is considered too restrictive, and out of touch with scientific practice: "None of the currently scientifically identified emergent properties (e.g. color patterns on mammals, flocking behaviors of birds, division of labour in social insects, etc.) can qualify as emergent on Kim's account" (Mitchell, 2009, p. 29). In order to illustrate the laxer requirements of what she dubs "scientific emergence" (p.34), Mitchell provides examples of phenomena which rely on dynamic complexity, self-organisation, as well as purported examples of downward causation in biological phenomena. Let us consider them in turn.

The first of these examples is the flocking of starlings, which is the result of a dynamic complexity, namely feedback loops and self-organization. Murmurations of starlings fly in an impressive display of apparent order, with large numbers of birds forming fluid, pulsating, and ever-changing shapes in the sky, fluctuating in density. It looks to the untrained eye as if the group acts as one single, coherent organism. This type of group behaviour serves to protect individual birds from predators, reducing predation through dilution (Goodenough et al., 2017), and the patterns created through group flight can serve to confuse predators, reducing overall predator success (Hogan et al., 2017). Computer simulations (Hildenbrandt et al., 2010), corroborated with empirical data tracking the individual positions of thousands of birds within their flock through stereo photography (Ballerini et al., 2008) show that the movements and effects seen at the group level can be explained through the individual behaviours of the starlings which constitute the group (a fact pointed out by Mitchell, 2009, p. 35). However, the complex patterns produced by the flocks are "not predictable by an aggregation of behaviors of individual [birds] in solo flight" (Mitchell, 2009, p. 35), calling as they do on feedback loops between the individuals in flight, leading to the self-organization of the group as a whole. Mitchell's contention is thus that microstructures which rely on feedback loops (and other nonlinear dynamics) are not aggregative, and therefore do not conform to the necessary requirements for reduction.

But what does it mean for the behaviours of the lower-level phenomenon to be non-aggregative? Aggregation calls back to the idea of "resultant" properties which are the consequence of the summation of properties of the underlying entities, and as such are decidedly not emergent. The typical example, mentioned earlier, is the weight of a pile of sand resulting from the addition of the weight of each sand particle. Yet according to Kim, restricting non-emergent properties to aggregations is too strict:

> There is no need to interpret the talk of "additivity" and "subtractability" literally; I believe these terms were used to indicate that resultant properties are simply and straightforwardly calculated and predicted from the base properties. But obviously ease and simplicity of calculation as such is of no relevance here; predictability is not lost or diminished if calculationally complex mathematical/logical procedures must be used. (1999, p. 7)

Keeping in mind that emergent properties are unpredictable and unexplainable through the entities, structure and interactions of the microstructure, it makes sense to say, conversely, that resultant properties are predictable and explainable. The appropriate metric then is not what particular mathematical operation is used to compute the result of the dynamic microstructure, but rather whether or not the phenomena—or specific aspects of the phenomena—are predictable or explainable through knowledge of the underlying entities and activities, regardless of how complex it is to calculate.

To begin with, one must determine what it is that is being explained by these models of starling flocks. The ultimate objective seems to be to understand every aspect of the dynamics of the interactions between individuals which would account for the observed behaviour at the group level (more on this shortly). But the models are unable to account for every aspect of the observed behaviour, and every research article points out which characteristics are captured by their model, and which are left out. For instance, Hildenbrandt et al. (2010) have shown that the density, shape, collective banking while turning, and many other aspects of the flocking behaviour are explained by their model. They understand these aspects to be explained because the visual comparison of their simulations with recordings of actual flocks above Rome reveal that they are strikingly similar. However, they also underline the limitations of their model, remarking for instance that the shape and density of flocks can only be compared qualitatively with actual flocks, in the absence (as yet) of quantitative data over time concerning these characteristics in an environmental context (2010, p. 1355). They furthermore suggest that their model has not taken into account perturbations, such as those which happen when the flock is confronted with a predator (see Hogan et al., 2017 for work on responses to predation), and other parameters that could be found elsewhere than Rome. We are, then, confronted with a situation described in the previous section, where the available explanations are incomplete due to simplifications and abstractions that were needed to carry out the research and provide explanations.

And yet, these incomplete representations of the world allow for the phenomenon, or at least many aspects of it, to be predictable and explainable through the microstructure. Their partialness does not preclude their explanatory power (for a more in-depth discussion regarding this issue see chapter 3). As

Mitchell and many others point out, the patterns and movements of the flocks are nothing more than the result of the interactions between individual birds, without the need to refer to any higher-level entity or property. Hildenbrandt and colleagues state that the flocking behaviour is explained by their model, which contains only inputs regarding the individual birds, and how they perceive only their closest neighbours, requiring "neither perception of the complete flock nor any leadership or complex cognition" (2010, p. 1356). Though there is no denying that these types of phenomena are understood to be "self-organizing," as Mitchell points out, and rely on feedback loops among the individuals, it remains that close study can reveal the underlying mechanisms required to account for the higher-level phenomenon, meaning that they are not novel or unpredictable with respect to their supervenience base in the sense described by Kim. This is not to say that pragmatically speaking, it is always a requirement to reduce phenomena, or that it always yields a better explanation to do so. Depending on the question asked or the research interests, it may be simpler, more practical, or simply more elegant to refer to the higher level without going into the weeds of the lower level. Here again, it is important to remember that I am not advocating any sort of eliminativism, or even a priority of lower-level explanations over higher-level; I am only defending the idea that in principle, functional reductions are always possible.

This remark furthermore applies to more than just murmurations of starlings. Self-organization, which Mitchell takes to be one of the foundations of emergence, has been defined as

> a process in which a pattern at the global level of a system emerges solely from numerous interactions among the lower-level components of a system. Moreover, the rules specifying interactions among the system's components are executed using only local information, without reference to the global pattern (Camazine et al., 2001).

It is thus explicitly understood that the higher-level phenomena are the result of nothing more than the lower-level entities and activities.

Nonetheless, there is talk of "emergence" in the passage, and in the literature in general; indeed, every single article referenced above which concerns starlings mentions 'emergent effects.' It seems, however, to be used in a very different way than it is by Kim. Ronald et al. (1999) spell out explicitly the fact that 'emergence,' as used in artificial life research (such as modelling starling murmurations) ought to be used to denote surprise at the apparent difference between the inputs to the constructed model and the outputs observed once the model is running. It is interesting to note that the only reason they are prescriptive in their definition, rather than descriptive, is that they feel the term has been overused in trivial cases and its significance therefore devalued. Thus, in the articles reviewed by Mitchell and here, though the ‘emergent’ phenomena are considered surprising, they not understood to be unexplainable or unpredictable, but are in fact reducible (in Kim’s sense): “self-organization theory suggests that much of complex group behavior may be coordinated by relatively simple interactions among the members of the group” (Couzin & Krause, 2003, p. 1). This tempers Mitchell’s claim that emergence is a concept which is alive and well in scientific publications: though the term is certainly widespread, at least in the context of flocks of sparrows and more generally as pertains to self-organisation, ‘emergence’ implies only that the higher-level phenomenon may be surprising, interesting, and complex, but nevertheless explainable through the interactions of the individuals which comprise it—in other words, are reducible to their supervenience base. As such, ‘emergence’ has come to mean something entirely different from that which Kim and many emergentists in philosophy are attempting to capture. Wimsatt (2000) makes this same point regarding the usage of “emergence” in scientific literature, concluding, in the same vein as Mitchell, that what is necessary is a redefinition of emergence such that it is “consistent with reductionism.” (p.271) As far as the usage in scientific literature is concerned, this is a reasonable project. However, insofar as Mitchell is attempting to critique Kim and to convince the reader that Kim’s functional reductionism is inoperable, it clearly misses the mark. Just as Wimsatt does, Mitchell calls for a broadening of our understanding of ‘emergence,’ but contrary to Wimsatt, she does not recognize that in doing so, she has come to include even phenomena which are reducible; as such, she can no longer claim that there is a strict opposition between emergence and reduction, since the former no longer precludes the latter, meaning that reduction has not been falsified through her examples of novelty and unpredictability.

2.3.3 Mitchell’s “scientific emergence”: downward causation

This permissive interpretation of the notion of emergence is also found in her understanding of downward causation, as seen in the example she puts forward: the foraging rate of honeybees. As she describes it, the foraging rate of individual honeybees is tuned to the “system-level property” of how much nectar is stored, a property of the hive, implying that the higher level (the hive) “*causally influences* in a feedback loop behavior at the lower level” (the individuals) (Mitchell, 2009, pp. 42–43). The way this causation happens is described in detail by Mitchell: certain bees fly out to harvest nectar, and return to the hive to unload it to a younger bee, who in turn carries the nectar to store it in an empty cell. As the hive becomes more and more full of nectar, the time it takes to find an empty cell increases. This in turn increases the wait time of the foragers, who must wait at the entrance of the hive until a young bee is available to carry their nectar into the hive. The probability of a given bee returning to its foraging activities is related to the wait time at the entrance of the hive (Mitchell, 2009, p. 43; see also Seeley, 1989). What this translates to is that the more nectar there is in the hive, the less foragers will be inclined to forage. Thus, according to Mitchell, there is a higher-level property—the quantity of nectar in the hive—which causes changes in the underlying components of the hive itself, namely the foragers.

In this example again, Mitchell proposes an interpretation of downward causation which, according to reductionists such as Kim, would not, in fact, be labelled downward causation. Recall that in Kim's example of mental properties, for the mental property to be emergent, the supervenience base P at t_1 cannot be the cause of the following emergent base P* at t_2, otherwise there is simply no reason to posit that the mental property M *also* causes P*, since that would imply overdetermination. So for downward causation to occur, M must cause P*, which in turn realizes M*. In the case of the foragers, let us label the higher level property of quantity of nectar in the hive as N at time t_1, realized by a combination of foraging speed and intra-hive distribution speed we can call FD.[10] At time t_2, the quantity of nectar N has ostensibly caused a decrease in foraging and distribution speed, which we can label FD*.

The fundamental problem with Mitchell's example is that the relation between FD and N is not identical to that between Kim's P and M; whereas P realizes M, FD in fact *causes* N (or at least, causes changes in N). Note that according to Kim, the emergence relation is (in some sense) a constitutive relation, implying that it is not causal (Kim, 1999, p. 32). Furthermore, as previously mentioned, Mitchell also draws on Craver (2007) for her notion of levels, who proposes that levels are determined by the constitution relation: higher level entities and phenomena are constituted by lower levels. The quantity of nectar in the hive (N) is constituted by the aggregate of the nectar found in each wax cell. Certainly, foraging and distribution (FD) cause changes in the quantity of nectar in the hive (N), doing so through the aggregate of the nectar found in each cell. As such the activities of the bees (FD) are not constitutive of the quantity of nectar (N). The scenario can thus be interpreted merely as a causal chain from FD to N* to FD*, with no supervenience, no inter-level complications, and therefore no need to posit downward causation (see Figure 2.3). So under this interpretation of the scenario, while it is true that the quanitity of nectar in the hive is the cause

[10] It is interesting to note here that FD is a dynamic process which happens over time, just as is foraging speed alone, the lower-level property Mitchell claims changes due to downward causation. Mitchell's earlier assertions regarding the impossibility of accounting for dynamic processes are here again shown to be exaggerated.

Figure 2.3 - On the left, Mitchell's interpretation of the causality involved in bee foraging. On the right, my interpretation. N is the quantity of nectar in the hive, FD the foraging and distribution speed, both at time t_1. Asterisks denote the same property at time t_2 Straight lines indicate supervenience, black arrows causality, and dotted grey arrows the supposed downward causation.

of a decrease in foraging speed, that causation does not have the requisite *downwardness* for the phenomenon to be labelled emergent: all the causes and effects are on the same level, since no constitutive relation is present, meaning that no inter-level relations are present.

One could point out that there is in fact a supervenience relation insofar as N supervenes on the aggregate quantity of nectar in all of the wax cells (call it NC). In this respect, N is at a higher level, supervening as it does on NC, satisfying the constitutive relation needed for level differentiation, and therefore opening the door to emergence. However, this move does not save downward causation for two reasons. The first is that though it is a supervenience relation, it is not an emergent one, N clearly being the resultant property of the aggregate of NC. The second is that there is no need to posit that N has any causal role, since Mitchell herself has shown that NC in fact does all the work needed to cause FD*, as illustrated in Figure 2.4: the likelihood of finding an empty cell is directly proportinal to NC, which causes changes in the distribution speed and therefore the foraging propensity (FD*). So while it is possible to find a constitutive relation, it does not save the downward causation.

But suppose we bear with Mitchell and accept that FD is a realizer of N; would this then be considered a case of downward causation? Interpreted through Kim’s lens, the answer is plainly no. Mitchell herself has shown that the rate of foraging is not changed directly by the higher-level property of quantity of nectar, but in fact through interactions between the individual bees. In other words, there is no need to posit

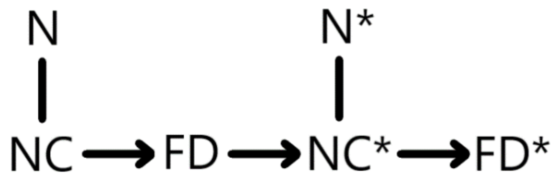

Figure 2.4 - Mitchell's foraging example recast to show the proper constitutive and causal relations, where N is the quantity of nectar in the hive, NC the quantity in each and every wax cell, and FD the foraging and distribution times. Lines denote supervenience, arrows denote causality.

downward causation from N to FD*, since the dynamics of FD are all that are required to cause FD*. Paraphrasing Kim: FD displaces N as a cause of any putative effect of N (1999, p. 32). There is therefore no need to talk of downward causation, since intra-level causation does the job.

To make this point more salient, we can simplify the phenomenon: suppose the foraging bees could simply determine on their own how much nectar there is in the hive, say by looking at the stocks, replacing the indirect information provided by the younger bees. This does not change the nature of the phenomenon in any significant way, merely removing a proxy for information gathering. Just as in Mitchell's example, the foragers could modulate their propensity to forage depending on the quantity of nectar in the hive. At a given time, a forager looks at the hive, sees very little nectar, and returns to forage; later, the bee sees that the hive is full of nectar, and does not return to forage. Would we be inclined to say that the quantity of nectar is an emergent property? Does it exert downward causation on the foragers? I believe that no one would be inclined to answer positively to either of these questions. The quantity of nectar is not a higher-level property, since it is not constituted by the activity of the foragers. It is, however, "the sum of the results of their individual behaviour" (as Mitchell herself puts it 2009, p. 43). What we have is a straightforward causal chain, with feedback from the quantity of nectar affecting the propensity to forage. Now if we simply add the indirect way of assessing the quantity of nectar through the time it takes the foragers to unload, rather than through directly looking at the stores, it seems that there is no fundamental difference in the assessment. Here as well, it is a straightforward causal chain from the foragers to the distributors, to the quantity of nectar, with feedback from the quantity of nectar affecting the distribution speed, which in turn affects the foraging rate. No downward causation; no emergence.

There is however a framework for understanding downward causation which could vindicate Mitchell's example of bee foraging. Malaterre (2011) analyzes claims of downward causation in cancer research through a manipulationist account of causation. He shows that researchers will sometimes group together chains of cause and effect between two or more variables into a single variable at a higher level. In the case of foraging, one could reinterpret the interactions between the young bees looking for a free wax cell and the distribution of those cells within the hive as a single variable which causes changes in in foraging propensity of foragers. These lower-level entities interact in a feedback, "naturally lead[ing] to an "upper-level" causal model that subsumes under an "upper-level" variable the numerous lower-level variables" (Malaterre, 2011, pp. 557–558). Though talk of downward causation in this sense can be warranted, it is important to note that it is not of the type outlined by Kim, since the causation is still understood as stemming from the lower level, with the higher-level entity used as a form of shorthand for the complex lower-level activities.

Mitchell is calling for a broadening of the term 'emergence,' but in so doing has diluted the meaning of the term to the point where it is compatible with reductionist positions. As was shown, the examples she has found in the scientific literature are all examples which reductionists would be happy to point to as examples of successful reductions: the movements of flocks of sparrows are explained by the interactions of individuals, and the downward causation attributed to the foraging propensity in honeybees is in fact an example of straightforward causation. She points to self-organizing phenomena as examples of emergent phenomena, all the while recognizing that those phenomena are explainable through the interactions of their constituents. As such, this reading of 'emergence' is no longer relevant to critiquing Kim, since its interpretation of novelty, unpredictability, and downward causation do not conform to the requirements laid out by Kim.

3.1. Concluding remarks

In sum, Mitchell's claims regarding the failings of reduction and the possibility of emergent phenomena are unconvincing. Her anti-reductionist arguments take aim at what she takes to be strict requirements concerning the uniqueness and completeness of the description of higher-level phenomena in terms of the lower level substrate, which she takes to enable reduction. According to her, these conditions cannot be satisfied by scientific explanations. We have seen, however, that Kim anticipates these possible objections, addressing them in his original paper. This demonstrates how Mitchell's interpretation of these requirements is overly stringent, and that they are not endorsed by reductionists themselves. As for the emergent phenomena, Mitchell calls for a broadening of the notion, but has broadened it to the extent that it includes phenomena which reductionists would applaud as having been reduced.

In this respect, it is hard to imagine what Mitchell would take as evidence that her approach has been falsified. If the dynamically complex systems, or self-organising behaviours, can and have been shown to be the exclusive result of the interactions of lower-level components, what more would be needed for Mitchell to concede that they have been reduced? And if Mitchell sees downward causation in situations where the very sources she references demonstrate straightforward causation, it is unclear what evidence could be brought forward to falsify the claim that downward causation is alive and well as a concept in biological scientific publications. In contrast, the position that functional reduction is possible in principle in all situations is a falsifiable position, since all that is required to falsify it is to show an example of an emergent phenomenon. This is why, though this chapter focused mainly on deconstructing the examples put forward by Mitchell, we can conclude that reductionism of the type advocated by Kim has been successfully defended. Mitchell claims that emergent phenomena exist and that downward causation exists but is unable to provide a convincing example; without an example, the position is hard to believe, flying in the face of strong intuitions such as the causal closure of the physical world and the unlikeliness of overdetermination.

One might be tempted to remark that I have merely shown that her arguments have failed, and have not provided a positive argument for reductionism. But recall that Mitchell set out to show that reductionism is impossible *in principle* in certain cases, a claim she takes to be necessary “to overturn causal completeness” (Mitchell, 2003, p. 184). By showing how her arguments fail to make an in-principle case against reductionism, I have also shown how they leave intact Kim’s arguments in favour of reductionism. Appealing to the complexity of phenomena, to the representational aspect of scientific explanations, or to the purported existence of “scientific emergence” have not proven sufficient for dismantling the case for the possibility of reduction in principle.

In conclusion, Mitchell’s vertical approach to fragmentation pluralism is to be discarded. However, she does have another card up her sleeve, namely her horizontal approach to fragmentation pluralism. This second defense of explanatory pluralism relies on many of the same premises as her anti-reductionism but puts forward arguments which do not rely on the possibility of reductionism, and as such, will be dealt with in the next chapter.

CHAPTER 3

HORIZONTAL APPROACH TO FRAGMENTATION PLURALISM: MULTIPLE MODELS

Mitchell's second broad line of arguments in defense of pluralism proposes that there can exist multiple explanations for a single general phenomenon. Whereas the anti-reductionist stance is a vertical approach to pluralism, arguing that the relations between higher- and lower-level phenomena can preclude unification, the pluralism addressed in this chapter is horizontal, since multiple explanations co-exist on a same level (Mitchell, 2003, p. 189). This argument relies on the difference between a given phenomenon understood as a particular concrete case, or as a general, abstract set of phenomena, as well as the partiality of scientific representations, and the distinction between models and explanations. I will unpack Mitchell's argument, and show how, though it may allow for a multiplicity of explanations, calling this 'pluralism' risks devaluing the term to the point where it loses virtually all scientific and philosophical interest. But first, I will begin by describing the example Mitchell proposes to illustrate this position, giving us something tangible to which we can apply the argument. To do so, I will gloss over the differences between general and particular phenomena, as well as models and explanations, leaving that discussion for the subsequent sub-sections.

3.1 Example: division of labour in social insects

Colonies of social insects such as bees or ants will typically operate with some form of division of labour. Individual insects in those groups will not perform every task required of the colony, instead focusing on one or a certain subset of tasks required by the colony as a whole. For instance, some bees will tend to forage, others handle the food, others tend to the queen, and those tasks will tend to change in a predictable way over the lifetime of the individual (see Winston, 1991). These phenomena all together form the abstracted generalized phenomenon of 'division of labour in social insects', which includes all cases of social insects which operate with some division of labour.

Mitchell reviews three different idealized models which have been proposed to account for the division of labour. Each of them focuses on a small subset of potential causes for the division of labour, deliberately ignoring all others. The first looks to genetic diversity among the individuals of the colony, finding that if genes are responsible for different thresholds of response to stimuli, and that those thresholds are randomly distributed among the population of the hive, then self-organization will result in a division of labour (Page & Mitchell, 1998). The second model assumes the same work algorithm for all individuals whereby work will be done if it seems needed, otherwise they will move on. If the nest architecture is such that individuals are born in the center of the nest, they will undertake different jobs throughout their lifetimes as they progress outward from their birthplace, meaning that the division of labour is related to the nest architecture (Tofts & Franks, 1992). The third model attributes the division instead to learning, coupled with asynchrony in the birth of new workers: each individual begins with the same learning algorithm, and learning is mediated by the individual's environment, which is dependent on their time of birth. These factors together change their behaviour over their lifetime, and lead to a division of labour (Deneubourg et al., 1987). These three models are merely a sampling of the possible explanations found in contemporary biology (Mitchell, 2003, p. 213).

When applied to a particular instance, Mitchell argues that these different models are likely to be integrated. Particular instances will typically not be the simple, idealized scenarios used as the basis for the models described: "the concrete explanatory situations on which we bring abstract models to bear are messy, perhaps unique products of historical contingencies and interacting, multiple causal factors" (Mitchell, 2003, p. 188). To account for all the causes of the division of labour in a given colony could require a combination of the self-organization models described earlier (and others), and their interactions. What's more, the multiplicity of actual divisions of labour in colonies means that the integration used for a given colony is not likely to work for another. For instance, in the case of honeybees, it is probable that all three models could play a role, since their genetic diversity, hive architecture and learning opportunities conform to the assumptions in all three models (Mitchell, 2003, p. 216). However, in the case of ant colonies the lack of genetic diversity will tend to exclude the genetic diversity model (p.217). To complicate matters even further, certain ant colonies exceptionally do have a significant amount of genetic diversity, implying that in these cases, genetic diversity may play a significant role (Boomsma et al., 1999 cited in Mitchell, 2003, p. 217). The result is that the models used to account for the division of labour vary from species to species, and even potentially from colony to colony.

3.2 Levels of abstraction: general and particular phenomena

3.2.1 Abstraction and idealization

Scientific explanations can target phenomena at different levels of granularity, going from particular instances of a phenomenon to general phenomena. At the finest grain, a 'particular instance' is understood here as an actual, presumably unique, "real case" of a phenomenon that has happened (Mitchell, 2003, p. 188). These particular instances are the result of a typically messy and complicated set of interrelated causal factors. Nevertheless, this complexity does not translate to a plurality of causal histories: “however complex, and however many contributing causes participated, there is only one causal history that, in fact, has generated a phenomenon to be explained” (Mitchell, 2002, p. 66). Any explanation that targets that particular instance will therefore attempt to capture that one causal history in all its messiness (Mitchell, 2009, p. 23). Thus, for explanatory purposes, “at the most concrete level in generating an explanation, a model may introduce all of the relevant features that uniquely characterize a given event” (Mitchell, 2003, p. 188). However, this specificity is done at the expense of the general applicability of the explanation, since the messiness implies a set of causes which, taken together, may be unique to that phenomenon.

Scientific explanations can also target more general phenomena, and in in focusing on such a large grain, lose the details of the particular instances which are encompassed. This is done through referring to a class of phenomena that resemble one another in some way. This approach overlooks certain aspects of the phenomena to be grouped together, such as certain properties or entities involved in particular instances. This is the case when talking about ‘the division of labour in social insects’, which overlooks the type of insect, labour, and division, to talk of all cases where something of the sort happens. These details are deliberately ignored through a form of abstraction, allowing for general applicability.

Abstraction need not only be applied to the explanandum phenomenon; it can also be applied to the explanans, but in a different way, yielding different results. This is done through a process often called idealization (Godfrey-Smith, 2009; Jones, 2005; Potochnik, 2017), which selectively looks at the effect of specific causes. In other words, confounding causes are experimentally controlled, leading to explanations which focus on the role of only a single, or a few, causes. This is not the only way that idealization can be carried out (see Frigg & Hartmann, 2020 for a comprehensive overview), but it is the way that Mitchell emphasizes for the sake of her argument. A model which ignores confounding factors in this way “accurately describes only ideal cases where those simplifying assumptions hold true, but only partially captures actual cases that are not ideal in this way” (Mitchell, 2002, p. 64). The resulting idealization is done deliberately at the expense of the identification of *all* possible—or actual—causes. There can therefore potentially be multiple idealizations for any given phenomenon since each idealization can focus on a different subset of causal factors. These multiple idealizations coexist through “a horizontal relationship of different questions addressed to the one and same phenomenon” (Mitchell, 2003, p. 189). The different questions emphasize different causes, leading to a multiplicity of idealized explanations or models (more on the difference between these two shortly).

Though Mitchell uses the terms “abstraction” and “idealization” apparently interchangeably, I will distinguish the two for clarity’s sake. The first kind of generalization, applied to the explanandum phenomenon, and which deliberately includes a broad range of phenomena by overlooking the details, I will call ‘abstraction’. The second, which deliberately excludes confounding causes, I will call ‘idealization’. As will be shown, these two ways of generalizing are what eventually lead Mitchell to make a case for pluralism.

3.2.2 Horizontal pluralism and integration

Mitchell’s argument for horizontal pluralism is that the general phenomenon created through abstraction can be explained through various idealizations. As more and more phenomena are subsumed under a single generalization, there can be more and more causes which account for it, leading to a proliferation of models, which each account for only one or a handful of causes each. Thus, according to Mitchell, in order to have a complete explanation of the general phenomenon, we will need a plurality of models to capture all the relevant possible causal histories; the coarse-grained explanandum phenomenon will need to call on the various models used to account for all the fine-grained explanandum phenomena included within it. Applied to the example of division of labour in social insects, this means that the multiple models surveyed will all capture some of the possible causes for the division of labour, and will be applicable to a subset of social insects. This also means that the models are not competing one with the other, since each of the models “describe[s] only what would happen in nonoverlapping ideal worlds” (Mitchell, 2003, p. 216). As such, they are in no position to compete as the *one true* explanation of the general, abstract phenomenon, since they explicitly and deliberately do not consider all possible causes. In this respect, at the general level, the plurality of models will remain since they all apply to the general phenomenon and are in no position to eliminate one another (Mitchell, 2003, p. 216). Because these explanations are all applicable to this same level of generality, Mitchell labels this defense of fragmentation pluralism a horizontal one.

On the other hand, according to Mitchell, the plurality of models is not warranted when applied to particular, concrete cases. Instead, “at the concrete explanatory level, […] integration is required” (Mitchell, 2003, p. 216). Thus, for any particular insect colony, it is possible that many of the models will be needed to account for the division of labour, and these models will need to be integrated in order to give a complete picture of the concrete phenomenon. For instance, as Mitchell remarks, honeybees are genetically diverse, have a concentrically structured hive, and individuals are born at different times: “thus what accounts for division of labour in the case of honeybees will be, presumably, a combination of genetic, learning, and architectural causal components” (Mitchell, 2003, p. 216). The assumption then is that at least the three models discussed earlier will be applied to the case of honeybees. At the concrete level, the models will need to be integrated, meaning that each partial explanation of the division of labour will need to somehow be coordinated with the others to explain all the contributing causes and how they interact (for more on “integration”, see chapter 5).

In Mitchell’s account, then, pluralism is warranted at the level of general phenomena, which will invariably call on multiple models (given sufficient complexity), but is rejected for particular phenomena. This implies that the integration which can be done at the concrete level precludes the possibility of pluralism for that level. In other words, integration is antithetical to pluralism at a given level of abstraction, since the phenomenon ends up being explained through a single explanation which gathers within it all the necessary models. At the general level, however, Mitchell claims that

> unification […] is unlikely to be very robust. The reason is found in the evolved complexity [EC – see Chapter 1] characterizing the domain of phenomena studied by biology. It is the diversity of the ‘solutions’ to adaptive problems and the historical contingencies influencing those variable paths that preclude global, theoretical unification. (Mitchell, 2002, p. 67)

Note that what Mitchell calls here “theoretical unification” is an account which would explain a general phenomenon, such as the division of labour in social insects, including all social insects, and all divisions of labour. But EC is such that the division of labour can be done in very many different ways, implying that any model which focuses on only a few causes of the division will be unable to explain all actual divisions of labour. This denial of the unification of the explanations only at the higher level but not at the lower level is why Mitchell’s pluralism is a fragmentation pluralism as opposed to an insular pluralism (see chapter 1).

In sum, according to Mitchell, there can exist a plurality of explanations for “what appeared at first sight to be the “same” phenomenon requiring a single explanation” (Mitchell, 2002, p. 67). It is in this respect that Mitchell can claim that the general phenomenon of division of labour in social insects admits of a plurality of explanations.

3.3 Do models explain a general phenomenon?

3.3.1 Targeting a general phenomenon

The fact that these multiple models exist cannot be called into question; but what is their relation to one another, and are they truly targeting the "same phenomenon"? This is a crucial element to the argument, since if it can be established that the explanations do not, in fact, target the same phenomenon, then there is no reason to posit explanatory pluralism: it would simply be a case of multiple explanations for just as many phenomena. This is a line of criticisms adopted by Ruphy (2013), which I will unpack in the following pages.

A first detail to address is that the difference between a particular and a general phenomenon is a relative one. Any general phenomenon can be particular with respect to an even more general phenomenon. And any particular phenomenon can be a general phenomenon of an even more particular instance. So for instance, we may be able to talk of the general phenomenon of division of labour in bees, a particular instance of which would be the division of labour in a specific bee species. On the other hand, it is a more particular instance of the division of labour in social insects, which would include not only bees but any social insect.[11] Thus all talk of 'a general phenomenon' is simply shorthand for 'such and such a phenomenon which is more general than the more specific instances which it subsumes'. With this in mind, we can turn to the thorny question of how generalizations are legitimized.

Let us begin with an absurd generalization simply to make a point. Suppose that we are presented with what are clearly two phenomena: the turning heads of sunflowers, and the turning heads of human spectators at a tennis match. It would come as no surprise that two different explanations are needed to account for these two phenomena, despite the very superficial similarity of heads being turned. Ruphy (2013, p. 177) adopts this tactic with respect to Mitchell's horizontal argument for pluralism. She contends that the pluralistic claim stems from an error in generalization: different social insects will warrant one—and only one—of the three explanations (assuming they exhaust the range of possible explanations), implying that each explanation in fact applies to different phenomena (presumably different species).

[11] There are of course limits: once the explanandum phenomenon is circumscribed to a specific event at a specific time, then it cannot be a generalization of anything else. The limit for generality however is not so clear; perhaps universal laws can be understood to apply at the most general level of abstraction: to the universe, or even to all possible universes. In any case, this is not the purview of biological explanations of behaviour, so we can safely return to our more concrete examples.

Consequently, the generalization of “division of labour in social insects” as a single (general) phenomenon was a mistake, since there is no single mechanistic explanation which accounts for it, despite the functional similarities in the particular phenomena. The superficial similarities of the phenomena simply do not warrant a legitimate scientific abstraction leading to a generalization. Just as one cannot abstract the general phenomenon ‘turning of heads’ from sunflowers and spectators at a tennis match, one cannot group together all division of labour in social insects if the various explanations are different.

What is required to validate Mitchell’s take on pluralism is a way to legitimize the generalization without relying on the superficial similarities of the particular instances. One way this could be done, as proposed by Ruphy, is simply by saying that an appropriate generalization is one which groups together phenomena which resemble one another in ways *relevant to the explanation sought*: that could be causal mechanisms, population distributions, a target for natural selection, or any other aspect which is in want of an explanation. In other words, the generalization is warranted precisely because it is possible to explain all the particular instances through a single explanation; and an explanation is seen as correct for a particular phenomenon if does indeed explain that phenomenon. Another way of understanding this idea is that if the generalization appears in laws or explanations that allow for sufficiently correct predictions or interventions (with criteria to be determined), then it is a legitimate generalization. Ruphy remarks that the circularity apparent in circumscribing the phenomenon through its explanation, and the recognition of proper explanations as targeting correct generalizations, is not problematic, and merely an indication that identifying appropriate generalizations invariably relies on our knowledge of those phenomena (2013, p. 177); generalisations once deemed legitimate may change as our understanding of the phenomenon increases (p. 178). Thus, according to Ruphy, Mitchell’s identification of the general phenomenon of division of labor in social insects is illegitimate precisely because of the plurality of explanations possible.

But this response is unsatisfying in at least one way. Certain generalizations *are* legitimate scientifically, despite being multiply explainable. This is similar to the cognitive science issue of multiple realizability, whereby it is taken for granted that a single mental state can be instantiated through multiple different brain states. In much the same way, a general phenomenon can be realized in a variety of ways, leading to multiple explanations of particular instances. These instances are legitimately similar in some respects but not in all the details. For instance, if one could say scientifically meaningful things about the division of labour in social insects in general, Ruphy would be hard-pressed to maintain that it is not a scientifically legitimate generalization. Judging by the quantity of scientific articles published about the (general) topic, it does seem as though this is a generalization recognized by the scientific community. Common questions relate to the difficulties in relating adaptationist and proximal mechanisms, and common themes are the multiple ways in which models relate sometimes only partially to actual particular instances (Beshers & Fewell, 2001; Duarte et al., 2011). Of note is that it seems entirely possible that the general phenomenon of division of labour in social insects could call on a single explanation of its phylogeny (see Duarte et al., 2011 for a proposal along these lines). Perhaps all social insects developed the division of labour due to similar selection pressures; whereas Ruphy emphasizes the multiple mechanistic proximal mechanisms, it is important to remember that (just as Tinbergen warned) there are other types of explanations which may apply to a different granularity, in this case ultimate explanations. If it is the case that there is an ultimate explanation for the division of labour in social insects, then Ruphy's idea that it is a bad generalization due to the fact that it calls on multiple explanations falls through.

We can furthermore note that Ruphy ignores the possibility raised by Mitchell of the integration of the various models, which muddies the waters of Ruphy's argument. Ruphy's approach works rather elegantly if each model applies only to a single case (or a single species), since the justification for the generalization then seems to rely only on spurious similarities. But if the multiple particular instances are all explained through integrations of the same pool of models, that in itself may be sufficient reason to justify the claim that they all resemble each other in significant ways. In other words, their explanations would not be so starkly independent from one another to warrant a sharp division between the phenomena.

It seems then that we have no strong arguments to doubt the scientific legitimacy of the general phenomenon of division of labour in social insects, despite not having strong arguments to validate it either. At the very least, we have good reason to suppose that there exist scientifically valid general phenomena which call on multiple models to explain the concrete, particular instances of the phenomenon—even in the unlikely event that it turns out that this does not apply to the division of labour in social insects. What is important to keep in mind, however, is that the generalization does deliberately group together phenomena which are known to be different in significant ways, and are known to be the product of various different causes. Mitchell's contention is that this general phenomenon requires multiple explanations, despite the fact that it looked like it was a single, unified phenomenon, which would "requir[e] a single explanation" (Mitchell, 2002, p. 67). But is claiming that there exist multiple models equivalent to claiming that there exist multiple explanations? In order to answer this question, we must first look at difference between models and explanations.

3.3.2 Models and explanations

Mitchell's writings sometimes refer to the multiplicity of models, and sometimes to the multiplicity of explanations, apparently somewhat interchangeably. Her argument for horizontal pluralism relies on the multiplicity of models accounting for a general phenomenon, and the idea that these models in some way *explain* that general phenomenon. There is, however, good reason to differentiate explanations and models, especially in the context of explanatory pluralism, where the issue is precisely to understand the relations between explanations targeting a given phenomenon.

There is a sizeable literature which concerns models in science, with debates raging regarding the ontology of models, how they relate to the world, and how they relate to explanations (Craver, 2006; Frigg & Hartmann, 2020; Hochstein, 2017; Morrison, 2011; Verreault-Julien, 2019; Weisberg, 2013). Mitchell does not address most of these issues, with her only explicit commitment being that models, assuming they have been empirically tested and shown to be valid, are "partial solutions to a biological question" (Mitchell, 2003, p. 217). By "partial," she means that they are idealizations, as described above: they push aside the multiple potential causes of a phenomenon to pick out only one or a handful of causes. Importantly for Mitchell's argument is that none of these models investigates all the causes together, implying that they are not subsumed under a single, overarching model. They furthermore "do not directly compete, since they describe only what would happen in non-overlapping ideal worlds" (Mitchell, 2002, p. 64). Since the models do not compete, no model will come to be seen as the one true model which pushes aside all the others. This means that the multiplicity of models seems to be here to stay.

Another of Mitchell's commitments, though this one implicit, is that models are explanatory. This is contrary to George Box's aphorism that 'essentially, all models are wrong, but some models are useful', but is in line with some contemporary research in philosophy. However, though the fact that at least some models explain is virtually uncontested, there is some debate about how this can be the case. Since models are idealized, it is not clear whether they "explain *despite* or *because* of the idealizations they involve" (Frigg & Hartmann, 2020 emphasis in original). Indeed, some argue that because models distort reality, they cannot be true; if models are false, and if explanations must furthermore be true, then there seems to be no reason to see models as explanatory (c.f. Reiss, 2012).[12] Others, however, have argued that idealized models can nevertheless be explanatory, either through the neglect of irrelevant factors (Elgin &

[12] Of note is that Longino's notion of 'conformation' seems like it would neatly resolve the debate since it does not rely on the models being true or false, but only more or less representative of the world as it is (see chapter 4).

Sober, 2002; Potochnik, 2017; Strevens, 2011; Weisberg, 2013), or precisely *because* they are idealized (Cartwright, 1983). Mitchell does not go in depth into this debate, though she does make explicit how models can be explanatory when applied to concrete cases. We shall see, however, that it is not entirely clear how models are meant to be explanatory of general phenomena.

Recall that for Mitchell, one of the roles of scientific explanations is to uncover the causes which brought about a phenomenon. In this respect, one fairly clear way that she understands models as explanatory is when they are applied to particular, concrete cases, shedding light on the causes that are at play in creating the division of labour in a specific case. This understanding of models is shared by some of the researchers working on the division of labour in social insects: for instance Beshers & Fewell (2001) see the various models they survey as being "explanatory" once "they can generate predictions that can be tested with experimental data" (p. 433). The idea here is that if they are not compared to empirical data, they remain in some sense speculative, and not explanatory. Thus, by applying them to concrete cases, the otherwise hypothetical models are revealed to be explanatory insofar as they can account for the observed reality. Models are therefore explanatory in Mitchell's sense, since they reveal the causal pathways accounting for a particular phenomenon.

However, when it comes to the general phenomenon, the explanatory role of models is understood in a looser way. The various models isolate one or a handful of causes which could account for certain particular instances, but not all of them. In this respect, they are "models of potentially contributing causes" (Mitchell, 2003, p. 217), offering only potential explanations of the general phenomenon, or what Mitchell calls "partial solutions" (p.217). For instance, the division of labour in social insects is potentially the result of genetic diversity, though for any particular instance, that may not be the case. Researchers working on the models see their work in this way themselves: having developed a new model, Beshers & Fewell point out that they have "expanded the range of possible explanations for division of labour" (2001, p. 431). Models are thus in a strange position, actually explaining certain particular cases, but only potentially explaining the general phenomenon.

Models have been compared with "how-possibly" explanations (see for instance Craver, 2006; Reiss, 2012; Reydon, 2012; Verreault-Julien, 2019). According to one influential understanding, these are explanations which describe only how a phenomenon could *possibly* be brought about (Hempel, 1965), contrasted with "how-actually" explanations, which identify the real causes (Craver, 2006).[13] How-possibly explanations are often put forward in the process of discovery, to narrow possibilities, or to suggest constraints on the how-actually explanation (Craver, 2007, p. 58). Because of this, it is expected that there will exist multiple how-possibly explanations, until the how-actually explanation is discovered. In fact, Craver described this process as a transition away from models: once we get from how-possibly to how-actually, the phenomenon is "not merely modeled but explained" (Craver, 2007, p. 58). In much the same way, the models put forward by Mitchell identify potential causes of a general phenomenon, and once they are applied to concrete cases and integrated (if need be) they are then transformed into how-actually explanations.

3.4 Horizontal pluralism as 'grocery list' pluralism

3.4.1 Models as items in a grocery list

In Mitchell's conception, then, models are seen as "partial" explanations of a general phenomenon. Because of this, she maintains that there is explanatory pluralism at the level of the general phenomenon:

> at the theoretical level [i.e. when applied to the general phenomenon], pluralism is sanctioned. Different idealized models do not directly refer to the same ideal systems. At the concrete explanatory level, on the other hand, integration is required. However complex, and however many contributing causes participated, there is only one causal history that, in fact, has generated a phenomenon to be explained. (Mitchell, 2002, p. 66)

But is it appropriate to call the multiplicity of possible explanations at the general level 'explanatory pluralism'? And if so, what sort of pluralism is it? The answers to these questions hinge on how, or to what extent, how-possibly explanations can be seen as explanatory of the general phenomenon. In this section, I will survey three possible interpretations for this claim, followed by a discussion of what this implies for explanatory pluralism. But first, I propose another phenomenon which could be the subject of multiple idealized models, allowing for a better understanding of what is at stake.

In order to better shed light on the interplay of the multiple models, let us compare the division of labour in social insects with another more familiar phenomenon. Suppose that we identify the general

[13] Though see Verreault-Julien (2019) for an extensive discussion regarding the different types of how-possibly explanations: while some authors see them as ways of showing that it is not impossible for a phenomenon to come about, others see them as potential explanations. Mitchell seems to adopt the latter conception for the role of models in biology.

phenomenon of ‘people going to the grocery store’, and we decide to investigate why people go there. We then construct many idealized models which could account for particular instances of people going to the grocery store. One model suggests that when people want to buy tomatoes, they go the grocery store. Another looks at celery as the cause of going to the grocery store, another at apples, and some even look at fruits and vegetables in general. Other models might identify that people go to the grocery store for work. All these models are idealized scenarios, which isolate a single cause or a handful of causes for going to the grocery store. For any particular, concrete case of a person going to the grocery store, it is likely that an integration of various models will be necessary, such as when a person goes to the grocery store to buy tomatoes *and* celery. Just as it is the case for the division of labour in social insects, we have, on the one hand, a general phenomenon which abstracts away from particular instances to encompass all humans and all grocery stores, and on the other hand, we have idealized models which potentially explain the general phenomenon, and actually explain particular cases when integrated. Granted, it is very unlikely that this phenomenon would be studied by biologists and as such should not be taken as a literal possibility. Even as far as scientific investigation in general is concerned, many details would need to be clarified: which people, which supermarkets, which timeframe, which experimental setup, etc.? The objective here is rather to use a very banal phenomenon to highlight certain issues with Mitchell’s account of horizontal pluralism. The important similarities for the analogy are the following: just as the division of labour in social insects covers many particular instances of groups or individuals acting for specific, sometimes overlapping causes, so the phenomenon of ‘people going to the grocery store’ covers many particular instances of groups or individuals acting for specific, sometimes overlapping causes. And if models can target specific causes and sometimes be integrated in order to do something like explain the general phenomenon, then the analogy holds. Now, the question to answer with respect to explanatory pluralism is whether, and in what respect, do the multiple models explain the general phenomenon. There are three possible interpretations.

3.4.2 Three interpretations of explanatory models

The first interpretation is that the models are not explanatory of the general phenomenon. Because they provide only a possible explanation, and not an actual explanation of the general phenomenon, they could be considered more simply as non-explanatory. This is intuitively plausible when one considers the grocery store example: if we were looking for an explanation of why people go to the grocery store, it is unlikely that anyone would consider any of the models in isolation as an explanation of the phenomenon. Though ‘wanting to buy tomatoes’ is a plausible explanation for a specific person to go to the grocery store, by no means can it be considered an explanation of all people going to all grocery stores at all times. Under this interpretation, then, the multiple models are not explanatory, and therefore there is no explanatory pluralism.

A second interpretation is that the models are explanatory of the general phenomenon, though only when taken all together. Under this interpretation, the models are partially explanatory of the general model, but can be understood as actually explanatory once all the partial explanations are considered at once. For instance, the general phenomenon of 'going to the grocery store' is explained by all the possible grocery lists. Thus, when asking 'why people go to the grocery store' in general, the explanation is that people have various grocery lists, and these lists, taken together, explain why people go to the grocery store. For any given person, there will be an integration of the various models (represented by items on the grocery list) to explain the phenomenon of that person going to the grocery store. But at the general level, there is no integration of the models since there is no attempt to relate the interaction of items (as is done in the third interpretation; see next paragraph), only a list of every item in the grocery store which could possibly be found on a grocery list. Under this interpretation, that list of every possible item on a grocery list is understood to be explanatory of the general phenomenon of people going to the grocery store. In the same way, the division of labour in social insects is explained through all the possible models which focus on specific causes, such as genes, learning mechanisms, nest architecture, etc. When taken all together, the models identify all the possible causes of the general phenomenon of division of labour in social insects. Considered in this way, it may be possible to claim that there is a form of explanatory pluralism insofar as the explanation for the general phenomenon does call on multiple models which are not integrated.

The third interpretation builds on the second, adding that the multiple models could be integrated, especially as science progresses. Indeed, researchers will not be satisfied with a grocery list of possible causes, working instead at understanding the interactions between the causes and weighing the impact of each. This is not a far cry from what Mitchell herself believes to be possible, considering her claim that the models can be integrated when applied to particular cases. The integration at the level of the general phenomenon could still be considered an explanation for the same reasons as those put forward in the previous paragraph. This proposal is similar to Thagard's (1998) Causal Network Instantiation in medical practice: medical researchers will produce a causal network identifying all the possible causes of a disease as well as their interactions and their respective weights, and will explain a particular case of a disease by instantiating the causal network, or in other words by identifying the relevant parts of the causal network for the patient in front of them. In this way we can see the integration of all the models as a unification of the explanations, where all models are used for the explanation, with the weight of each measured as a proportion of their use in all concrete cases. So, for instance, one could explain that a certain proportion of people go to the grocery store to buy tomatoes, whereas another proportion goes to buy celery. This could be a way of understanding Mitchell's claim that there is an expectation that a single explanation will be put forward to account for the general phenomenon. Though explanations of specific cases still call on particular instantiations (in Thagard's sense) of the causal network which incorporates all possible causes, they still all start from the same overarching, integrated explanation which includes all possible causes and interactions among the causes. Under this interpretation however, one could maintain that there is no explanatory pluralism, since the multiple models are all integrated into one overarching explanation for the general phenomenon.[14]

3.4.3 Grocery list pluralism

Of the three proposed interpretations, only the second makes room for any kind of explanatory pluralism. Whereas the first interpretation considers that the multiple models do not explain, the third proposes that they explain once integrated into a single explanation. The second interpretation does allow for some semblance of explanatory pluralism, though admittedly of a peculiar—and perhaps trivial—sort.

It is peculiar insofar as the multiple models which are the basis for the pluralism claim are not quite considered explanatory when taken in isolation, since they highlight only potential causes, therefore only ever *potentially* explaining the general phenomenon. And this is the important point to remember under this interpretation: though the individual models can be explanatory when applied to particular cases, they cannot be actually explanatory when applied to the general phenomenon, only potentially explanatory, which is to say: not explanatory. After all, in Mitchell's example, the idealizations are such that the models do apply to the division of labour in social insects, but only when certain conditions are met: whereas one model is applicable only when genetic diversity is present, another applies only in cases where the nest

[14] Another possibility relating models to general phenomena is to create a model (or explanation) which is just as general as the phenomenon: the general explanation for the division of labour of insects in general is that such a division is more efficient for resource and effort expenditure for the colony in general (Lucas & Ben-Shahar, 2021). Though this is an interesting proposal relating to the granularity of explanandum phenomena and explanans, I do not believe this is an issue for pluralism since there is only one explanation at stake. Pluralists might be tempted to add this to the list of possible models for the phenomenon, but it does not change Mitchell's argument in any substantive way.

architecture conforms to specific requirements. Each model specifies the circumstances necessary for the causes they highlight to actually be in play. It is due to the potential character of the explanatory power of models that Mitchell can claim that there is explanatory pluralism: the general phenomenon, by virtue of its generality will always be the result of multiple possible causes, implying multiple possible models. Of course the identification of a general phenomenon and its possible causes is a valuable contribution to scientific research in and of itself, but it is not actually explanatory of any given case.

And it is in this respect that this type of explanatory pluralism can be characterized as trivially true. After all, the substantive claim behind this interpretation is that for a given general phenomenon, there are multiple potential causes which can bring it about. But when a general phenomenon is demarcated through abstraction applied to the entities and outcomes that are involved, as is the case with the division of labour in social insects or humans going to the grocery store, it is unsurprising that many causes could be involved. And in this grocery-list interpretation of pluralism, all we have is a list of the possible causes, with no integration, and no attempt at placing them all into a causal network. Mitchell assumes that the expectation should be that such a phenomenon will be explained through a single, unified explanation, (which is somehow more unified than the Causal Instantiation Network mentioned earlier) but it is unclear who would entertain such an expectation, nor why. The assumption that the division of labour in social insects—just as the reasons for going to a grocery store—ought to be accounted for with a single explanation seems to simply be a mistake under this reading.

3.5 Concluding remarks

With these considerations in mind, we can finally tackle the main question: does Mitchell's horizontal approach to pluralism make a strong case for explanatory pluralism? In order to properly answer this question, we must evaluate three aspects: (1) whether we are indeed dealing with a single phenomenon; (2) whether there exist multiple explanations of that phenomenon; and (3) what the relation between the multiple explanations is.

With respect to (1) whether the division of labour in social insects is indeed a single phenomenon, and as previously discussed, the answer is not so easy to come by. The criteria for accepting the legitimacy of a general phenomenon in science are unclear, though we have explored some possibilities earlier. What we can say for certain is that general phenomena of the type Mitchell mobilizes are created through abstracting away from the particulars of actual cases, looking at categories of entities (such as social insects) and general behaviours or effects (such as the division of labour). Of course, not anything goes: one cannot simply create a general phenomenon by thinking about reality in a certain way, or categorize by fiat. The particular phenomena must be in some sufficient sense be similar, perhaps through similarity of function, of mechanism, or other such relevant characteristics (see Bogen & Woodward, 1988; Hacking, 1983; McAllister, 1997 for discussions on this topic). Furthermore, judging by publications of the topic, the division of labour in social insects is an object of research in biology, lending credence to its legitimacy, in part because of attempts to provide a single explanation of its phylogeny (see e.g. Duarte et al., 2011). It is important however to point out that, by definition, it encompasses many concrete phenomena, each with its distinct characteristics, entities, and outcomes.

As to (2), whether there exist multiple explanations, here again a clear answer is difficult to come by. It will depend on one's commitment to whether models of the sort described by Mitchell can be counted as explanations or not. As discussed, the most plausible interpretation, which is in line both with Mitchell's writings and with explanatory pluralism, is that the models potentially explain subsets of the general phenomenon, and actually explain it when taken all together. But they actually explain only through a (potentially very long) list of disjunctive possibilities: people go to the grocery store to buy tomatoes and/or celery and/or fruits and vegetables in general and/or… Social insects divide their labour due to genetic predispositions and/or nest architecture and/or learning and/or… In any case, what is important to note here is that to save the pluralism, it is necessary to view the multiplicity of models as *potential* explanations at the level of the general phenomenon.

What (1) and (2) imply with respect to the relations between the explanations (3) is that it is unsurprising that multiple models, each focusing on one or a few potential causes, could co-exist to explain a set of phenomena which are brought together because of superficial similarities. If this is what the horizontal approach to pluralism boils down to, it seems a far cry from the promises of explanatory pluralism as a novel and potentially revolutionary way of understanding science and its relation to the world. Instead, we are left with a 'grocery list' type of pluralism, which maintains that general phenomena can be the result of multiple potential causes, which can be isolated through models. Those models are idealized, which is what allows them to co-exist, but it is also what distances them from how-actually explanations. In virtue of idealizing the actual world, they remain models rather than explanations. It is *because* they idealize the actual world that their multiplicity is made possible, but it is also because of this idealization that the plurality is trivially true. By ignoring causal factors, models both do not actually explain, and make room for different models which ignore other causal factors. The plurality boils down to a list of possible causes for a given phenomenon.

In conclusion, the horizontal approach to fragmentation pluralism is at best trivially true, claiming that multiple potential explanations can co-exist when applied to a general phenomenon, and at worst, is an erroneous conclusion based on an overstatement of the explanatory power of scientific models. However, one central concept which has been left largely unspecified up to this point is 'integration'. It is of central importance both to insular and fragmentation pluralists, since in both cases, the claim is that integration is limited in its application, leading to various forms of pluralism. In chapter 5, different integrative strategies will be surveyed, and examples of successful integrations in biological explanations of behaviour will be presented in order to better understand what integration is, and how it is done. Before moving on to integration however, we will look at a more forceful version of explanatory pluralism in the next chapter: Longino's insular pluralism.

CHAPTER 4

LONGINO'S INSULAR PLURALISM

Helen Longino has developed a form of explanatory pluralism over the course of a few decades, laid out in a few different works, namely *The Fate of Knowledge* (2002), the introduction to *Scientific Pluralism* (2006b with Kellert and Waters), and most recently *Studying Human Behavior* (2013). The last of these is particularly interesting with respect to this thesis, being Longino's latest and best case for insular pluralism, as well as focusing on biological explanations of behaviour. Being insular, the pluralism she defends is stronger than Mitchell's. The latter's fragmentation pluralism denies the possibility of unity of science, but allows, and even encourages, integration of concrete explanations. While Longino also denies the possibility of unity of science, she goes further in highlighting the barriers to integration, and furthermore proposes that lack of integration across scientific explanations is not a problem, perhaps even a boon. Indeed, Longino criticizes Mitchell's pluralism as being "weak in holding that for any causal process there is a uniquely correct causal story" (2002, p. 199), and for clinging to monist intuitions about the possibility of integration: "this presupposes a commensurability that may not obtain in all cases, and does not obtain in the case of behavior" (2013, p. 147). Thus, Longino's pluralism is more forceful, allowing for even more disunity, and is defended in different ways.

This chapter will describe Longino's insular pluralism (section 4.1), then critically assess the three building blocks of her arguments (sections 4.2, 4.3 and 4.4), and end by raising some questions about the possibility of insular pluralism (section 4.5).

4.1 Longino's pluralism

Longino's explanatory pluralism is a form of insular pluralism, which proposes that multiple scientific approaches can put forward their own explanations for a given phenomenon. According to her approach, it is entirely possible that those explanations will in some sense be incompatible, and that the incompatibility will not be seen as a problem to be resolved, but instead as a success of science. As Longino puts it:

> Strong or substantial forms of pluralism hold that there are some phenomena and investigative contexts characterized by an ineliminable plurality of theories, models, or hypotheses, and that this situation should not be judged a failure, but should be understood and incorporated into philosophers' understanding of scientific success. (2013, p. 137)

This passage highlights the two-fold nature of Longino's scientific pluralism: it is both descriptive and prescriptive. She and other pluralists argue that whether pluralism or monism is the best view of science is "an empirical question" (Kellert et al., 2006b, p. xii; see also Longino, 2013, p. 144), underlining how they believe that pluralism is best seen as a descriptive enterprise. Yet at the same time, they also urge philosophers and scientists alike to adopt the "pluralist stance", an attitude which puts to the fore the diversity of approaches and explanations in the sciences, and propose that we should see such a diversity in a positive light, which in turn avoids "senseless controversies that do not lead to progress" (Kellert et al., 2006b, pp. xii–xv). Thus, Longino's description of science is such that there exists an ineliminable plurality of explanations targeting specific phenomena, and that this plurality should be judged favourably.

With all this in mind, what is needed to vindicate firstly the descriptive aspect of Longino's insular pluralism is to show that it is possible for good scientific explanations to target the same phenomenon and be incompatible. And secondly, with respect to the prescriptive aspect, it must be shown that there is good reason to think that the incompatibility is permanent. This permanence of incompatibility is important both to distinguish insular pluralism from the mere temporary plurality of explanations, and to justify the shift to a pluralist stance regarding the multiplicity of conflicting explanations. There are therefore three major building blocks to Longino's insular pluralism:

> Insular pluralism: multiple scientific explanations can target a single phenomenon and be permanently incompatible, and this is a positive aspect of science.

I. Single target: the multiple explanations must target a single phenomenon.
II. Incompatibility: the multiple explanations must be incompatible.
III. Incommensurability: the incompatibility must not be resolvable.

These three aspects of insular pluralism will be tackled in turn in the following sections, showing both that they are necessary, as well as what Longino puts forward with respect to each. I will also raise certain complications which arise from these arguments. As will become clear, I tend to agree for the most part with what Longino proposes, and I believe that there is great value in her ideas. More specifically, her lucid understanding of what constitutes a single phenomenon, as well as her analysis of how incompatibilities can occur among scientific explanations are both insightful and useful for my own integrative monism, as will be shown in chapter 7. Where we diverge is in her proposal that incommensurability can occur, leading to permanent incompatibilities, an idea which I will argue is wrongheaded throughout the remaining chapters of this thesis. But before delving into criticisms, let us first look at the three building blocks for Longino's insular pluralism.

4.2 Identifying one and only one phenomenon

A necessary element for any pluralist argument is to show that multiple explanations do indeed target a single phenomenon. In a sense, this echoes the difficulties seen in the previous chapter regarding the classification of general phenomena, but at an even more concrete scale, namely the identification and measurement of specific, concrete phenomena, especially across research and measurement contexts. This first step is to ensure that the explanations truly are targeting the same phenomenon, otherwise we would be faced with a multiplicity of explanations merely targeting a multiplicity of different phenomena. For instance, Longino (2001) describes the myriad ways that aggression can be defined and measured, and points out that, supposing that a genetic and an environmental account explain a behaviour in incompatible ways, "unless [they] are accounts of the identical phenomenon, it's not clear what it means for them to constitute competitive accounts" (p.695). Thus, an apparent plurality of explanations about aggressive behaviour may turn out to instead be explanations about slightly different phenomena, dissolving the apparent incompatibility among explanations, all the while opening doors to future research which could explore the differences among the explanations and among the phenomena, and their repercussions. Adequately isolating what counts as "one" behaviour phenomenon is therefore a necessary step for any claim to explanatory pluralism.

The identification of a single explanandum phenomenon is an obvious first step in principle but can be a difficult one in practice, especially across studies, labs and researchers. In practice, the justification for the classification of behaviours is often left unexplained. In this section I will dig a little deeper than Longino into this aspect, to look at three steps that are involved in circumscribing a behaviour in order to use it as an explanandum phenomenon in a biological explanation: individuation, operationalization and measurement. These are here presented as if they were three chronological steps in the process, but as will be shown later, the reality is that they all depend in some measure on one another, and the actual process of defining a specific behaviour calls on all three processes in parallel.

4.2.1 Individuating "one" behaviour

Individuating various phenomena into classifiable behaviours is a more difficult task than one may imagine. Individuation is often an implicit part of the research that relates to the assumed granularity of behaviours.

For example, in Class et al.’s (2014) article about handling behaviour in wild passerines, “biting, pecking, flapping its wings” are all understood as being an “aggressive response” (p.429). Those three types of movement are clearly morphologically quite different, but are nevertheless understood to exemplify a single behaviour. Though it is intuitively apparent that this lumping together of movements into a single type of behaviour is not arbitrary, it is interesting to note that no justification is explicitly given. It is very likely that a justification does exist, whether it is predicated on previous studies establishing a relevant correlation between the three types of movement, or on the field expertise of the researchers, or any number of legitimate processes. My point is twofold: first, this is not typically something that is deeply considered or reconsidered every time biologists carry out research of this sort. It is to some extent taken for granted that all these steps in identifying a phenomenon work, or have worked in previous research, and we don’t need to overly worry ourselves about it. And second, because it is not thoroughly considered every time, this opens the door to possible misalignment between studies, and potential difficulties for comparisons and external validity, though this is no doubt relatively rare in rigorous, published research. I am not however suggesting that the individuation is merely intuitive, arbitrary, or subjective, only that the objective criteria are not easy to come by, and are not merely a question of looking at an animal’s behaviour: some work needs to be done.

The first challenge is to look at the activity of the entity under study and separate the various phenomena into more-or-less discrete units of behaviour. In order to target a behaviour for explanation, it must be what Rose (2001) calls “reified”, abstracting away from the messy reality of the dynamic variety of particular movements to “convert […] a dynamic process into a static phenomenon, a phenotype” (p.418). Reification is thus a two-step process: first, the messy and continuous stream of movements of an agent is cut up into parts which are each supposed to represent some coherent unit. This sort of delineation can be hierarchical, with certain behaviours counted as a subset of others, as in the case of “biting, pecking, flapping its wings” all falling under “aggressive response”: each of the three movements can be considered a behaviour in its own right, despite the fact that aggressive response as well is considered to be a single behaviour, and in many cases its granularity is likely more suited for use in research.

It has been pointed out that folk psychology is often the unofficial starting point for this “cutting up” of movements. The researchers’ understanding of a behaviour is colloquially explained through the intentional stance thought to underlie the movements of organisms (C. Allen & Bekoff, 1997; Weber, 2012). For instance, a bird’s pecking and biting would be perceived as being the result of the beliefs and desires which lead to aggressive behaviour. Although this may seem too unsystematic for serious scientific research, as well as raising problems relating to the unknowability of the mental states, it may be a necessary first step to begin making sense of classifications of behaviour, allowing preliminary answers to questions such as 'what is the animal/person/organism/entity trying to do?" Moving away from informal ways of individuation, some philosophers have argued that behaviours are classified by their function, some emphasizing evolutionary function (Griffiths, 2009; Rosenberg, 2006), others a biological version of Cummins-style functions, which take into account the role of the behaviour within a larger system (Weber, 2012). Both these approaches are challenging to operationalize insofar as they appeal to only indirectly observable properties, namely functions. Furthermore, as will be shown in the next subsection, these proposals, although not entirely wrong, give an incomplete picture of the possible ways of classifying behaviours.

The second step of the reification process involves using the individual behaviour thus cut out of the stream of movements for comparisons across contexts, individuals, and species, thus allowing generalizations. These generalizations in turn lead to the foundations of biological work into patterns of behaviour in an individual (Sih et al., 2004), identifications of homologies in behaviour (Ereshefsky, 2007), and comparative behavioural biology in general. Of course, any assertion of similarity is limited: drawing on work of zoologist Robert Hinde (1970), Allen & Bekoff note that "some lumping together of distinguishable actions is essential, because each action is probably unique in its exact form" (1997, p. 47). In the example of Class et al.'s (2014) research on wild passerines, each individual bird no doubt acted in slightly different ways when it came to biting, pecking or flapping wings (some biting more than pecking, some doing so in slightly different ways, etc.) but those individual differences are glossed over to allow for more broad measurement of 'aggressive response', and comparisons across individuals. Though it is not the case in Class et al.'s (2014) research, reification is potentially problematic if it leads to what Rose calls "arbitrary agglomeration" (2001, p. 418), whereby phenomena are lumped together that ought instead to be considered separate. More broadly, what is problematic in virtually all of the cases is that the methodology for individuation is implicit and unjustified, raising questions about not only its legitimacy but even how to judge the legitimacy in the first place.

Making explicit the implicit methodologies and gaining a clearer understanding of the individuation of behaviours raises the familiar philosophical questions of classification in the life sciences and in sciences more generally. What makes an individuation legitimate, or more legitimate than another? Are the different behaviours assumed to be natural kinds, or do they obey some other, perhaps more pragmatic classificatory practice? As is shown in the next section, the individuation of behaviours goes hand in hand with their operationalization.

4.2.2 Operationalizing a behaviour

Very similar to individuating behaviours, researchers must also operationalize them in order to recognize them in the world and thus classify them accordingly. Whereas individuation is more of an implicit, intuitive assumption, operationalization is an explicit, systematic endeavour, which attempts to describe a behaviour such that it can be recognized and therefore classified through observation. But when looking at phenomena that are to be classified as a given behaviour, what similarities are we looking for? This question is not so simple to answer. Compare these three scenarios: (i) a bird flaps its wings when handled by a researcher; (ii) a bird flaps its wings after a rainfall; (iii) a bird pecks and bites when handled by a researcher. In some respects, (i) and (ii) are similar because the movements themselves (flapping wings) are identical, but in other respects (i) and (iii) are similar because of the apparent objective—assumed through the context—of the behaviour. These problems become even more significant when the comparisons are across vastly different species with different morphologies and ecosystems.

Ethologists are perhaps the most systematic among biologists in operationalizing behaviour thanks to the ethograms they produce. An ethogram is "a behavioral catalog that presents information about an action's morphology and gives the action a name." (C. Allen & Bekoff, 1997, p. 40) There is however some disagreement regarding whether the action's description ought to be contextual or acontextual. Some ethologists claim that behaviours ought to be classified contextually, referring to either the function, consequence, immediate causation, or history of the behaviour under observation (Hinde, 1970). This approach thus widens the possible descriptions for a behaviour looking beyond the mere descriptions of movements. Other ethologists however argue that ethograms should instead classify behaviours acontextually, describing only the movements themselves, thus avoiding potentially faulty interpretations (Golani, 1992). Indeed, contextual reporting relies on the expectation that different researchers or observers will infer the same objectives or causes for the action, needing to extract relevant contextual information from a string of phenomena, an optimistic proposition at best. Acontextual reporting on the other hand runs the risk of lumping together similar-looking movements that in fact play very different roles for the organism. For instance, in the examples (i) and (ii) earlier, although the morphology of the action is identical (flapping wings), most biological explanations will see the action as different (aggressive response vs. drying wings), except perhaps the most proximal ones which will focus only on the specific physiological and neurological events which immediately precede the action. Contextual reporting on the other hand will classify (i) and (iii) as belonging to the same type, as no doubt Class et al. (2014) would do, since biting pecking and wing-flapping are all conceptualized as part of the overarching "aggressive response".

As has been shown in this section, operationalization and classification go hand in hand. If the classification of behaviours is contextual, the operationalization will be done in terms of context, history, function, etc. If, on the other hand, the classification is acontextual, the operationalization will refer only to specific bodily movements. Conversely, if the operationalization is based on the context, history, function, etc. then it will more easily lead to a contextual classification, whereas operationalization by bodily movements will lead to acontextual classification. Thus, these choices have direct repercussions on how behaviours are conceptualized, and how they can be compared within and across time, individuals, and species. In practice, biologists will tend to move back and forth between these kinds of conceptualisations in their field notes, as well as use more acontextual reporting in figures, reserving the contextual interpretation of behaviours for the discussion sections of their papers. Needless to say, this variety potentially complicates discussions between researchers who may not have the same criteria for the operationalization of behaviours.

4.2.3 Measuring behaviour

The previous steps eventually allow for the quantification of individual behaviours, leading to the measurement of the phenomenon. Measurement can come in many forms and with various objectives, and it is interesting to note that Class et al.'s (2014) measurement scheme is in fact relatively rare because it is has been developed relatively recently; more traditional ethological measurements relied on frequency and timing of particular behaviours over the course of specified timeframes (Réale, personal communication). In any case, Class and colleague's measurement scheme will serve as an illustration of the relevant point. First, the measurements can be either along a continuous scale, be discrete, or categorical. For instance, in Class et al.'s (2014) research, aggressive response is measured on a scale of 1 to 5, with the tacit understanding that it is a somewhat arbitrary cutting up of a gradual scale. This example also demonstrates that measurement along such a scale may also entail the researcher's somewhat subjective evaluation of the behaviour since there are no clear indications of the difference between each of the five gradations—though intersubjective evaluations are generally identical. Measurements can also be discrete, as is the case with breath rate, yielding a number of breaths per second (Class et al., 2014, p. 429). In this second case, each breath is a counted as a discrete unit. Categorical measurements are also used, in cases where the behaviour is either present or not, for instance when measuring whether or not a bird has constructed a nest.

Second, the measurement can either attempt to measure directly the behaviour, or through various proxies. For instance, aggression in birds can be measured through direct observation of pecking, biting or wing-flapping, but in the case of humans, it is sometimes measured through the conviction of a violent crime, a proxy measurement (Longino, 2001). Research on humans furthermore allows for measurements otherwise unavailable in the case of other organisms, such as self-report, projective tests or interview measures (Suris et al., 2004). These methods therefore use proxies (i.e. the post-facto communication or answers to questionnaires) for the measurement, which rely on the lucidity, integrity and honesty of the person communicating the self-report for correct measurement. The use of proxies can therefore be problematic for accuracy, but it need not be: for instance, in the case of birds building nests, the presence of a nest could be considered an accurate measurement, without the need for observation of the actions that lead to the construction of the nest.

Third, when the measured behaviour is understood to contain many other behaviours, that subordination relation can lead to two different ways of counting. In one case, any of the subordinate behaviours count towards the overarching behaviour. Class et al.'s (2014) research illustrate this situation, whereby any and all of three behaviours (pecking, biting, wing-flapping) count towards an aggressive response. In the second case, the subordinate behaviours do not independently count towards the overarching behaviour; instead it is their conjunction that is important. This is the case of nest-building, where the individual actions that lead towards having a nest only count as nest-building insofar as a nest is built.[15]

And finally, the purpose of the measurement can be either to measure the behaviour of a specific individual, or to measure the distribution or average of a behaviour in a certain population. When measuring a population, sampling becomes an important methodological issue since the size and choice of the sample can have effects on the results. A sample size which is too small relative to the studied

[15] Or at the very least there needs to be the "intention" to build the nest, regardless of whether the behaviour is completed or not. If a bird is in the process of building a nest and dies, it still was exhibiting nest-building behaviour. However, if all it does is move twigs around, that behaviour does not count as nest-building. Needless to say, problems may arise when trying to differentiate these two scenarios in practice.

population obviously can lead to extrapolation errors. But errors pertaining to the source of the sample can be more difficult to uncover. For instance, certain earlier studies on aggressive behaviour in humans relied exclusively or almost exclusively on prison populations, giving erroneous generalizations to the population at large (Longino, 2001; Suris et al., 2004).

The biggest concerns when it comes to measurement is of course accuracy and precision. As is clear from the above discussion, measurement of behaviour is not always a straightforward affair. Decisions regarding how to place behaviours on a scale, which proxies or subordinate behaviours count, and in which ways, and finally typical sampling pitfalls all contribute to possible lack of accuracy and/or precision in measurement. It should be made clear however that these are not generally major methodological problems for studies taken individually; in other words, any single study can use its own measurement scheme with relative impunity as long as it is internally consistent, and that the conclusions do not overstep the bounds of the premises and methods. However, when it comes to comparing different studies to draw more general or stronger conclusions, the differences in measurement schemes can result in incommensurability, which is particularly problematic when unacknowledged or glossed over.

4.2.4 Bringing it all together

Looking closely at behaviour as an explanandum phenomenon and how it is characterized and measured raises two interrelated points. The first is that although the individuation, operationalization and measurement of behaviours can be conceptually separated, that separation is hard to carry over to practice. Indeed, the individuation of behaviours shapes the target of operationalization and measurement. Operationalization, for its part, can act as a framework for individuation, as was shown when comparing contextual and acontextual ethograms, and has a direct impact on measurement, defining what phenomenon counts as a unit to measure. Finally, measurement can determine individuation and operationalization by precluding behaviours which are unmeasurable or measured only with difficulty.

The second point relates to the interdependence of all these steps. Because of the difficulty in systematizing the process, it can be challenging to legitimize the choices made by the researchers to cut the blooming buzzing confusion of organisms in motion into distinct and measurable behaviours. If each step is done without proper attention to detail and especially if the justifications for the steps are implicit and rely on intuitions, even small errors could compound and result in an explanandum phenomenon that fails to circumscribe a legitimate object of study. Indeed, the danger is for the behaviour to be classified by fiat, leading to an explanandum phenomenon that defies explanation, or worse yet, that leads to erroneous conclusions. A further difficulty which has been raised previously is that comparisons may be compromised. When it comes time to compare individuals or species through particular studies, if the individuation, operationalization and measurement are done in different ways, the result can be research that seems to be addressing the same phenomenon when in fact they are talking past each other.

Research on aggression is a striking example of these potential problems. Aggression is a notoriously difficult concept to individuate. Aggressive behaviour is typically assumed to be a “multifaceted construct that may be expressed behaviorally in a myriad of ways” (Parrott & Giancola, 2007, p. 281). Operationalization carries over this conceptual vagueness; in the case of research on humans for instance, various factors are taken to indicate aggression, including verbal outbursts, physical violence, anger or irritability, or specific actions in a controlled laboratory setting (Longino, 2001; Suris et al., 2004). These can be measured through self-report, projective tests, behavioural laboratory measures or interview measures (Suris et al., 2004). But not only are there a plurality of measurement methods, the scales and units of measurement themselves are legion: Suris et al. (2004) provide an overview of no less than 60 formal scales which allow measurement of aggression in humans, a list they furthermore point out is not exhaustive (p.221). It goes without saying that there are many more measurement schemes that apply to aggressive behaviour in animals, as has been suggested in the earlier parts of this chapter. This "kaleidoscopic patchwork" (Longino, 2001, p. 694) of concepts and measurements implies that it is difficult to be sure that the various research programs really are about a single phenomenon. The limits, identification and measurement of aggressive behaviour are therefore unclear and difficult to compare one with the other, and it is easy to imagine that this same problem is present for many other phenomena.

4.2.5 Comparisons and external validity remain possible

Though these are serious issues with which to contend, they must nevertheless be kept in perspective: researchers can and do utilise different operationalisations and measurements schemes all the while targeting a similar enough phenomenon to be able to compare and contrast results.

This is an important point to make because it may be tempting to jump to the conclusion that any difference in measurement or operationalisation implies that researchers are working on different phenomena. This is a point that Sullivan (2009) makes, practically leading to the denial of any possibility of comparison across contexts. Sullivan proposes an anti-reduction and anti-integration argument which is predicated on these very problems of the operationalization and measurement of phenomena. She puts forward what is essentially one and the same argument against both Bickle's (2003) "ruthless" reductionism and Craver's (2007) "mosaic unity" through integration: if the multiple labs that are researching what is thought to be the same phenomenon are in fact using different experimental protocols, then they are not measuring the same thing, and therefore all possibility of comparing results across labs is doomed. Indeed, according to Sullivan, "unless the behavioral protocol and intervention techniques used were identical", then all that we can say when faced with laboratory results is that the "interventions undertaken by an investigator (or multiple investigators) directly explain the behavioral data observed by *that* investigator in *that* laboratory" (2009, p. 518 emphasis in original). This measurement problem then transfers over to the more fundamental problem of isolation of a phenomenon: faced with the multiplicity of experimental protocols in neuroscience labs, Sullivan proposes that "it is not clear that neuroscientists working within the same field are even talking about the same phenomenon." (p.526) This proposal of a plurality of phenomena thus raises the same problems covered in this section, which Sullivan (erroneously) suggests could be a compelling argument for pluralism.[16]

Ironically, though this argument may seem at first glance to corroborate a form of pluralism, it is better understood as antithetical to Longino's approach. This is because while Longino's pluralism is conditional on the multiplicity of *explanantia* for a given phenomenon, Sullivan's approach relies on the multiplicity of *explanandum phenomena*. Longino, for her part, rightly points out the challenges inherent to isolating a given phenomenon, but allows for leeway in the identification and definition of the phenomenon, which in turn allows for situations where a given phenomenon is targeted by multiple explanations (as will become clear in section 4.3). Sullivan's proposal, in contrast, proposes criteria that are so strict as to seemingly make impossible any generalization or external validity to results obtained through an experimental design. By her lights, every slight difference between experimental protocols implies the targeting of a different phenomenon, making links with other labs virtually impossible, and links to the more general phenomenon the experiment is attempting to explain tenuous at best. Outside of the scope of the specific laboratory setting and experimental protocol, it is "an open question" (Sullivan, 2009, pp. 519–522) whether we can say anything of note. Thus, rather than highlighting the multiplicity of

[16] She does later develop her 'coordinated pluralism' (Sullivan, 2017), which values the integration of methods and results from various laboratories, and proposes specific ways to do so, contradicting to some extent her strong conclusion that identical measurement setups are a *sine qua non* to comparisons.

explanations, Sullivan's approach instead implies that insular pluralism is mere illusion predicated on an actual multiplicity of phenomena. If we accept Sullivan's critique, we can thus conclude that while Longino may *think* we are faced with a plurality of explanations, it is in fact a plurality of phenomena.

Now, though it is true that the issues of identification, operationalization and measurement of phenomena pose some challenges, those difficulties should not be overstated, at the risk of denying any possibility of experimental science explaining anything but their own laboratory environment. It is not the case that any difference at all in any of the three steps laid out in this section will lead to significant differences in the isolation of a behaviour, or of a phenomenon more generally. If this were the case, then there would be virtually no possibility of communication across labs, and, perhaps even more importantly, no possibility of applying the findings in the lab to real-world contexts. Science would be limited to explaining specific events in specific labs, a prospect which would make science of little to no interest to anyone.

Clearly, Longino's approach is more reasonable: some leeway in definitions, operationalization and measurement is permitted without destroying all hope of comparison between results. The practical application of experimental results in the real world, as well as the progress in understanding of phenomena through the pooling of findings by different laboratories (Craver, 2007) show that overly stringent requirements on the classification of explanandum phenomena does not reflect reality. Furthermore, the mere *existence* of meta-analyses—though rife with their own potential difficulties and biases (see e.g. Eysenck, 1994; Greco et al., 2013; Metelli & Chaimani, 2020)—shows that the differences in experimental protocols and labs does not imply that each lab is working on their very own phenomenon; some comparison and cross-checking is indeed possible, and can yield fruitful results. To be sure, the question of just *how much* leeway is permitted is a vexed one, and has been the subject of debate among philosophers of science probing the ways external validity can be determined (Cartwright, 2009; Guala, 2003; Reiss, 2015, 2019; Steel, 2008). What is important for the sake of Longino's pluralism (and indeed for the sake of my own monism) is that though there may be disagreements about the ways in which external validity is to be understood, there is no disagreement that external validity and cross-laboratory comparisons are possible and legitimate. Different researchers can and do target the same phenomenon, which is the first step in identifying whether or not incompatibility occurs between explanations.

4.3 Incompatible explanations

The second step is to show that the explanations are incompatible in one way or another. This is necessary since if on the contrary they are compatible, then monists can rejoice, since there is no conflict in need to resolution, and it can be seen as a first step towards the unification of the diverse explanations, and perhaps even the end result of a unified science. For instance, reductionists—taken as anti-pluralists par excellence—will want to show how explanations at various levels are all compatible, by showing how they all explain the same causal functions (see Kim's approach covered in chapter 2). In contrast, insular pluralists want to show how there exist "multiple irreducible models or explanations" (Kellert et al., 2006b, p. xiv) for at least certain phenomena (though not necessarily for *every* phenomenon). What this amounts to is that they

> do not assume that the plurality of accounts should be consistent, that all truths from one accepted account must be translatable into truths of the other accepted accounts. Perhaps the approaches and accounts within the plurality cannot be combined and perhaps they even disagree with one another about certain points. (*Ibid.*)

Of course, finding incompatible scientific explanations is not the biggest challenge, since essentially all parties—pluralists and monist alike—agree that at any given time there can be multiple hypotheses or models which attempt to account for a given phenomenon. A monist would consider the incompatibility of explanations merely the result of science running its course, and entertaining multiple hypotheses, most of which are simply wrong, until the one true explanation is identified. Thus, what the pluralist needs to show over and above the incompatibility itself is that this plurality is not merely the result of the multiplicity of how-possibly explanations, but instead a normal consequence of different approaches putting forward correct explanations. So, once it has been shown that the multiple explanations are indeed targeting the same phenomenon—or close enough—, it must be shown that the explanations are also incompatible with one another.

The most obvious cases of incompatible explanations are those where different explanations contradict each other explicitly regarding specific and easily comparable elements of an explanation. This can be the case when two evolutionary explanations propose different selection forces for the same trait in the same species (c.f. Alcock, 1987; S. J. Gould, 1987; Mitchell, 1992 regarding the evolutionary origins of clitoral orgasms), or when two explanations in neurobiology put forward different neurological mechanisms to account for the same behaviour (Parmigiani et al., 1999). The more difficult cases are those where the contradiction is implicit, and apparent only insofar as the explanations are not compatible, though they are not directly contradicting one another. Cases such as these can be found in debates between genetic approaches and social-environmental ones, where geneticists accuse the other side of overstating the influence of the environmental causes (Longino, 2013, pp. 46–47); while geneticists are not in a position to argue that the environmental researchers are wrong in their explanation, they can propose that not taking into account genetics can lead to an overestimation of certain environmental effects. Because the explanations from each approach does not directly contradict the other (as would be the case with competing evolutionary accounts), the incompatibility is only implicit. Other examples can include explanations at different levels of organisation, where each purports to explain the phenomenon in ways that make the explanation at the other level either impossible or its importance overstated (Mitchell, 2002).

As previously mentioned, finding incompatible explanations is not the difficult part, since science proceeds through the testing of hypotheses about a given phenomenon, and at any given time there can be multiple hypotheses which are entertained simultaneously. It is no surprise that the history of science is rife with conflict and competition among various approaches. The real challenge for pluralists such as Longino is in showing *why* such an incompatibility occurs such that it can be interpreted as not being a problem for scientific goals, and not being a problem in need of resolution.

In view of such an objective, the greater part of Longino's (2013) book is a careful examination of five different approaches to studying the proximal mechanisms behind human behaviour, spanning from genes up to environmental causes: quantitative behavioural genetics, molecular behavioural genetics, neurobiological approaches, social-environmental approaches, and finally integrative approaches. For each of these approaches, she teases apart the different methods, scopes and assumptions which are explicitly or implicitly used or held by the researchers involved in the discipline. To give a sample of the work done by Longino, three of those approaches will be summarized in the following sections: quantitative behavioural genetics, molecular behavioural genetics, and social-environmental approaches.

4.3.1 Quantitative behavioural genetics: methods, scope and assumptions

The field of quantitative behavioural genetics analyzes the variation in behaviour across a given population, and tries to sort how much of that variation is due to genetic or environmental factors. The distinction between genes and environment is further refined by separating additive and nonadditive genetic variance, shared and non-shared environmental variance, as well as calculating gene-environment interactions (Gregory et al., 2011). After plotting the various phenotypes for a given behaviour observed in a population, researchers try to answer the question “What does genetic variation in the population contribute to this behavioural variation?” (Longino, 2013, p. 25). Their research however does not in fact look at the genetic sequences of individuals in the population under study; instead, for research on humans, it typically looks at twin studies and adoption studies in simple heritability studies. Monozygotic twins (with identical genes) and dizygotic twins (with, on average, 50% of shared genes) are looked at in shared environments (reared together) or nonshared environments (when one twin is adopted) to show to what extent their behaviour varies in concordance with genetic and environmental factors (e.g. Bouchard, 2004; Bouchard, Lykken, McGue, Segal, & Tellegen, 1990). Behavioural geneticists also use longitudinal studies to look for changes over time, as well as multivariate analysis to look for covariance of traits (Longino, 2013, p. 29).

What is important to realize, and which virtually all behavioural geneticists readily admit, is that the results of behavioural genetic research are far more limited than is intuitively apparent (and often portrayed in articles or books aimed at the layperson). First, the analysis of variation does not give information regarding *how* the genetic or environmental factors produce the behaviour, but only *how much* they contribute to the variance observed for that behaviour in that population (Sober, 2001). Second, the results apply only to the specific genetic baggage of the population under study (which is unknown except for the extent of its similarity to other members of the population—which is to say, in most cases: the other twin) and the specific environments in which the twins live (O’Connor, 2014, p. 247). This is because this type of analysis “does not ascertain how much genes *in general* matter, or how much the environment *in general* matters. What the experiment investigates is a *specific* set of genetic factors and a *specific* environmental treatment.” (Sober, 2001, p. 56) In other words, if the population used for the study consists solely of twins, it is possible that genetic and environmental factors specific to those twins (e.g. being treated in a similar manner) may limit the external validity of the results. This implies that generalizations to other populations, environments, the species at large, or claims about particular individuals in that population are sometimes claimed to be tenuous at best (see Lewontin, 1974; Longino, 2013). Finally, the analysis only takes into account *variation* in behaviour, not behaviour itself; a distinction which has repercussions on the interpretations of the results. For instance, if all subjects under study exhibit the same trait, the analysis will not reveal whether this is due to genetic or environmental factors; thus, having two arms in a population where everyone has two arms yields a heritability of zero even though it is quite clearly related to one’s genes (Sober, 2001, p. 55). The results furthermore do not directly explain an individual’s genetic predisposition to a behaviour, since it is a comparative population-level analysis, and looks at links between variations in genetic baggage and corresponding variations in phenotypes, not particular genes and their expression in particular individuals. In sum, although behavioural genetics has provided fairly simple and powerful ways of researching heritability, it has serious limitations; as the behavioural geneticist Eric Turkheimer remarks concerning predicting behaviours from genes: "the most obvious example of something resembling genetic prediction is prediction based on observation of an identical twin, with the non-negligible caveats that it isn’t entirely genetic and it isn’t prediction" (2015, p.S33).

4.3.2 Molecular behavioural genetics: methods, scope and assumptions

Surpassing some of the limitations of behavioural genetics, molecular behavioural genetics endeavours to describe the individual-level mechanisms that lead from genes up to behaviour. Researchers attempt to associate specific genetic markers or allelic variations with incidences of a certain behaviour, and ideally show that there is furthermore a causal role. Various methods have been developed to establish the correlations, chronologically starting with linkage analysis, which relies on within-family similarities in genes and behaviours, but is mostly useful only for genes of large effect (e.g. Down’s syndrome) (Turkheimer, 2015, p. S34). Next is candidate-gene association studies, which measure differences in a gene to see if it correlates with a specific outcome across non-related individuals. And finally, genome-wide association studies attempt to associate individual units of DNA, single nucleotide polymorphisms (SNPs), with traits of interest (Turkheimer, 2015, p. S35). The methods have become more and more powerful, yet continue to show that individual genes that play a large role in any complex trait are exceedingly rare, and that the effect of individual units of DNA is vanishingly small (Longino, 2013, p. 61; Turkheimer, 2015).

As is the case with behavioural genetics, establishing that variation in an allele tracks variation in behaviour does not imply a direct causal role, but only the gene’s involvement more generally. One of the ways the hypothesis of the causal role can be strengthened is by revealing another step along the causal chain. For instance, if a researcher can show that a particular gene will have a causal role in a physiological trait (e.g. serotonin metabolism) through experimental manipulation, and that physiological trait is known to have a causal role in a behaviour (e.g. aggression), then that gene can be said to have at least some causal effect on the behaviour (Longino, 2013, p. 54). Other causal pathways include development, where the genes are shown to lead to, or enable, a certain developmental pathway for the organism, which leads to certain behaviours; this is the research carried out by developmental biologists.

Nonetheless, even when it is the case that the intermediaries in a causal pathway can be revealed, the general conclusion of molecular genetic research on behaviour is that the links to be made between genes and behaviour are complex and difficult to pin down with any certainty. In contrast, in classical genetics, "reliable developmental processes were simply presupposed" (Wereha & Racine, 2012, p. 562), leading to the straightforward idea that genes determine phenotypes. Though the supposition that classical geneticists held to the idea of ‘one gene, one phenotype’ may have been overstated (Waters, 2007), it is now clear that any complex phenotypic trait, including psychiatric (Collins & Sullivan, 2013) and behavioural traits (Longino, 2013, p. 54), are polygenetic.

4.3.3 Social-environmental approaches: methods, scope and assumptions

While both genetic approaches attempt to differentiate between the genetic and the environmental influences on behaviours, the social-environmental approach focuses only on the environmental factors. In the case of human behaviour (the focus of Longino’s research), researchers will look at macro-level variables such as social class or cultural identity and micro-level variables such as media exposure or specific life-history experiences, and attempt to determine their influence in the prevalence of specific behaviours or patterns of behaviours. This association between environmental factors and behaviour can be retrospective or prospective. Retrospective studies will look at a population already identified as having a particular trait or behaviour, and compare their environmental background with a control population. Prospective studies will deliberately modify the environment in a given population to determine wither the change has any effect on future behaviours (Longino, 2013, pp. 44–45).

While researchers attempt to isolate factors in the environment which could account for specific behaviours, social-environmental approaches have been criticized for neglecting the possible genetic factors at play in their studies. McGue (1994) and Scarr (1999) each review different studies looking at divorce rates and school achievement, concluding that “restriction of study variables to environmental factors fails to reveal the most likely causal factor—genes” (Longino, 2013, p. 47). However, as Longino points out, while it is true that they are unable to take into account the genetic components of behaviours, the point of the research is to understand the role of competing environmental variables only, and as such the biological aspects are outside of their purview (p.48). To do so, they must assume that the genetic variation in the population studied is either sufficiently uniform or sufficiently random for the differences between individuals to be inconsequential in the first case, or to even out in the second. Thus, while it is true that the social-environmental studies are limited in their scope by their interests and methodology, the larger point Longino is making is that all approaches are faced with their own limits, which stem from their methods, scope and assumptions.

4.3.4 Incompatibilities

We thus have a picture of a few different approaches to explanations of behaviour, which according to Longino explains why the explanations given are incompatible. The first reason explaining the incompatibility is the general idea that these approaches simply do not ask the same questions, and therefore do not arrive at the same answer. For instance, while the environmental approach asks about which environmental factors are most important for a given behaviour, the quantitative behavioural genetic approach will instead ask what portion of the variation is attributable to genetic variation. In practice, Longino points out that this leads to each approach characterizing themselves as the embattled minority, claiming that the rival approach is "more generally, if wrongly, accepted" (2013, p. 138) and calling for the supremacy of their approach. She goes on to show examples of publications in each approach trying to discredit their opponents, putting forward their own discipline as the best way to explain behaviour. Yet as Beatty (1997) suggests, despite the language used in the rival publications, these arguments are more aptly seen as disagreements regarding the *relative significance* of each theory, and not about discrediting an entire discipline. The fight for publications, for research grants, and for notoriety can lead to an exaggeration regarding the merits of one's field, which often is done by disparaging others. Thus, while it is true that the fact that the approaches do not ask the same questions can lead to conflict, this in and of itself is not proof of *incompatibility* in explanations: as Longino pointed out with respect to the socio-environmental approaches, each approach has its own scope, and is limited in what it can take into account. As such, when an environmental approach states that such and such an environment is central to the development of a given trait, they are not claiming thereby that genes do not play a role, but instead that among different possible environments, this particular environment is more inclined to bring about the behaviour under study. So, while it is possible that asking different questions about a phenomenon could bring about incompatibility, it is not the case that it always will.

Her second argument is similar to the first, but digs deeper into the reasons why different approaches could have incompatible explanations, looking into the way they parse the context of the explanandum phenomenon. Longino proposes that the incompatibility is not only because of the questions asked, but also because of the parsing of the causal space of the phenomenon to explain, which is needed because of the dynamic and constitutive complexity of the phenomena to explain (see chapter 1, section 1.3). What this means is that for a given behaviour in need of explanation, there is a space of possible causes, including allele pairs, the intrauterine environment, physiology (such as hormone secretion), shared and non-shared environment, among others. Any and all of these causes could have an effect on the behaviour, and the scope, methods, and assumptions of each approach will be such that they look only at certain parts of the causal space, and will even sometimes cut it up differently from one another. For instance, molecular genetic approaches look only at genes or alleles which would cause a role in the expression of phenotypic traits, but their experimental protocols “are not designed to distinguish among hypotheses about nongenetic causal factors or between genetic hypotheses and hypotheses concerning specific nongenetic factors” (Longino, 2013, p. 59). In contrast, environmental approaches will simply not have the tools to factor in possible genetic causes for the behaviours under study. Another example given by Longino is how, while quantitative behavioural geneticists will see the intrauterine environment as an environmental factor (since it is not included in the genes carried by the twins), environmental approaches will see it as a non-environmental factor, since it is not a measurable environmental factor that they can consider (Longino, 2013, p. 127). As such, the explanations proffered by each approach can be incompatible with one another, since they do not consider or measure the causal space in the same way, leading some to highlight certain environmental causes as the main determiners of a given behaviour, while others will point to specific genes or genetic variations as the main drivers of the behaviour.

Up to this point, I agree with most of what Longino has proposed, namely the difficulties—but necessity—of singling out a single phenomenon in order to properly assess if explanations truly are incompatible with one another, as well as the analysis of the differential parsing of the causal space, which accounts for the incompatibilities. Her research highlights the challenges and potential pitfalls of studying behaviour through a plurality of approaches, and her work will be foundational to my own monist position covered in chapter 7. Where we part ways however is when Longino proposes that the incompatibilities in scientific explanations could be permanent due to incommensurability, and a staple of successful science.

4.4 Incommensurable explanations

This brings us to the third and last building block of Longino’s pluralism: the incompatibility must be shown to be permanent. This, according to insular pluralists, is an important point which differentiates them from more moderate forms of pluralism. Longino characterizes some pluralists such as Mitchell (2002) or Kitcher (2001) as accepting a temporary pluralism which will ultimately be eliminated, and collapse into monism. In other words, they would accept some incompatibility in explanations, but attribute it to the immaturity of a science, with the understanding that all incompatibility will be resolved at a later date. In contrast, and in order to completely avoid all monism, Longino commits to the fact that certain explanations will be impossible to reconcile one with the other, implying that the pluralism will be ineliminable. What's more, this ineliminability of incompatible explanations for a given phenomenon is not seen as a failure, or even a problem, for science, and is instead the result of science producing a plurality of incompatible yet successful explanations. Of course, this prima facia poses a problem for those who expect scientific explanations to tell us the truth about how the world works. After all, if truthful scientific explanations describe how the world really is, and there is only one world to explain, how could there be multiple incompatible explanations? To defend such a position, Longino proposes a form of incommensurability of scientific explanations (2013, p. 147).

Incommensurability was brought to the forefront of philosophy of science in the twentieth century through the work of Thomas S. Kuhn and Paul Feyerabend. Kuhn's famous book *The Structure of Scientific Revolutions* (1996, originally published in 1962) proposed that science progressed through revolutions, with one theory coming to supplant another, rather than simply progressing through the accumulation of knowledge. According to Kuhn, when two theories are in competition, they are so radically different as to be incommensurable, meaning that there is no way of finding a common measure between the two, such that we would be able to make sense of one theory through the concepts of the other. Feyerabend had been using the concept of incommensurability more than a decade before Kuhn to "challenge different forms of conceptual conservatism in science and philosophy" (Oberheim & Hoyningen-Huene, 2018). In 1962, he too introduced the term "incommensurability", this time to attack reductionist and positivist ideals regarding inter-theory communication. He argued that if the theories are mutually exclusive of one another, then they are conceptually incompatible, and therefore incommensurable (P. K. Feyerabend, 1962).

Longino's own version of incommensurability relies on many of the same intuitions, proposing that the incompatibility of explanations highlighted in the previous section can be permanent if and when the approaches are unable to find common ground in the comparison and understanding of each other's results. To defend her position, Longino puts forward two arguments, the first of which highlights how the differential parsing of the causal space will be such that differing approaches will be unreconcilable. The second argument for its part, explains how each explanation can nevertheless be said to be successful, implying that the incommensurability is not problematic, but merely a feature of the way in which science operates: through a plurality of correct but incompatible explanations.

4.4.1 Parsing of the causal space

Longino's first argument relies on the different ways the approaches understand, measure, and represent the phenomena of interest. In *The Fate of Knowledge*, Longino suggests that "given that a phenomenon is modeled slightly differently in different approaches, quantitative measures will vary between approaches, so that comparing data descriptions will show inconsistencies between approaches"(2002, p. 93). This idea is fleshed out and pushed further in her more recent *Studying Human Behavior*, where she argues that "each approach measures variation in differently parsed causal space. These different parsings result in incommensurabilities among the approaches" (2013, p. 127). In other words, what is implied by the differential parsing of the causal space seen in the previous section is that the varying methods and scopes will sometimes yield blind spots regarding the effects of other potential causal pathways. For example, whereas environmental approaches will single out environmental causes, they will be unable to even consider the effect of genes. They consider and measure different environmental factors, but they have no tools to conceptualize or measure genetic influences, meaning that the scientific explanations they produce will not only be incompatible with genetic explanations, but will also be in no position to evaluate or integrate those explanations. Because their methods preclude the possibility of integrating other causes into the explanation, "research pursued under the aegis of any of these approaches pushes them in nonreconcilable directions" (2013, p. 126). As such, according to Longino, not only will the explanations stemming from different approaches be potentially incompatible with one another, they will also be incommensurable because there will be no way to meaningfully compare or integrate results.

4.4.2 Conformation

Her second argument relies on the abandonment of 'truth' as the only measure of success for scientific explanations. As seen in chapter 1, pluralists in general, and Longino in particular, conceive of scientific explanations as representations, as opposed to as set of propositions about a phenomenon (2002, p. 113). Explanations will sometimes be in the shape of visual representations, diagrams, models, and many other kinds of non-propositional representations. Longino contrasts this with propositions which are taken to be either strictly true or false (more on whether this distinction holds shortly). In contrast, these non-propositional representations are not amenable to the binary of truth or falsity, since they (often deliberately) idealize, distort, and/or omit certain information about the phenomenon or object targeted by the explanation, meaning that they will rarely if ever be strictly 'true' to the object they are representing (see section 1.4 and chapter 3 for further discussion). For example, the twin-helix model of DNA is a representation of actual DNA, but never in fact represents actual DNA (especially actual, specific bits of DNA) in all its details. Some models will be particularly detailed, including the precise geometry of each atom within the structure and the length of covalent H-bonds involved, but others will gloss over many points, idealize many aspects. Yet both these approaches give us a representation which simplifies (sometimes more, sometimes less) the object enough to allow us to better understand it, and to make it applicable to a wide range of phenomena (in this case DNA in general). And yet they are nonetheless considered a successful scientific representation.

To make sense of the fact that these representations can still be considered successful, Longino proposes to evaluate this success through a continuum of "conformation." Drawing on her previous work (2002, pp. 114–118), she proposes that conformation encompasses "truth (literal) at one extreme, but also isomorphism, homomorphism, similarity, approximation, and other relations that name forms of success in representation" (2013, pp. 147–148). In this view, because scientific explanations are representations which need not be an exact correspondence with the world, one can have a plurality of explanations which each partially correspond to the world in their own specific ways, implying that despite their incompatibility, they are each, in their own way and in their own respects, 'correct' insofar as they conform to their subject matter. So, for instance, the environmental approaches will produce an explanation about which environmental factors are most important, while a molecular genetic approach will produce an explanation which highlights the role of a SNP. Each of these explanations is not merely 'true', since they omit, idealize and/or distort the object of the representation (each omits the other approach's side of the explanation), but they are successful insofar as they do correspond in some respects to the object. Thus each conforms in some respects to reality and is therefore a successful explanation, without invalidating the other.

Now, with respect to the distinction between propositions and representations: Longino herself concludes her remarks concerning the conformation of representations by suggesting that propositions as well can be understood as conforming more or less to the referent (2002, pp. 118–119). For instance, the proposition 'this bird is exhibiting aggressive behaviour' could conform only partly to reality if aggressivity is measured on a scale, and the bird is only mildly aggressive, or aggressive only at certain times. Propositions are therefore seen as one option among many for representing the information shared through scientific explanations, which, just like other representations, rely on conformation for denoting their success.

Conformation not only implies that each representation must be partial and can therefore make room for other representations, it can also preclude commensurability through the way conformation is measured. According to Longino, though conformation denotes some correspondence to reality, the success of that correspondence is only measured from *within* a given approach: “correctness must be relativized to the initial parsing of the causal space” (2013, p. 148). As such, the conformation of an explanation stemming from a socio-environmental approach is determined not by the correspondence of the explanation and reality in some absolute sense, but conformation with respect to the way the phenomenon is understood and measured, taking into account the specific parsing of the causal space which the approach has adopted. What’s more, “different kinds of conformation will be mandated by different pragmatic aims.” (p.149) Certain explanations will be more useful for certain interventions, and that usefulness will be an element in the evaluation of conformation. As such, we cannot compare the degrees of conformation of representations coming from different approaches, because each approach determines for itself how to measure the conformation: “conformation to degree *n* in one setup is not equivalent to conformation to degree *n* in a different one.” (2013, p.148) Consequently, because each approach is only capable of measuring the degree of conformation of its own explanations, in its own unique way, it is not possible to build bridges with other approaches: “in comparing claims issuing from different approaches, it is necessary to use at least a common degree of conformation.” (2013, p.148) The idiosyncratic nature of the measure of conformation is thus another element which Longino puts forth to argue for the incommensurability of scientific explanations.

In sum, according to Longino, scientific explanations are best seen as representations which can be more or less successful at portraying the world as it actually is : conformation “treats success as at least in part a relation between content and some object or set of objects distinct from that content” (2002, p. 116). In other words, while conformation leaves space for scientific explanations which are not strictly speaking entirely ‘true’, it does nevertheless rely on the fact that scientific explanations are about the world, and remains therefore non-relativistic. The success of that relation is furthermore measured only from within the various approaches, meaning that each approach is unable to make sense of the explanations stemming from other approaches.

4.4.3 The cartography analogy

To better explain this way of seeing scientific explanations, Longino calls on the cartography analogy, a staple in pluralist writings (see among others Giere, 1999; Kitcher, 2001; Longino, 2002; see also Ruphy, 2013 for a discussion about the analogy). Just as maps represent only a partial view of the terrain they are meant to cover, so scientific explanations are only partial representations of the object they are explaining. Indeed, maps will focus only on certain elements of the terrain they are meant to represent which will vary depending on the purpose of the map (Longino, 2002, p. 116). This is what Kitcher calls the "intended content" (2001, p.57): the entities and relations which the map is created to represent, which are deliberately limited in scope. For instance, a map of the Montreal metro (subway) will have as intentional content the adjacency and order of the different stations, but deliberately distort the precise geographical location of each station in order to more clearly represent the intended content. It will be much easier to see which is the next stop if all the stops are represented as equidistant along a mostly straight lines, with no streets on the map, rather than trying to parse a topographical map which contains all the streets, street names, bus lines, etc., and which tries to accurately portray where the subway lines have actually been dug. The intended content is therefore composed of all and only the elements and relations which the map has been created to represent.

Another crucial aspect of maps are the conventions used (Kitcher, 2001, p. 57; Longino, 2002, p. 116). Because the map represents aspects of the world which are not the map itself, an idiom is developed which allows for the correspondence between the visuals of the map and the features of the terrain being mapped. For instance, a topographical map will translate altitude into contour lines, and may shade as green the parts which are forested. The metro map will have a white point which corresponds to each station, with lines of a given colour between stations which can be accessed without transferring between trains. The fact that such lines and points are used, and the chosen colour, are determined somewhat arbitrarily, as long as the conventions allow for ease of understanding. Once the convention is set however, it must be respected, and the success of the map is determined by comparing the intended content, as translated by the convention, with the terrain itself.

One important point which follows from this is the fact that any given map is only ever a partial representation. Indeed, it will not contain all the information it is possible to garner about a given terrain, as it would be too much to parse. Maps convey information about some features of a given terrain, but they do not contain within them *all* of the information about the terrain. It is in this sense that they can be said to be partial, since no map will represent all of the elements and relations of a terrain as they actually are, "because that would duplicate the terrain being [and would therefore] be useless" (Longino, 2002, p. 116). As mentioned earlier, Mitchell (2009, pp. 116–117) makes a similar point by calling on the short story by Jorge Luis Borges aptly called "On exactitude in science"(1975), which tells of a map which is in a one-to-one correspondence with the terrain it portrays, but is therefore too large to be of any use (see 1.4.1). Maps therefore are seen to be good at what they do in virtue of the fact that they are partial representations of the terrain, not despite that fact.

What this implies regarding scientific explanations is that, just like maps, explanations are partial representations of the explanandum phenomenon, and rely on intended content and conventions to judge their conformation. And just like maps, the conformation to the object of study will not be a matter of truth or falsity, but rather degree of conformation. Furthermore, because the measure of conformation for any given explanation depends on "different measurement setups" and "pragmatic aims" (Longino, 2013, p.148-149) within the approaches, comparison of conformation across contexts is impossible. This is analogous to the intended content and conventions used in a map: scientific explanations, just like maps, focus on specific aspects of the explanandum phenomenon, find specific ways to measure it, and seek out answers to questions specific to the approach. What this entails is that each approach, or each map-maker, using their own methods, will be unable to judge whether or not the explanation from the rival approach or map-maker conforms or not, leading to incommensurability. And finally, just like maps, we can have a plurality of scientific explanations about a single phenomenon and not consider that plurality to be a problem in need of resolution; just as having a map of the Montreal metro does not make a map of the Montreal streets useless or problematic, so having an environmental explanation of a given behaviour does not make a genetic explanation useless or problematic.

4.4.4 Insular pluralism

What these two arguments for incommensurability yield are scientific approaches creating a multiplicity of representations of causal interactions for a given phenomenon, which can all be successful despite being incompatible. Longino sees this in a positive light: "Epistemologically, we may learn more about a system by utilizing multiple partial representations, each of which enables us to go further in our study than would the attempt to obtain a complete representation of all causal interactions" (2013, p. 147). This, then, is why Longino claims that the plurality of explanations is not seen as problematic, but only the result of successful science running its course. There is furthermore no reason to believe that this plurality will disappear, because the incommensurability precludes the possibility of eliminating or integrating explanations, each approach having their own methods, scope, and assumptions, and their own measures of conformation which are not applicable to those of the other approaches. According to Longino, the position to adopt is therefore insular pluralism, which allows for the multiplicity of incompatible explanations for a given phenomenon, and sees this plurality as the mark of successful science.

4.5 Is insular pluralism tenable?

Longino presents a strong and interesting case for insular pluralism, which in many ways is more robust than Mitchell's fragmentation pluralism, despite being more forceful. Her arguments rest mainly on the fact that once we are sure that the explanandum phenomenon has been correctly identified, different approaches will have different methods, scopes and assumptions, which will lead them to produce incompatible explanations for that same phenomenon. Longino stresses that the way the various approaches parse the causal space makes it impossible to compare and integrate results across approaches. The success of the explanations is furthermore measured only from within each approach, using their own measures to evaluate just how much the explanations conform to the explanandum phenomenon. This therefore leads to incommensurability, implying that the pluralism found at the level of scientific explanations of behaviour is here to stay, and should be seen as the result of successful scientific research.

There are however good reasons to think that the incommensurability Longino puts forward will not stand the test of time. While it is true that different approaches parse the causal space in different ways, leading to incompatibilities, it is not the case that this parsing is static. Through changes in the way one sees the phenomenon itself and its causal history, it will be possible to make the different approaches commensurable. Ironically, it is precisely the kind of work that Longino has done, laying bare the methods, scopes, and assumptions of different approaches, which allows for the building of bridges between approaches, and the eventual breakdown of incommensurability. This opens the door to the integration of various explanations, and indeed, as will be shown in chapters 5 and 6, strategies for integration abound. The tools used for integration are furthermore constantly being added to, pointing to the fact that there is no a priori limit to the possibilities of integration.

The fundamental reason why integration will always be possible is because scientific explanations are successful insofar as they are representations of reality. And reality, as pluralists themselves agree, is not plural: there is only one world to explain, and for any given phenomenon, only one causal history which is the target of explanation. As such, even if we agree that the degree of conformation of an explanation is, at least to some extent, measured through the lens of the approach which produced the explanation, it is nevertheless the case that it is successful only insofar as it does indeed conform to reality itself. As such, each explanation implies some empirical commitments with respect to the measures and predictions related to the explanation, and ontological commitments with respect to the entities and relations posited within the explanation. As will be shown in chapter 7, I propose that these commitments are the Achilles' heel of explanatory pluralists and are the basis for my defense of integrative monism.

The remaining chapters are thus both a critique of insular pluralism, as well as a defense of my own form of monism. Chapters 5 and 6 tackle the integration of scientific explanations, showing the myriad ways that incommensurability can be—and has been—broken down. This paves the way for my defense of integrative monism, found in chapter 7.

CHAPTER 5

THE INTEGRATIVE TOOLKIT

Scientific integration is garnering attention in philosophy of science, with more and more publications appearing since the beginning of the millennium. In the second half of the twentieth century, there was a push for more interdisciplinary research to be done in academia, leading some to talk about the integration of various fields (see for instance the contributions to Bechtel, 1986). At the time the emphasis was on the integration of theories or entire fields or disciplines, or the eventual creation of new theories which would be able to bridge the gaps (Darden & Maull, 1977); more recently, the spotlight has shifted from theories to other units of scientific research, such as the explanations themselves, the methods, the tools, etc. (Brigandt, 2010) These epistemic units (theories, explanations, methods, etc.) are combined in such a way that they contribute together to a better understanding of a given phenomenon. It seems to be unanimously agreed that integration, when it happens, is a good thing, though this apparent agreement is tempered by a few caveats.

The first is that it is not always clear what is understood by 'integration', an especially pressing issue when it comes to prescriptive stances which encourage researchers to pursue integration. In the case of biological research, integration is a term "used to cover a multi-faceted dynamic, in which methods, bodies of data and explanations are synthesized in order to understand and intervene more effectively in biological systems." (O'Malley & Soyer, 2012, p. 59) The second caveat is that despite generally seeing integration in a positive light when it does happen, not everyone agrees that it is a necessary or ubiquitous desiderata for scientific research, namely those pluralists who would see it as a (potentially) fruitless and wasteful pursuit (Kellert et al., 2006b; Longino, 2013). Of note is that even Mitchell, who explicitly promotes what she calls integrative pluralism, sees integration as a good strategy only in certain circumstances, rejecting it in others (Mitchell, 2003, 2009; see also chapters 2 & 3 of this thesis).

One objective in this chapter is to address the first of those caveats by giving a sense of what is meant by integration. The sheer number of publications and references to integration of scientific explanations, theories, methods, etc., be it in the philosophy of science literature or the scientific literature itself, is such that it would be impossible to cover them all in a single chapter. My goal instead will be to give a glimpse into the variety of ways of doing integration in the biological sciences to show how there is seemingly no limit to the creativity employed in creating integrative solutions. Section 5.1 begins with a general account of integration. I then cover accounts of theory integration (section 5.2) and explanatory integration (5.3), with an emphasis on how the integration is understood to be carried out, and what results are expected. I will then turn to two examples of integration in practice. The first example concerns alcohol dependence, a phenomenon which is the topic of much work in behaviour research, namely by Kendler and colleagues (Kendler et al., 2002, 2006; Kendler & Myers, 2010) to show how integration is done in this specific case, and how it allows for a better understanding of the phenomenon (section 5.4). The second example is more general, and relates to Tinbergen's four questions, which are often taken as an example of the relative independence of different explanations in biology. I will show how in fact, that purported independence is eroding in current biological research, thanks to explanatory integration (section 5.5). Both these cases serve to show how the claimed pluralism associated with these examples does not hold up to integrative strategies.

My second objective throughout the chapter is to show how these integrative strategies are in fact so many ways of breaking down purported incommensurabilities. Indeed, for integration to even be possible, the various approaches from which stem the integrating epistemic units must be able to communicate and understand one another. As will be shown, integrative strategies rest on tools which are meant to facilitate this commensurability, building bridges between different ways of parsing and understanding an explanandum phenomenon and its explanans. By multiplying integrative strategies, researchers are building what I call an "integrative toolkit", adding resources which enable and facilitate the discovery of common ground across approaches. Each of these tools chips away at Longino' insular pluralism, more specifically at her ideas regarding the incommensurability of scientific explanations. The ever-expanding integrative toolkit shows the flaws in the insular pluralist's portrayal of approaches unable to find a common measure. This paves the way for my argument found in chapter 7, where I propose that there is no in-principle limit to integration.

5.1 A general account of integration

Integration is a difficult term to define precisely, as it can be used to describe all manner of practices used to join together epistemic units stemming from scientific research, such as—among others—theories, explanations, models, data, or results. Of course, the difficulty lies in characterizing more specifically what is understood as 'joining together', and how or to what extent each epistemic unit is involved. What's more, in a textbook dedicated to interdisciplinary research, the authors suggest that "in today's research practice, interdisciplinary integration [i.e. integration of epistemic units stemming from various disciplines] often occurs but is rarely described" (Rutting et al., 2020, p. 42). As such, it can be a challenge to find examples of integration, as they are not always named as such, and once found, it can be difficult to know precisely how it was carried out.

O'Rourke et al. (2016) have done tremendous work in attempting to understand and synthesize what is understood by 'integration' both in the context of interdisciplinary studies, as well as in philosophy of biology. They remark that integration is often described not entirely concretely, but through an array of metaphors, describing how epistemic units come to be fused, melded, blended, amalgamated, harnessed and knitted together. As they point out, "one striking aspect of [these metaphors] is that they are all approximations of a process in which different things are combined into one" (p.67, footnote 19), which to them is the essence of integration. Their understanding of integration additionally puts to the fore the *process* by which epistemic units come to be combined, rather than the *result* of that process. Ultimately, their proposal, which I adopt here, is to understand integration as "a generic combination process the details of which are determined by the specific contexts in which particular instances of integration occur." (p.67) Their definition is deliberately large, to adumbrate essentially all aspects and all ways of conceptualizing integration in the life sciences. It also puts to the fore the contextual element, highlighting how the details of integration depend to a great extent on what it is which is being integrated, as well as the research context. As such, there is no a priori way of determining the range of integrative processes which could be invented, since each integrative strategy will depend on the context. In line with this idea, much of this chapter is devoted to enumerating the multiple ways that integration has been understood or carried out, illustrating the multiple possibilities of integrative processes.

Their work however makes no distinction between integration at large and explanatory integration. Yet many researchers propose that explanatory integration is a new and more specific way of characterizing a type of integration. Three elements often recur in discussions regarding explanatory integration. The first is that it can be done both within and across fields. According to Brigandt (2013), integration "can refer to either the integration of different scientific fields or the formation of an integrative account that combines a variety of different ideas", the latter of which can be done within a single scientific discipline, by for instance combining explanations at different levels (p.461). A second important feature of explanatory integration is that it is typically understood to apply to many different epistemic units, and not only to theories or disciplines. Indeed, as mentioned at the outset of this chapter, earlier discussions surrounding integration tended to focus on the integration of theories or of disciplines (c.f. Bechtel, 1986), often with an emphasis on reduction. More recent takes on integration often see it as an alternative—or even in opposition—to more traditional intertheoretic reduction as we will see below. The third element also often places it in opposition to reduction: integration is understood as a reciprocal relation, rather than an asymmetric one (Love & Lugar, 2013, p. 548). No field, or explanatory element entering into the integration, is understood a priori as being superior to the others, nor in a position to eliminate or otherwise eclipse the other elements. Whereas reduction is often understood to give the greatest weight to the more fundamental theory, in the case of integration, it is understood that all parts will contribute, and that no theory, field, explanation, etc. will be in a position to eliminate the others. Explanatory integration can consequently be understood to be the bringing together of epistemic units (such as theories, explanations, methods, results, etc.) from different approaches, for the creation of a novel explanation, with no approach understood a priori as being superior to any other.

As will be shown through the description of the variety of ways of doing integration, integration in general, including explanatory integration, can have consequences on the explanations put forth for a given phenomenon, leading to the elimination of explanatory pluralism. Integration implies that explanations can be integrated one with other, but it can also be used to create novel explanations, or create the epistemic units necessary for the creation of new explanations. Integration thus breaks down explanatory pluralism because it creates the tools necessary for resolving the incompatibilities in explanations highlighted by Longino (see chapter 4). This does not necessarily mean that the explanations themselves will be integrated—though that is possible—, rather that epistemic units of all sorts can be combined to sidestep any purported incommensurability, and to reconcile previously incompatible explanations.

The two following sections cover both integration at large, and explanatory integration. The first deals specifically with theory integration and disciplinary integration, as they are the original way in which integration was understood in the philosophical literature. The second section covers explanatory integration, the more recent take on integration, which covers all manners of ways in which epistemic units of all sizes (such as explanations, models, data, results, etc.) are made to interact to yield integrated explanations. The articles and examples covered in the following sections are related not only to the biology of behaviour, but also similar phenomena and fields, the literature specifically on integration of biological explanations of behaviour being too sparse.

As will become clear, the corpus covers many aspects of integration, some authors talking about specific methods for creating integrations, others proposing general ways of conceptualising it, others still putting forward ways of categorizing types of integration. More often than not, these three aspects are conflated, or at least not explicitly teased apart; there is of course significant overlap in these three objectives, and authors sometimes switch freely from one to the other. Nevertheless, with respect to the concrete strategies for integration, it will become clear that all approaches concur that there must be an emphasis on aligning vocabularies, concepts, or ontologies, between the various explanations, in order to allow for a common understanding of what is at stake. This can apply either to the explanandum phenomenon or the explanans, or to both at once, in order to make sure that researchers are talking about the same phenomenon (or aspects of the phenomenon), and that the notions used in the context of one explanation do indeed correspond to notions used in the other.

5.2 Integration of theories and disciplines

As mentioned earlier, one of the more 'traditional' ways of integrating epistemic units has been through theoretical or disciplinary unifications. In the last decade or two however, philosophers have sometimes seen integration as an alternative to these approaches, focusing on the fact that many integrations happen at smaller scales than entire theories or fields. The three most well-known ways of approaching theoretical integration are described in what follows, namely reductionism, unificationism, and interfield theories.

5.2.1 Intertheoretic reduction

Intertheoretic reduction is the idea that a theory (typically at higher levels of organisation) can or will be 'reduced' to another theory (typically at lower levels) (E. Nagel, 1961; Schaffner, 1967, 1974). This type of epistemic reduction has been described in chapter 2, and I will therefore not be going over that same ground. Intertheoretic reduction can be seen as a form of theory integration, insofar as it is a process which relates one theory to another, by showing how one can be derived from the other.

And yet, reductionism is sometimes seen as antithetical to explanatory integrative approaches. Indeed, for some, integration "is a replacement for the traditional idea of reduction" (Brigandt, 2013, p. 461; see also Mitchell, 2003, p. 192), among other reasons because—as mentioned earlier—integration tends to be understood as a reciprocal relation rather than an asymmetric one. Another reason is that reduction is understood as a stance which aims at total reduction of all theories to a single, fundamental theory

(Cusimano & Sterner, 2019, p. 55), whereas explanatory integration typically eschews such vast goals, focusing on local integrations.

However there are a few considerations which make this position less convincing. The first is that reductionism need not imply a completely asymmetric relation between the theories. Many reductionists do not claim that we can do away with higher-level theories, even once they have been reduced. Nagel himself agrees that "explanations at higher levels often remain practically preferable and for many purposes indispensable" (1998, p. 2). Of note however is that the stronger version of reductionism, eliminativism, is not a form of integration. Because eliminativism seeks to abolish the higher-level theory once it has been reduced to the lower-level theory (P. M. Churchland, 1981), it is best seen as opposing integration. Integration is such that both theories would need to be preserved, and together act to explain the phenomena. Eliminating one of the two theories therefore would rightly be seen as eschewing integration in favour of a single, superior, theory.

The second consideration suggesting that reduction is a form of integration is that one could suppose that intertheoretical reduction is possible in some cases but not in others, meaning that the aim of unifying science through reduction is abandoned, without the need to abandon reduction in general. This also implies that reduction is not in opposition to integration, but rather one integrative strategy among others. In view of this, Mitchell's idea that reductionism is not to be abandoned by those looking for integration, but only recognized as one among many research strategies (Mitchell, 2009, pp. 22-23,44) seems to be the right approach.

The third consideration which highlights how reductionism and integration could go hand in hand is the fact that new developments propose that reductionism can also be applied to smaller epistemic units than theories, such as parts of theories, explanations, mechanisms, etc. It is true however that in this case we are no longer talking about intertheoretic reduction as envisioned by Nagel and others. This type of reduction when applied to explanations (as opposed to theory reduction) was covered in some detail in chapter 2, and will be addressed later in this chapter.

The point to underline for the purposes of this chapter is how intertheoretic reduction happens; in other words, how the potential incommensurability between theories is overcome. The challenge for any reduction is to establish correspondences between elements in either theory, such that one theory is in a position to be derived by the other. Without such a correspondence, it is *prima facie* impossible for one theory to be shown to be the logical consequence of the other (E. Nagel, 1961, p. 352). Nagel distinguishes between two cases where reductions can occur. The first is called "homogeneous", and concerns reductions where the vocabulary used in both is approximately similar, or in other words homogeneous. Because they share the same language, reduction is a relatively straightforward affair, with one theory coming to encompass the other. Nagel gives the example of Newtonian mechanics absorbing Galileo's laws (which concerned only terrestrial motion), thereby explaining the movement of both terrestrial and celestial motion through a single theory. But because each theory used the same vocabulary of motion, mass, etc. communication and correspondence between concepts was relatively simple (1961, p. 339). The second kind of case is "heterogeneous", and concerns theories which do not share a common language. In these cases, it will be necessary to create "bridge principles" which allow for the connectability of the concepts from each theory, ultimately leading to the derivability of the reduced theory from the reducing theory (1961, p.453-454). Nagel proposes three ways for these connections to be created: (1) by establishing logical connections between terms, such as synonymy or one-way entailment, (2) by conventions, such that the correspondence is created by deliberate fiat, or (3) by factual or material correspondence between a theoretical concept and a state of affairs (p.354). For instance, the concept of 'temperature' in Boyle-Charles' law can be made to correspond to 'mean kinetic energy of molecules' as derived from statistical mechanics through (2) or (3), despite the fact that each of these concepts makes no mention of the other. By creating bridge principles, the terms used in the Boyles-Charles law and in statistical mechanics can be connected to those used in the other, allowing for derivation, and hence reduction.

Schaffner (1967, 1969, 1974), applying these ideas to biology, refined Nagel's model to allow for corrections in the reduced theory before reduction. In doing so, he addressed a problem which was identified in Nagel's account, namely that it is entirely possible that the reduced theory will contain false statements, which should not (or could not) be derived from a correct reducing theory. Classical genetics' inheritance laws, for instance, do not account for certain exceptions, and therefore need to be modified before reduction (see Brigandt & Love, 2017 for more details). Corrections are therefore made to the (typically) higher-level theory before being reduced.

In sum, intertheoretic reduction contain tools for breaking down incommensurabilities. By creating bridge principles, and/or by correcting certain problems in the reduced theory, it is possible to find correspondences between the theories. This acknowledges that it is not always a straightforward affair, and that the commensurability of the two theories will need to be actively built, not taken for granted.

5.2.2 Unificationism

Unificationism, as defended by Kitcher (1981, 1989), is the idea that scientific explanations are meant to explain as many phenomena as possible with as few premises and brute facts as possible. Our most successful scientific explanations unify disparate phenomena, showing how they can in fact be explained by a few relatively simple argument patterns. Following Hempel (1965), explanations are here understood as arguments, or more specifically, argument patterns: schematic sentences which sketch out the type of explanation a given theory will provide for any given phenomenon. To make this abstraction a little more concrete, Kitcher (1989) illustrates it with the progression of classical genetics (p. 438-442). Mendel's original idea was that the transmission of phenotypes could be accounted by supposing the existence of dominant and recessive alleles, such that you could predict and explain why certain proportions of descendants had such-and-such a phenotype. The problem was that this argument pattern did not account for many phenotypes which did not perfectly fit onto the simple dominant-recessive explanatory schema. Morgan later refined the Mendelian explanation by adding in the linkage relations among loci, explaining why some alleles were more likely to be transmitted together to the descendants. Later still, the Watson-Crick model accounted for even more phenomena by referencing the DNA sequences, transcription mechanisms, details of cell biology, and many other factors, showing how many of the exceptions to Morgan's model could be accounted for. This simplified history of classical genetics shows that over the space of a few decades, the argument patterns used to explain trait transmission were modified such that they could explain more and more phenomena using one single explanatory schema which changed across successive developments, referencing the underlying mechanisms thought to account for phenotypes. The Watson-Crick model is therefore superior insofar as it unified more disparate phenomena under a single scientific explanation. Unificationism takes this to be the goal of scientific explanation.

This relates to integration since one of the ways that unification can proceed is through the integration of theories, leading to the explanation of a broader range of phenomena. For instance, while Mendel's Laws allowed for the explanation of the transmission of many traits, it was ultimately with the integration of theories stemming from cytology and molecular biology that greater coverage was achieved. Indeed, the Watson-Crick double-helix model came from a collection of information and data from biochemists and chemists such as Rosalind Franklin, and contributed to Mendel's theory by bringing in information from outside of classical genetics.

Unificationism in and of itself is not integration however, for two reasons. The first is that unificationism concerns the capacity to explain more phenomena through fewer argument patterns, rather than a process of incorporating epistemic units from elsewhere. This implies that, while some unification may involve integration, some will not; it is possible to extend the range of phenomena explained only through intra-theoretical progress. As Kitcher points out, Mendel's dominant-recessive model was further refined to take into account different loci, accounting for epistasis (Kitcher, 1989, p. 440), without the need to integrate epistemic units from within or without the field or theory. The second is that unificationism is silent when it comes to the process of integrating theories or explanations, aside from the general idea of accounting for more phenomena with few premises. For instance, though it does tell us why the acceptance of the Watson-Crick in classical genetics is considered a success, it does not tell us how it was done, nor does it tell us how future integrations could proceed, except to specify the result: greater unification of disparate phenomena under as few argument patterns as possible. In this respect, it is more aptly seen as an account of why integration might be desirable for scientific explanations, as opposed to providing a guide for integration itself.

5.2.3 Interfield theories

The last approach to theory integration covered here are interfield theories, proposed by Darden and Maull (1977). It must be said that their approach is explicitly not about the integration of theories themselves, but it does result in a theory which is able to build bridges between different fields. Interfield theories provide some of the answers that were lacking in Kitcher's unificationist account, explaining in a little more detail, for instance, how it is that cytology and Mendelian inheritance came to be integrated, thus allowing for unification.

Fields are described by Darden and Maull as areas of science which share certain elements, such as a central problem, accepted techniques, theories, and/or facts which are used in attempts to solve the problem, as well as expectations regarding explanatory factors and goals (1977, p.44). Different fields, understood in this broad manner, can be interested in the same phenomenon, or phenomena which are closely linked. When this is the case, the different fields can often enter into certain relations.

> [An] interfield theory is likely to be generated when background knowledge indicates that relations already exist between the fields, when the fields share an interest in explaining different aspects of the same phenomenon, and when questions arise about that phenomenon within a field which cannot be answered with the techniques and concepts of that field. (p.50)

Thus, when a field finds itself unable to account for all the questions it has, it is natural for researchers to turn to other fields, and in so doing, create an interfield theory which combines explanations from each field.

Darden and Maull suggest four specific ways that the interfield theory can "make explicit and explain relations between fields" (p.48), as well as an example for each. First, one field can specify the physical location of entities or processes from another field, for example when cytology confirmed the location of genes for the Mendelian theory of heredity. Second, a field may specify "the physical nature of an entity or process postulated in another field", such as when biochemistry described what a repressor is, which was an entity in the operon theory. Third, one field may explain the structure of entities or processes whose function is explained through another field. This is the case for all fields which explain the constitutive characteristics of elements in other fields, such as physical chemistry detailing the structure of molecules which have a biochemical function. And finally, one field can specify the causes or effects of entities or processes used in another field, for instance when "the theory of allosteric regulation provides a causal explanation of the interaction between the physiochemical structure of certain enzymes and a characteristic biochemical pattern of their activity." (p.49) These four ways that interfield theories can clarify the links between fields are general ways of understanding the methods and role of integration.

To illustrate their approach, the authors use the same example covered above regarding unificationism, which is to say the integration of cytology and genetics into the chromosome theory of Mendelian heredity. Whereas cytology was a field which used the microscope to try to find where in the germ cells the hereditary material was located, genetics was a field which hypothesized the existence of (what came to be called) genes to account for the heritability patterns encountered in the breeding of different species. Each field had questions which they were unable to answer from within their field. Cytologists "had no way of investigating the functioning of chromosomes in producing *individual* hereditary characteristics", while geneticists were "unable to answer the question: where are the genes located?" (p.52) These questions gave rise to the integration of both fields to create the interfield chromosome theory of Mendelian heredity, which drew on cytology to account for the mechanisms involved in heredity, and drew on heredity studies to account for the inherited characteristics. This new theory allowed for solutions to some unanswered questions from both fields, but also opened the door to novel predictions, such as the proposal that some genes would be linked in inheritance, explaining why some patterns of inheritance seemed unusual (p.53).

We see through this example a little more concretely how the fields of cytology and genetics came to be integrated. Both fields were preserved, and a new theory was created which incorporated information, methods and concepts from both fields to allow for a better and more comprehensive understanding of the phenomenon under study. Darden and Maull point out that their approach is decidedly not reductive (nor anti-reductive) since the chromosome theory is not seen by anyone as reducing cytology to genetics, nor vice versa (p.60-61). It can be seen instead as a form of integration, bringing together various fields which are interested in similar enough phenomena that they can jointly create a new theory which not only answers questions which each on its own cannot do, but also creates avenues for new questions and eventual answers.

Darden and Maull's article goes further than unificationism, by giving a more concrete understanding of just how the different fields can influence one another through the creation of an interfield theory. And it is an alternative to reductionism insofar as neither theory is understood to be reduced to the other; instead, they work together to create something new. The four ways in which interfield theories make explicit the relations between fields, namely through the specification of the location, nature, structure, or cause and effect of an entity or process from another field, are valuable additions to the integrative toolkit. And these relations they have identified can presumably be carried out in multiple ways, depending on the fields involved and the phenomena at stake.

That being said, many integrations do not rely on the construction of a new theory, and instead are limited to smaller epistemic units, which can result in the creation of novel epistemic units, such as explanations, models, data, etc.. The next section explores more specifically these issues.

5.3 Explanatory integration

Recent work in philosophy of biology has shifted away from discussions about theory integration to move instead to more local and specific integrations, which yield novel explanations. These ways of approaching integration highlight the variety of methods and results which are possible when focusing in on the details of how fields and disciplines can cooperate to produce explanations together.

5.3.1 Explanatory reduction

Just like integration itself, explanatory reduction has been seen as an alternative to intertheoretic reduction. By applying to smaller epistemic units than entire theories and abandoning the view that reduction could lead to a unity of science through a Theory of Everything, explanatory reduction has far more modest goals, and is therefore adopted more widely. As it applies to many different parts and types of explanations, there is a great variety of explanatory reductionist models, or proposals. Many of these have been touched upon in chapter 2, namely Kim's (1999) functionalist reductionism, which can be seen as a form of explanatory reduction, since it applies not to theories, but to functions and the constitutive parts which bring about those functions. Other kinds of explanatory reduction were also covered in section 2.1.1, including Waters' (1990) way of reducing phenomena through the difference-making principle, or Kauffman (1971) and Wimsatt's (1976b) proposal that reduction can proceed by finding the most relevant causal or constitutive components in a lower level, accounting for the phenomenon. Oppenheim and Putnam's (1958) take on reductionism could also be seen as a form of explanatory reductionism since they understand a theory to be reduced when the reducing theory successfully explains all of the phenomena explained by the reduced theory. In other words, their theory reduction emphasizes the role of explanations rather then theories. Many of these ideas have been carried over into a more fully fleshed account of multi-level mechanistic explanation, as will be explained in a later section.

One approach that is notable in its originality is Weber's (2005) research on experimental biology. He proposes that there are cases where "experimental biologists directly explain biological phenomena by applying laws and theories from physics and chemistry to the specific kinds of systems that they study." (p.49) The resulting explanation is reductive insofar as it uses more fundamental, lower-level laws, but maintains the reference to the higher-level structures or phenomena. While most explanatory reductionist approaches rely in some sense on an explanation of the whole in terms of its parts, Weber's reductionism is novel insofar as it relies only on lower-level fundamental laws being used in explanations of phenomena at a higher level, suggesting that we have not exhausted the ways that explanatory integration can bridge higher and lower levels.

5.3.2 Mitchell's three types of integration

As reviewed in detail in chapters 2 and 3, Sandra Mitchell proposes a form of pluralism which sees integration in a positive light, though with some limitations. Mitchell (2003, p.192-194) identifies three ways in which the integration of explanations can happen. The first is mechanical rules, which apply when establishing the "joint effects of independent additive causal processes" (2003, p. 192). This is the most straightforward way of integrating explanations, and the most concrete method for integration seen in this chapter so far. Each explanation can simply be added to one another, in the same way electromagnetic and gravitational vectors can be added. According to Mitchell, the limit of such an approach is that it is not applicable to nonlinear or nonadditive interactions, since they cannot be simply summed up. It is not clear however why mechanical rules would be limited in this way; Mitchell explains that when nonlinear interactions are in play, emergent phenomena may come about, but as was shown in chapter 2, her characterization of "scientific emergence" through dynamic complexity is overstated, and does not lead to intractable phenomena. While it is a truism that nonlinear interactions are not merely additive, it is not the case that this will make them impervious to integration, and to some form of mechanical integration. Indeed, the various integrations made possible through mechanistic explanations is discussed at greater length in the next subsection, with the help of Craver's (2007) interlevel mosaic.

The second is local theoretical unification, which develops models that encompass many features of complex processes. According to Mitchell, this looks a lot like Kitcher's unificationism, but with certain restrictions that make the unification stop short of universal unification (since, after all, Mitchell is a pluralist). Those restrictions have been covered in detail previously (chapters 2 & 3), so I will not be rehearsing them here. Mitchell uses the example of Leibold et al.'s (1997) work on the modelling of trophic structures of ecosystems, where they attempt to integrate "bottom-up" and "top-down" perspectives into a single coherent model. Much like Darden and Maull's (1977) interfield theories, the new model attempts to use findings and methods from both perspectives to create a new theory which will better account for the diversity of phenomena observed. And just as is the case with interfield theories, how this integration can be carried out will depend on the context.

And the third is explanatory, concrete integration, which is a local integration of multiple partial explanations of a specific, complex phenomenon, but is not generalizable to other phenomena. This was seen in chapter 3 concerning the division of labour in social insects: many models are created which only partially account for any given situation, meaning that their integration will be necessary when applied to a concrete case. The resulting explanation will be specific to the case at hand, but more importantly, even the methods used for integration can be unique. In discussing this particular type of integration, Mitchell gives the example of the modelling of the complex ecosystem of Lake Erie, one of the Great Lakes of Canada. The model needs to incorporate information from (at the very least) dozens of factors such as the behaviour of the different crustaceans, of the different fish, the impact of chemicals, and the local, seasonal variation in solar radiation. “Features of the method of integration of these multiple factors for a single lake may be local to Lake Erie or may be symptomatic of a class of situations, but are unlikely to be global and algorithmic.” (Mitchell, 2003, p. 194) In this respect, Mitchell’s explanatory, concrete integration is best seen as the characterisation of a class of situations where integration can be beneficial, rather than an actual “type” of integration.

Mitchell’s three types of integration are not to be understood as an exhaustive list. She herself reviews some other proposals for integration, among others Darden and Maull’s interfield theories and Kitcher’s unificationism. Considering her integrative pluralist position, there is good reason to believe that she would see other types of explanatory integrations in a positive light.

5.3.3 Interlevel mosaic

Mechanistic explanations have been mentioned throughout this thesis as a notable way of integrating explanations. This and similar ideas have been suggested by many researchers in the last decades (see among others Bechtel & Richardson, 2010b; Delehanty, 2005; Glennan, 2002; Machamer et al., 2000). Craver’s *Explaining the Brain: mechanisms and the mosaic unity of neuroscience* (2007) is one of the most fleshed out recent descriptions of how these integrations can happen. Though Craver looks specifically at research in neuroscience, it is quite clear that this also applies to mechanistic explanations in biology in general, and behavioural biology more specifically.

Craver’s general idea is that, within neuroscience, not only is it the case that explanations *can* span multiple levels, but they *must* if they are to be a good explanation (p.10). There are therefore links that must be made between explanations at various levels in order to appropriately explain a given phenomenon: “One establishes interlevel explanatory linkages by describing mechanisms, by identifying the appropriate entities and activities, by showing how they are organized together, and by showing, most importantly, that each of these features of the mechanism is relevant to the explanandum phenomenon.” (p.267) Craver’s contention is that these interlevel explanations act together and interact one with the other to produce a more complete picture of the phenomenon, and that in fact, without the contributions from multiple levels, the description of the mechanisms would be incomplete.

Neuroscience is, by its very nature, a unification of many different fields. But the relationship between the fields is not one of reduction:

> It is not the case that theories at one level are reduced to theories at another. Rather, different fields add constraints that shape the space of possible mechanisms for a phenomenon. Constraints from different fields are the tiles that fill in the mechanism sketch to produce an explanatory *mosaic*. (p.18-19)

These constraints can be intra-level, meaning that explanations from one field will restrict what is considered a plausible mechanism for all fields concerned with that level (p.247). They can also be interlevel: "Upward-looking Interlevel integration involves showing that an item is a component in a higher-level mechanism. The downward-looking aspect of Interlevel integration involves describing lower-level mechanisms for a higher-level phenomenon." (p.257) According to Craver, a phenomenon is explained once the mechanisms within a level, and across levels, are integrated.

Craver's illustration of this kind of integration is the work that was done linking long-term potentiation (LTP) with learning and memory. Contrary to what was expected by reductionists, the discoveries were made at multiple levels and through various discoveries, with fields and explanations at various levels influencing one another. Most telling is Bliss, Gardner-Medwin and Lomo's (1973) article which argues for the relevance of LTP to learning and memory, calling on results from multiple fields, including experimental physiology, biochemistry, psychiatry, physiology, and computer science. All these explanations of different aspects of different parts of the overall mechanism relevant to learning and memory, stemming from other fields, shape the possible mechanisms of LTP, through constraints on what the likely nature, structure, or role of LTP is in learning and memory. This leads to an understanding of LTP as a constituent of a larger mechanism: "As of 1973, LTP was no longer proposed as identical to or an example of memory, but rather as a component of a multilevel memory mechanism." (Craver, 2007, p. 243) And it is this integration across fields which makes the resulting explanation so successful.

Craver not only incorporates the common notion that mechanistic explanations focus on the constitution of the mechanisms producing the phenomenon under study, but he also adds the notion of "constraint", highlighting an interesting way that explanations can interact.

5.3.4 Transitory integration

Ingo Brigandt (2010) proposes that integration can happen when different fields join forces to solve a particular problem, in a particular way. In doing so, he positions himself against both reductionistic approaches, as well as Darden and Maull's (1977) interfield theories approach. Contra reductionism, Brigandt suggests that explanatory integrations show that it is not always the lower levels that are taken as more explanatorily fundamental. And against interfield theories, he proposes that "successful integration may result from various smaller epistemic units—individual methods, concepts, models, explanations—being linked in an appropriate fashion" (2010, p.22), creating smaller, and more ephemeral, integrations between fields, as opposed to a full-fledged interfield theory.

Building on work done by Allan Love (2003, 2006, 2008), Brigandt examines research in evolutionary developmental biology (evo-devo) which focuses on the origin of novelties. A novelty, or evolutionary innovation, refers to a morphological feature which appears in a lineage, despite the fact that it is not found in an ancestral species. The appearance of feathers or the vertebrate jaw are examples of novelties, and pose a challenge to evolutionary theory, since while it can account for the selection among competing phenotypes, it cannot account for the appearance of novel phenotypes. Brigandt proposes that the explanation of novelty can be understood as a "problem agenda" (2010; see also Love, 2008) consisting of a set of related questions which cannot be addressed through a single field, as well as criteria for the adequacy of an explanation. In this case, evo-devo must marshal explanations from phylogeny, paleontology, ecology, biogeography, and of course developmental biology to account for the historical, geographical, environmental, and genetic aspects the explanations require.

Arguing against reductionism, Brigandt's contention is that it is developmental biology which carries the explanatory force (2010, p.9), and not a more fundamental field. The crucial question for an explanation of novelty "is how genotypic variation translates into phenotypic variation—which is the domain of developmental biology" (p.10) as opposed to the more fundamental fields interested (for instance) solely in genetics. Accounting for genetic variation is not seen as the problem in need of solving; rather it is how the genotypic variation gets translated into the phenotypic variation which can explain the appearance of novelties. But Brigandt points out that this priority of developmental biology is contingent on the problem agenda, and could very well be different for another problem; in other words, no field is intrinsically more important than other; they are only instrumentally better with respect to a given issue (p.16), as determined by the criteria of explanatory adequacy (p.17).

Now, while Brigandt highlights the different contributions that each field can bring to the table, the integrative aspect of a desired resulting explanation is somewhat vague. Brigandt calls on the necessity of "combining" different fields (p.15), giving certain examples, but leaving the reader with few clues as to how this is to be understood in a more general manner. Similar to Craver's (2007) mosaic unity, one gets the sense that each field introduces constraints that each other field must take into account. For instance, developmental biology can specify existing constraints in ancestral species, limiting the possible phenotypes, as well as describing how those constraints could have been broken (Brigandt 2010, p.11-12). Paleontology, for its part, "adds a historical-temporal scale to phylogenies", constraining hypotheses about species relations (p.8). Brigandt's important contribution is in highlighting that integrations can be transient: "it may be sufficient for a genuine explanatory integration of disciplines to relate and integrate items of knowledge from traditional disciplines solely for the purposes of a specific problem" (p.19). The result is not a new theory, or even a new field, but only a punctual bringing together of epistemic units for the resolution of a particular problem through the creation of a novel explanation.

In sum, Brigandt (2010) argues for a view of explanatory integration which conceives integration as a transient unification of epistemic units from various disciplines, focused on a specific problem or epistemic goal. In this respect, the idea of permanent and stable unification of disciplines is discarded, and which disciplines or epistemic units are considered more fundamental will vary with the problem pursued.

5.3.5 Data integration

Though not a type of *explanatory* integration per se, another type of integration which must be mentioned but cannot be covered in full is data integration. As pointed out by O'Malley and Soyer (2012), data integration and explanatory integration are distinct. Whereas explanatory integration "refers to both the synthesis of different explanations as well as the import of explanatory (and predictive) models from other research into a specified domain of inquiry" (p.61), data integration concerns more specifically the bringing together of datasets from various sources, with the objective of forming a body of information "that can be treated as a unified whole" (p.61). Thanks in part to the rise of data-mining techniques, science increasingly relies on large datasets, posing a challenge for efforts to integrate them, due to the various ways that information can be parsed, measured, and encoded. Leonelli (2013) discusses how data "curators" take up this challenge, attempting to sort and label information in databanks such that it will be usable by a wide variety of researchers. This is no simple feat, involving more than mere aggregation of information; rather, data must be "carefully selected, formatted, classified and integrated in order to be retrieved and used by the scientists who may need them." (p.506) To facilitate this task, data curators are developing standardised ways of storing the data, including the specification of the way the data was produced in the first place, as well as confidence rankings of the information gathered (p.507). Leonelli (2013) identifies three ways that this data integration can be carried out.

Inter-level data integration allows data from different levels of organisation to be used across levels. Research on model organisms, for instance, such as *E. coli*, *C. elegans*, or *Arabidopsis*, is done by many fields interested in many different scales, from genes all the way up to phenotypes and interactions with the environment. Each of these fields inputs data into a central databank, in the hopes that it can be used by other fields, with the objective of understanding "the biology of the organism as a holistic whole rather than as an ensemble of disconnected parts" (p.507). To make sure that this data is usable by all the different fields, data curators work on software and modelling tools which allow for better visualisations of the data. They also attempt to create standardized keywords and concepts through consultations and meetings with researchers in all fields, facilitating the use of the database for all levels of inquiry (p.508).

Cross-species integration of data raises many of the same issues, but

> this task is made even harder by the terminological, conceptual and methodological differences between communities working on different organisms, as well as differences in perceptions of what counts as good evidence and the degree to which specific traits are conserved across species through their evolutionary history. (p.510)

To build bridges between the data for different species involves not only aligning concepts and ontologies as is the case with inter-level integration, but also an iterative process of comparing data and results of experiments as applied to each species every step of the way, to ensure that the integration is carried out successfully. Leonelli gives the example of flowering time in both the *Miscanthus* and *Arabidopsis* plant species, and how it is not only the data which is shared, but also the results of the various experiments aimed at modifying the flowering time (p.509).

The third kind of data integration Leonelli calls "translational integration", which arises when a pressing social need calls for the integration of data from various scientific and non-scientific sources. The characteristic aspect of this integration is that it rests not on the search for knowledge for the sake of knowledge, but the search for efficient interventions relating to specific problems. This implies negotiations with stakeholders, which comes to determine what type of data is most apt to produce interventions within a desirable timeframe. Leonelli gives the example of a mold infestation in a forest in the UK, which prompted meetings between the Forestry Commission, private landowners, and scientists, among others. Data from multiple sources came to bear on the problem and needed to be integrated, including aerial maps of the infestation, genome sequencing, linkage analyses, PCR-based diagnoses, as well as the “possible ecological, economic and societal implications of each mode of intervention under consideration” (p.511). Leonelli proposes that this type of integration is distinct because of its commitment to rapidly producing results that have an effect on human wellbeing, though the form of integration may not differ from the others.

While data integration itself is not synonymous with the integration of explanations, it obviously can have great implications regarding the possibility of integrated explanations. As O’Malley and Soyer (2012) point out, it is not the case that data integration will “follow sheepishly behind theory” (p.62); rather, in many cases, it is the integration of data which opens the door to the integration of other epistemic units, such as explanations, models, and even potentially theories. It is also very likely that the integration of data could lead to novel explanations which cover phenomena which were not originally intended in any of the original datasets, showing once again how data integration can be a starting point for expanded understanding of biological phenomena.

Data integration, understood as a broad technique, is in full development, with the rise of bioinformatics, data mining techniques, ever-expanding datasets, as well as the rise of AI-assisted parsing of the information. This suggests that many of the challenges facing data integration will eventually be overcome. This section has by no means given justice to the wealth of research and projects carried out in this domain, but gives a glimpse into the potential of data integration as relates to explanatory integration more generally.

5.4 Example: integration of mechanisms

I will now look at two examples of integration in action, the first centered on a specific phenomenon, and the second relating to the more general idea of the purported independence of explanations. Both cases highlight how integration is a tool which can be understood to be breaking down explanatory pluralism by joining together various epistemic units.

This first example of integration relies on work by Kenneth S. Kendler, and shows how multiple different explanations of various aspects of a given phenomenon can be integrated to better explain that phenomenon. Kendler is a noted psychiatrist who has worked on the genetics of psychiatric and substance abuse disorders. He uses methods from psychiatry, behavioural genetics, and molecular genetics to attempt to understand the causes of disorders. He brands his approach as pluralistic, arguing that no single approach to a given phenomenon will be superior, and that the best explanations will be the result of the integration of multiple explanations.

5.4.1 Mechanisms of alcohol dependence

Kendler (2008) proposes to abandon the traditional ideal of finding fundamental laws for psychiatry, and instead turn towards a "mechanistic approach" which attempts to integrate the causes from multiple levels into a coherent whole. His approach considers that the best way to study the psychological mechanisms for a given disorder is through a decomposition of the multiple causal pathways which can lead to it, with different approaches attempting to explain the various parts in isolation. This step is then followed by a re-composition through the integration of the multiple explanations. While this can be relatively straightforward in simple systems where the component parts are easily isolable and have additive effects, the same idea applies to systems which incorporate causal loops: "more complex mechanisms can be much more challenging, but the basic principle still holds." (p.3) The objective of integration is a better understanding of the phenomenon: "Ultimately we face the task of figuring out how the entire system works" (p.3). Integration then is the tool which permits a more complete understanding of the complex, multi-level systems which account for psychiatric disorders.

Kendler uses alcohol dependence as an example of this approach. He first describes what is known about the multiple mechanisms which enter into the disorder, including genetic risk factors, biological factors such as alcohol metabolism, personality traits, environmental factors such as peer substance use, and cultural factors such as the acceptability of public drunkenness. He then shows how in this case at least, the genetic effects are not aggregative, and instead are included in causal loops. For instance, people who have genetically reduced sensitivity to alcohol's effects are more likely to drink frequently, and as such have a greater risk of developing a dependence: "So genes influence subjective ethanol effects, which influence alcohol expectations, which in turn loop into the environment, influencing consumption patterns, which in turn affect the risk of alcohol dependence." (p.4) Other ways in which the multiple causal pathways affect each other are highlighted, such as the fact that risk-taking adolescents tend to seek out similar peers, creating an environment for themselves which promotes drug-taking behaviours; or how the offspring of heavy-drinking parents, who likely have a genetic predisposition to alcohol dependence, can consciously decide to avoid alcohol because of their awareness of the risk (p.5). These multiple causal pathways are first studied in isolation, but the best understanding of the causal pathways leading to alcohol dependence will necessarily integrate all of the causes, showing how each affects the others. Though this may become complex when feedback loops are involved, it is nevertheless tractable in principle, and remains the ultimate objective. Kendler expects the mechanisms studied by psychiatrists to be riddled with such cross-level and intra-level interactions and causal loops, highlighting the need for integration.

5.4.2 Kendler's (methodological) pluralism

Kendler defends what he calls "empirically based pluralism" (2012), and is cited by both Mitchell (2009) and Longino (2013) as an example of a scientific researcher defending pluralism. His main target is reductionistic approaches, which according to him would privilege a single type of mechanism, or a single type of analysis as the only good way to arrive at a successful explanation; instead, he rightly contends that "it is not possible *a priori* to identify one privileged level that can unambiguously be used as the bases for developing a nosologic system." (2012, p. 12) However, a close look at his approach shows that the pluralism proposed is neither fragmentary nor insular, and in fact is best understood as not even being a form of explanatory pluralism as described in chapter 1 of this thesis. The pluralism proposed is in fact more appropriately seen as a form of methodological pluralism, encouraging research on a given phenomenon using multiple different methods, such as those from genetic approaches, biological approaches, socio-environmental approaches, etc. But despite the methodological pluralism, there is no fragmentary or insular pluralism since Kendler emphasizes that all these methods will yield explanations which can be integrated one with the other.[17] Because both fragmentary and insular pluralism rely on the fact that certain explanations will be impossible to integrate—either precluding the unity of science for fragmentation pluralism, or allowing for conflicting and incommensurable explanations for insular pluralism—an approach which sees no limits to integration cannot be understood as pluralistic in either of these senses. Indeed, Kendler does not propose any limits to the possibilities of integration, and

[17] The only explanatory pluralism that could fit the bill is a form of type pluralism, under the assumption that the different methods could yield different types of explanations, though (as is typical with type pluralism) this is never explicitly addressed by Kendler.

suggests instead that integration ought to be the ultimate objective for any successful account of a psychiatric disorder.

This example shows how explanatory integration can and does get used to account more fully for a single phenomenon. Kendler's approach underlines how the integration of explanations of causal mechanisms, while not always additive, and seldom simple to do, is not only possible, but also the ultimate objective for a successful account of the phenomenon under study.

5.5 Example: Tinbergen's four questions

This second example widens the focus to look not at a particular phenomenon, but instead at ways in which integration builds bridges between purportedly independent explanations of any one phenomenon, showing how pluralism does not hold in the face of ongoing research.

Biological sciences, and behavioural biology more specifically, has a history of proposals aiming to highlight the relative independence of different explanations. These proposals are often gathered under the wider denomination of the "levels of analysis" account (not to be confused with the "levels of organisation" or "levels of mechanism" mentioned elsewhere in this thesis). Ernst Mayr (1961) famously proposed that explanations in biology can evoke two types of causation: ultimate and proximate. In the case of behaviour, proximate causation asks the "how" questions, calling on the immediate mechanisms that brought about the behaviour, such as development, the environment, etc. Ultimate causes are historical, and explain "why" the behaviour is there to begin with, typically calling on natural selection, genetic drift, or other such mechanisms. Tinbergen's four questions (1963) subdivide the ultimate and proximate explanations into two further levels. Ultimate explanations can call on either evolutionary origins or current adaptive value, whereas proximate explanations look at ontogenetic processes or mechanisms. Sherman (1988) added that mechanisms could be further subdivided into those that target physiology and those that target cognition.

These ways of classifying the questions asked by biologists have often been interpreted as a form of explanatory pluralism, wherein the answers to each of the questions are relatively independent from each other. Indeed, Sherman claims that "every hypothesis within biology is subsumed within this framework; competition between alternatives appropriately occurs within and not among levels" (1988, p. 616). Tinbergen himself however encourages their "integration" (1963, p. 411), and more and more researchers are calling for a new understanding of the “levels of analysis” which instead recognizes just how interrelated they all are (Laland et al., 2013, 2014).

In a similar vein to Longino’s work in *Studying Human Behavior* (2013) wherein she highlights the scopes, methods, and assumptions of various approaches, I will describe how and why behavioural ecologists have long made use of assumptions in their methods, to facilitate their research. These gambits—as they are called in the literature—are sometimes implicit, sometimes explicit assumptions about certain aspects which are considered to be outside of the purview of the approach, as delimited by the levels of analysis account. However, as will become obvious, the assumptions can affect the resulting explanations. The interest of this example is to show how making these assumptions explicit opens the door to the integration of explanations stemming from other approaches, namely molecular biology and psychology, and how this work breaks down the purported independence of the levels of analysis.

5.5.1 The phenotypic gambit

Behavioural ecology is the study of the evolutionary basis of animal behaviour due to ecological pressures. In this respect, it traditionally looks to uncover the two ultimate causations, namely the current adaptive value of a trait, and its evolutionary history. But some authors have highlighted the fact that these questions rely on assumptions regarding the proximate mechanisms, the recognition of which can have important repercussions on the ultimate explanations.

Researchers begin by observing certain behaviours in a given population, then providing an “economic analysis of costs and benefits” (J. R. Krebs & Davies, 1993, p. 48) using broad environmental categories in simplified models, to show how (or if) the behaviour is optimal with respect to fitness for a given environment, through what is called “optimality models”. For instance, Krebs & Davies (1993, pp. 49–53) review Kacelnik’s (1984) experiments showing the optimality equations describing the foraging behaviour of starlings bringing back invertebrates (i.e. food) for their nestlings. Starlings become less and less efficient at hunting for prey as their load increases, meaning that there is a point at which it is more efficient to return to the nest and empty their load, rather than continue hunting at a diminished speed. Furthermore, as Kacelnik’s calculations of the diminishing rates of return predicted, the size of the maximum load is related to the distance from the nest: the further the starling is, the more the load brought back to the nest will be heavy. Part of the challenge of doing such an analysis is finding which factors are indeed counted as costs and benefits. For instance, it was discovered that while the foraging behaviour of starlings is related to the net rate of energy delivery to the nestlings and not on the energetic efficiency of the parents' foraging, that of bees is based on the energetic efficiency of the individual nectar-gatherer, and not merely on the gross quantity of nectar brought back to the hive (J. R. Krebs & Davies, 1993, p. 54).

When attempting to determine the optimality of a given behaviour, behavioural ecologists will often resort to the "phenotypic gambit", a term coined by Grafen (1984). To carry out the phenotypic gambit, the researcher (i) elaborates a strategy set, which is "a list or set of (perhaps all) possible states of the character of interest." (p.63), along with (ii) a rule for determining the success of a strategy, such as number of offspring or inclusive fitness. This allows for what should be the full spectrum of (realistically) possible behaviours, and a way of quantifying their relative success. Once this step is complete,

> the phenotypic gambit is to examine the evolutionary basis of a character as if the very simplest genetic system controlled it: as if there were a haploid locus at which each distinct strategy was represented by a distinct allele, as if the payoff rule gave the number of offspring for each allele, and as if enough mutation occurred to allow each strategy the chance to invade. (Grafen, 1984, p.63-64)

The inputs for the calculations are idealized, under the assumption that, given enough time (Hammerstein, 1996), each behavioural strategy is just as possible as any other, thus allowing researchers to effectively study the phenotypic traits and their evolution without recourse to the genes themselves (Huneman, 2014, p. 168). As Grafen points out, "taken literally, the gambit is usually false: few species studied by behavioural ecologists are haploid" (p.64). Nevertheless, Grafen argues it is a necessary gambit insofar as behavioural ecologists cannot simply wait until geneticists validate every study in population genetics, not to mention that most of the time, the gambit works (c.f. Réale & Festa-Bianchet, 2000).

On the other hand, some researchers have shown instances where the gambit fails. For instance, Hadfield et al. (2007) have demonstrated that a shared natal environment can sometimes lead to similar phenotypes in birds, such as plumage colour and skeletal traits, obscuring genetic differences. This was done using methods not typically associated with behavioural ecology: the researchers experimentally manipulated the environment of blue tit chicks by swapping them with other nests at birth, then, "using a methodology developed for the comparison of geometrical subspaces, [they tested] whether the phenotypic gambit can be made" for the group of traits under study (p.550). Through the integration of methods from other approaches, their results showed concretely how similar phenotypes can be the result of dissimilar genetics, exposing instances where the phenotypic gambit is strictly speaking false.

Bull and Wang (2010) for their part used experimental evolution (laboratory settings made to explore evolutionary dynamics) and genetic tests on model organisms (in this case certain microbes) to test the gambit. Because model organisms have well-researched genetics, they were able to do away with the phenotypic gambit and keep an eye on the genetics while manipulating the environment to test optimality models, so that "the genetic pathways of evolution [could] be identified." (p.2). Their research showed that there are times when the optimality models fail precisely because of an incompatibility between the assumed and actual genetics: the phenotype assumed to be optimal was not selected for because of constraints at the genetic level. However they also showed that there are times when the phenotype *is* selected for, but not through the expected genes: "Experimental adaptations that led to the expected phenotypic changes by mutations outside the candidate regions may provide the means for reshaping our understanding of those classic genetic systems." (p.17) What this means is that the integration of genetics into optimality models lends more credibility to both: "inclusion of genetics enhances the understanding attained with optimality models, and experimental tests of optimality can inform genetics" (Bull and Wang, p.2). Research such as this shows the possibilities and value of integrating genetics into behavioural ecology.

Indeed, this development of molecular behavioural genetics, though ostensibly uncovering proximal mechanisms responsible for given behaviours, can in fact have a great impact on answers to questions relating to adaptation and evolution. First, regarding adaptive value, molecular genetics reveals explicitly what the phenotypic gambit glossed over: there is (almost) never a simple correspondence between genotypes and phenotypes, and not all phenotypes have an equal chance of being realized because of the way they may be constrained by the underlying genes, which in turn are constrained by ultimate explanations. Second, it informs questions relating to ontogeny since it specifies some of the building blocks necessary for development, namely the genes, and could eventually highlight environmental interactions within ontogeny. And third, molecular genetics can cast doubt on evolutionary hypotheses or suggest new avenues of inquiry, such as when the genetic research pinpoints certain historical times of strong selection. For example, work on the FOXP2 gene and the dating of its spread in human populations can suggest strong selection for this gene involved in language production and comprehension, sometime during the last 200 000 years (Enard et al., 2009; Preuss, 2012), implying that evolutionary explanations must account for that particular environment of evolutionary adaptedness. These links between molecular biology and behavioural ecology show how ultimate and proximal questions can be linked through integration.

5.5.2 The behavioural gambit

A second gambit has been pointed out more recently: the “behavioural gambit”. Certain behavioural ecologists have suggested that looking at the underlying mechanisms of the behaviours under study can yield fruitful insights (Giraldeau & Dubois, 2008). Fawcett et al. (2013) propose that the gambit in this case is that each behaviour taken as an input of an optimality equation is usually implicitly or explicitly assumed to stem from a single physiological or cognitive mechanism specific to that behaviour, and as such, is the result of natural selection rather than individual adjusting to its environment over the course of its own lifetime. The fact that it is an *assumption* is most striking when considering behaviours that are learned over the lifetime of the individual, rather than the direct result of natural selection acting on genes. As the authors point out, this neglect of the fact that certain behaviours are learned also implies that the limitations in what can be learned are also obscured. Instead,

> this approach invokes an additional, unstated assumption: that the psychological mechanisms underlying flexible decision making do not constrain the expression of adaptive behavior and allow animals to reach the optimal solution to a given problem (Fawcett et al., 2013, p. 2).

Just as was the case with the phenotypic gambit, this assumption leaves unconstrained the possible behaviours, as opposed to recognizing that certain strategies may be constrained by the cognitive capacities of the individuals and the species. The problem arises when optimality models do not fit with the observed data, indicating that there may be a problem with the assumptions.

Acknowledging this gambit and instead trying to reveal the underlying mechanism—or mechanisms—can lead to new discoveries, and better optimality models. For instance, Kacelnick (2012) argues that rather than assume that all behaviours are the result of specific, heritable mechanisms, many are likely to be the result of a more general learning mechanism. He uses the example of optimizing foraging to show how the great variety of environments, food sources and resource types make the idea of specific mechanisms unlikely: to account for the behaviour of an individual in such a wide variety of situations, the organism would need “a large library of heuristics as well as a large number of subsidiary rules to select the optimal parameters and then rank the performance of each rule.” (p.30) This is seen as a problem because it is assumed that the mechanisms ought to be more parsimonious so as not to overburden cognition. And this problem is compounded with the framing problem: how does the individual recognize which rule is appropriate for each situation? Rather than attempt to account for such complicated cognitive machinery, Kacelnick reaches out to psychological research to propose that the phenomena are better accounted for through *reinforcement learning*, a general learning mechanism which pushes the organism to repeat successful strategies, and avoid unsuccessful ones (p.31-32). By focusing on the mechanisms underlying the behaviour, one can see why not all behaviours have an equal chance of being selected (since some may simply not be attainable through that mechanism), and why some sub-optimal strategies can surface. This is because general mechanisms such as reinforcement learning will not optimize for every specific situation, and may stop at a satisfactory solution (known as satisficing) in many cases (Fawcett et al., 2013, p. 6).

In sum, recognizing and challenging the behavioural gambit can lead to the realization that a single mechanism can account for a variety of behaviours, which can explain why certain optimality models fail to capture the observed behaviours.

5.5.3 Integration breaks down the four questions

The interest in finding cases where the phenotypic and behavioural gambits fail is not to show that optimality models need to be discarded, but rather to demonstrate two important points about the possibility and value of integration.

The first is that the recognition of the gambit opens the door to integration, and that integration is indeed possible. Had the gambits not been recognized as such, and pointed out by various researchers, the assumptions may have stayed implicit, or if (as in the case of the phenotypic gambit) they were explicit but neglected, there may never have been research carried out on the limitations created by those assumptions. And as shown through the examples above, the research has been successful at integrating explanations from various approaches, and replacing gambits with explanations. When it comes to the phenotypic gambit, the developments in genetic analysis, as well as that of disciplines such as evolutionary developmental biology and epigenetics have shown that the correspondence between phenotypes and genes is not always straightforward, highlighting ways that the gambit can be overcome. As for the behavioural gambit, the integration of explanations stemming from experimental psychology replace the simplifying assumptions about specific behavioural mechanisms.

The second important takeaway from this discussion surrounding gambits is how the integration of explanations from other approaches enriches our understanding of behavioural biology. Without looking into the genetics underlying the phenotypes, certain behaviours—or the absence of certain behaviours—could have stayed a mystery. The same goes for the cognitive mechanisms underlying the behaviours: calling on cognitive mechanisms allows for more precise and biologically realistic explanations (Fawcett et al., 2013). The gambits show that although the ultimate causation responsible for the behaviours under study *can* be researched by neglecting the proximal mechanisms, the latter could still have an effect on the validity of the ultimate explanations. Indeed, although these research strategies may be fruitful much of the time, their recognition as a gambit demonstrates that they are only a partial explanation, which could be changed or eventually shown to be wrong once the proximate mechanisms are brought to light. In other words, the explanations relating to the behavioural adaptations of populations are constrained by the proximate processes, despite the fact that those constraints often may not change the results of behavioural ecology research. The independence of the types of explanations put forth by Tinbergen's four questions is thus called into question.

And this is the more general point to be made regarding the gambits: Tinbergen's four questions are not in fact independent. Though Tinbergen himself (and likely most, if not all, behavioural ecologists) knew that the independence was only relative, new research has shown how assumptions made by the various involved disciplines can be broken down, and replaced by explanations from other sources, leading to integrative accounts which more accurately portray the phenomena. This has led Kevin Laland and many of his collaborators to argue that in light of these new discoveries, Tinbergen's four questions ought to be abandoned in favour of the recognition that all the processes involved in all explanations are influenced by, and even in feedback loops with, other processes at other levels of analysis (Laland et al., 2013, 2014; Rittschof & Robinson, 2014). Instead, biologists should recognize that processes which seem relatively independent are exceptional cases where the feedback loops have negligible effect (Laland et al., 2011).

The levels of analysis accounts of explanations in biology have been a useful tool for sorting out precisely what researchers are trying to demonstrate, but it would be a mistake to understand the divisions as anything more than useful heuristics. Each level can have an impact on the others, and the examples given above show how integration makes it such that indeed they often do.

5.6 The integrative toolkit

This chapter has proposed a survey of many of the kinds of integration found in the literature. Some types of integration are understood to be possible for theories, including reduction and unification, or even the creation of novel interfield theories. Others apply more locally to explanations, such as explanatory reductions. Most explanatory integration concerns the development of new models or explanations which incorporate elements from various approaches. This can be done by showing the interactions of various causes—be they additive or non-additive—or through the recognition of constraints imposed by explanations coming from other fields or approaches. And finally, data integration is becoming an important tool for the sharing of information across laboratories and research teams, which comes with its own set of challenges. These integrative strategies, or conceptualisations, represent only those which have so far seen attempts at philosophical analysis, and there is little doubt that more fine-grained analysis of the scientific literature could reveal many more.

The picture that is revealed through this overview is that there are many ways of carrying out integration, and that new ways are constantly being developed. When researchers are faced with specific problems which they cannot address from within their approach, either because their predictions do not match the observations, or because they are unable to even make predictions, they will not stand idly by. Instead, they will seek out new methods, new explanations, and notably: seek to integrate them into their own work in order to solve the problem they are faced with. The development of these integrative strategies can be seen as the creation of tools for the integrative toolkit: each new way of carrying out integration could potentially be used by others, and adds to the pool of strategies. This ever-expanding toolkit represents the many ways which incommensurability can be broken down between approaches.

If we agree that integration is possible, does that mean that it is desirable? In short: yes, it is desirable, though it is not the only desirable avenue. In other words, integration will always be valuable for certain aspects of explanation, but not for all.

As was implied in many passages of this chapter, what integration allows is for a more complete picture of the studied phenomenon. As was shown most explicitly through Craver's mechanistic interlevel mosaic and Kendler's work on alcohol dependence, the integration of mechanisms uncovers causes at multiple levels and coming from multiple perspectives, and shows how they are related to one another. Without clarifying these relations among the causes, we will only ever have a partial view of the phenomenon, with many unexplainable observations. As Kendler suggests, any explanation of a complex mechanism "will require the integration of multiple explanatory perspectives." (2008, p.9) All other integrative strategies have the same ultimate objective: to give a better and more complete understanding of a complex phenomenon which requires contributions from multiple approaches. Whether it is through the integration of theories, through the creation of novel interfield theories, the transitory bringing together of epistemic elements, or the pooling of data, all forms of integration ultimately lead to the creation of new explanations, or the application of an old explanation to a explanandum phenomena, which add to our knowledge of the phenomenon at stake. In this respect, integration will always bring added value to explanations by replacing assumptions with explanations, black boxes with mechanisms, and shedding light where before there was none.

Of course, in practice integration is not the best strategy every step of the way. As many researchers have pointed out, integration becomes pertinent and possible only once sufficient research has been carried out in relative isolation on the various parts of the phenomenon under study (see among others Bechtel & Richardson, 2010b; Brigandt, 2013; Craver, 2007; Kendler, 2008). Even strong reductionists such as Nagel conceded that premature attempts at reduction could hinder the practical development of science (1961, p. 362). But this does not imply that the integration will be impossible, only that more research needs to be carried out before it is possible and desirable. And as biologists Laubichler et al. (2018) remark:

> there is value in striving for theoretical unification and integration, both for explanatory and for practical reasons and […] even if we do not reach our theoretical goals yet, the formal clarifications related to ontologies and data models that are necessary to connect different types of data and models at any scale are an important first step toward reaching this goal eventually. (p.8)

Thus, even in cases where research in isolation may be more fruitful, integration, and attempts at integration, remain a valuable enterprise. In any case, my point is not that integration is the best way to proceed in all situations; as will be argued later, my position is only that integration will always be in principle possible, and that fragmentation and insular pluralism are therefore misguided.

In sum, despite its typical association with pluralism, integration can be seen as a way of countering pluralist claims. Mitchell proposes that "the arguments [she has] given for expecting pluralism imply that the types of integration within science will also be varied and diverse. No single theoretical framework, no single algorithm, will suffice." (Mitchell, 2003, p.194) While it may be the case that her pluralism implies variety in integrative strategies, it is quite clear that the inverse does not hold; integration can instead be seen as a way of countering pluralism. This idea is also expressed by Laubichler and colleagues:

> we argue that […] explanatory pluralism is no longer adequate for several areas of biology because of (1) the data revolution and (2) the computational revolution within the life sciences that have brought the goal of theory integration within reach again." (2018, p. 7)

According to this view, new developments in data integration, among others, make integration at large more and more likely, casting doubt on pluralist ideals. For if it is possible to integrate the plurality of theories and explanations, then pluralism will not hold. Thus, while methodological pluralism may be valuable many steps of the way, explanations ought ultimately to be integrated in order to give the most complete explanation possible.

To showcase a novel contribution to the integrative toolkit, chapter 6 presents an attempt at opening the door to integration in behavioural biology, through the analysis of the concept of 'behaviour' as used by biologists. And as I will argue in chapter 7, the proliferation of integrative strategies suggests that integration will always be possible in principle, thanks to the creativity of researchers contributing to the integrative toolkit.

CHAPTER 6

INTEGRATION THROUGH CONCEPT GRADUALISM: DEFINING BEHAVIOUR

This chapter presents a concrete example of work done to build bridges between various approaches, with the aim of favoring integration.[18] The first step in most integrative strategies is to ensure that researchers are talking about the same phenomenon, ensuring that the epistemic units at stake really are about the same thing. One of the ways this can be done is by showing how concepts used in different disciplines are different or similar, through various techniques. A manual for interdisciplinary research describes one of these techniques in the following way: "Differences or oppositions in disciplinary concepts can sometimes be addressed when one extends the meaning of an idea beyond the domain of the discipline into the domain of another discipline." (Rutting et al., 2020, p. 44) This is what I have done here with the concept of 'behaviour', ubiquitous in many areas of biology, but rarely if ever defined, and often used in different ways, sometimes leading to conflicts and misunderstandings.

Many disciplines in biology explain the behaviour of organisms through various means. Yet the very concept of 'behaviour' is more often than not left unexamined by biologists, who tend to rely on a common-sense or intuitive understanding of what behaviour is—for instance a response to stimuli—and what phenomena ought to count as behaviour (Levitis et al., 2009). This has resulted in a lack of consensus regarding a definition of behaviour, as well as conflicting intuitions regarding classifications of phenomena, and ultimately a plurality of ways of formulating behaviour-related explanations.

The concept of 'behaviour' is central to—and could even be said to define—many scientific disciplines, in particular biological disciplines such as behavioural ecology, evolutionary psychology, ethology or neurobiology. The concept has played a major role in the history of these disciplines, with such landmark work as Darwin's research on animal and human behaviour (Darwin, 1859, 1871), the development of ethology (Tinbergen, 1963, 1976), as well as more controversial work on human behaviour such as sociobiology (Wilson, 2000), or even eugenics (Galton, 1883). The concept has also received much attention from philosophers, leading to work on the nature/nurture debates (Griffiths, 2002; Tabery, 2014), and in the philosophy of psychology (P. M. Churchland, 1981a; Dennett, 1998; Longino, 2001; Sober &

[18] Most of this chapter was previously published as Muszynski, Eric & Malaterre, Christophe (2019), Best Behaviour: A proposal for a non-binary conceptualization of behaviour in biology, in *Studies in the History and Philosophy of Biol & Biomed Sci* vol.79.

Wilson, 2003). This concept can be understood as an “epistemic hub”, and thus could “mediate between [the various approaches] and thus provide important opportunities for exchange of information and integration of causal explanations” (Kutschenko, 2011). Indeed, whether a phenomenon can be classified as a behaviour or not can have important repercussions for the type of research that can legitimately be carried out about the phenomenon, as well as telling us something about the explanations we can expect. As mentioned in the previous chapter (section 5.5.2) behaviours are typically thought to be the result of natural selection and therefore adapted for a particular environment, just as any other trait. Explanations of behaviour also typically call on a certain responsiveness to the environment which is usually not thought to be possible for other types of physiological or developmental mechanisms. So, for instance, when plant biologists use methods from behavioural ecology to produce explanations about plant behaviour (as does, e.g., the Cahill lab at the University of Alberta), one can wonder whether plants truly do exhibit behaviour, and therefore whether those explanations are legitimate. Or more modestly: what their definition of behaviour is, and whether it departs from the understanding of the rest of biologists. If, on the other hand, a certain phenomenon is not taken to be a behaviour for animals, then this could close the door to neurobiological research, under the assumption that it would be a purely developmental or physiological phenomenon, and not a cognitive one. The possibility of classifying a phenomenon as behaviour can thus have important repercussions for research, opening or shutting the door to the integration of theories, explanations, and methods across various approaches.

In this chapter I look at the use of the concept of ‘behaviour’ in biological disciplines, in an attempt to understand what it is about their object of study—if anything—that is common. This project is first and foremost descriptive, analyzing and explicating the concept of ‘behaviour’ through its use in biological research, and fits within the metaphilosophical framework of conceptual explication or engineering. Looking specifically at uses in biology, I circumscribe what it is that feeds the intuitions of biologists when they label a phenomenon as a behaviour, or conversely, when they take it to not be a behaviour. Though this would classically be laid out in a list of necessary and sufficient conditions, I argue instead for a definition of behaviour as a spectrum, and propose the dimensions along which the intuitions of biologists are developed. This graded multi-dimensional definition leads to an explicit reconceptualization of behaviour in a biological framework. This type of work aims at breaking down ostensible incommensurabilities between various approaches, which might otherwise be talking past each other. In this respect, it can be seen as a first step towards eliminating the plurality of conceptualizations and explanations of behaviour by opening the door to integrations.

To this aim, I first review the recent debates about defining behaviour and identify the lack of consensus as notably stemming from the assumption that behaviour is categorical (section 6.1). By analysing the various definitional elements that have been previously proposed in the literature, I argue that the attribution of the label 'behaviour' stems from the knowledge and intuitions we have about how the phenomenon of interest is mechanistically produced (section 6.2). More specifically, I identify three major characteristics of the mechanistic explanation that play a key role in this respect: the complexity of the mechanism, the stability of its constitutive entities, and the quantity and significance of input variables (section 6.3). I show (section 6.4) how paradigmatic cases of behaviour are those which rank high in all three characteristics. I then look at how coarse cross-discipline attributions of behaviour compare to one another (Section 6.5). Finally, in section 6.6, I explain why this reconceptualization of 'behaviour' matters for scientific integration.

6.1 The problem of defining 'behaviour'

A first difficulty regarding definitions of 'behaviour' is that common-sense definitions separate the notion into two distinct meanings: (1) the way in which an animal or person behaves in response to a particular situation or stimulus, or (2) the way in which a machine or natural phenomenon works or functions, as in "the erratic behaviour of the old car" (OED online, 2017). My aim is to explore the conceptualisation of 'behaviour' in the biological scientific literature, looking at the definitions and uses in that context more specifically. Indeed, my contention is that sense (1) is closer to what is typically found in biology, though sense (2) is also used, and could be considered an extension, or marginal case, for my proposed conceptualisation of 'behaviour' (see Lazzeri, 2014 for a discussion of this issue). I will address this point in more detail in section 6.5.

Looking more specifically at definitions of behaviour within the biological literature reveals a dearth of definitions, and a lack of consensus. Levitis and colleagues (2009) did an extensive review of texts in biology[19], finding over 25 operationally distinct definitions, and over 100 sources which they believed should have included a definition, but didn't. Of the definitions that are put forward, some are too vague or too large to properly circumscribe the phenomenon. For instance, Tinbergen (1976) defines behaviour as "The total movements made by the intact animal", whereas Davis (1966, pp. 4–5) proposes: "What an

[19] Though we criticize several of their fundamental assumptions, (Levitis et al., 2009) is one of the rare pieces of research done specifically about the definition of behaviour in biology, and as such has been valuable to our own research.

animal (or plant) does." Others are more precise in their description, such as Beck et al. (1991, glossary) who propose to define behaviour as the "externally visible activity of an animal, in which a coordinated pattern of sensory, motor and associated neural activity responds to changing external or internal conditions." I will proceed with a more in-depth analysis of these definitions in the next section, but for the moment, suffice to say that their variety is indicative of a problem with defining behaviour in biology, or at the very least a lack of clarity, which is reflected in the disagreements regarding the classification of phenomena. This is a significant issue since a misconception regarding what counts as behaviour, and more generally disagreement about the proper use of a term or the proper referent of a term can mean that researchers can have difficulty knowing whether their explanations truly are targeting one phenomenon, similar phenomena, or wildly different phenomena. This of course can be a barrier to integration, as discussed in chapter 4. And as mentioned in chapter 5, integration can be a tool for breaking down explanatory pluralism, and it is therefore important for my anti-pluralist position to understand what, if any, are barriers to integration, and whether and how they can be overcome.

Indeed, biologists disagree regarding which phenomena ought to be classified as behaviour. Levitis et al. (2009) performed a survey, asking biologists which of a list of phenomena ought to be attributed the label of behaviour. While some phenomena were the object of wide agreement, no phenomenon was unanimously considered behaviour or non-behaviour, and many indicated major divergences of opinion. For instance, "A spider builds its web" was labeled as behaviour by 97% of respondents, whereas "A beetle is swept away by a strong current" scored only 5%. However, examples such as "A plant bends its leaves towards a light source" (48%) or "A rat has a dislike for salty food" (66%) were not so clear-cut. There is therefore a lack of consensus regarding which phenomena ought to be labelled a behaviour. This can be understood as a problem, since as mentioned above, whether something is considered a behaviour can have an impact on the methods and explanations which can legitimately be used to explain it, and has an impact on which—if any—integrative tools can be brought to bear.

Overall, two major problems stand out. The first, and relatively simple, one relates to properly circumscribing the phenomenon that is to be classified (a point also raised in chapter 4). This may seem obvious, but is often overlooked, leading to differing views about what may seem, at first sight, to be the same phenomenon. For instance, in Levitis et al.'s (2009) survey, one of the cases presented is "A rat has a dislike for salty food", which two thirds of respondents classified as a behaviour and one third as non-behaviour. This disagreement may very well have stemmed from confusion regarding the phenomenon at stake: does it refer to the action of avoiding salty food ("A rat tastes some food, spits it out, and does not eat the rest"), or merely to a non-actualized propensity ("A rat, in its cage, is said to dislike salty food, though no food is present")? While the former is quite clearly a behaviour, the latter is not so obvious. It is therefore imperative to be explicit, so as to reduce the chance of divergent classification attributable to misunderstandings regarding the phenomenon at stake.

This can be done in three simple steps, by specifying: (i) the entity in question (e.g., is this a single organism, a group of organisms, a part of an organism?); (ii) what the entity is doing, through the use of an active verb (this therefore excludes events which happen *to* the entity in question); (iii) the appropriate context, so as to state more precisely what the *doing* is about. Consider: "A goose and its flock fly in a V formation." This can reformulated either as "A goose flies alongside other geese, creating a V formation" or as "A flock of geese flies in V formation". In the first case, the phenomenon of interest is one that pertains to a single organism, whereas in the second case, it concerns a group of organisms, each case possibly leading to different reasons for attributing or not behaviour. The types of explanations which can therefore be used may differ, meaning that clarifying the explanandum phenomenon may go a long way in resolving apparent pluralism. "A beetle is swept away by a strong current" is also awkwardly formulated for a behaviour, but for a different reason: here, it is the use of a passive verb that introduces a confusion about who is at the origin of the action. Reformulated with an active verb (ii) as "A strong current sweeps away a beetle" clarifies the phenomenon of interest. The context (iii) also plays an important role in certain cases. Consider: "A spider flies off a branch along with its web, caught by the wind". This should probably not be labeled as 'behaviour' if what is meant is "A spider falls to the ground along with its web; both were caught by the wind", but could very well be an instance of behaviour if reformulated as "A spider catches its web in the wind and electric fields, thus getting carried away once resources are scarce" (as some spiders are capable of doing: see Morley & Robert, 2018). Of course, the reformulation of the phenomenon alone should not be understood to determine whether the phenomenon is or is not an instance of behaviour; it merely ensures that communication across researchers is as unequivocal as possible, after which labeling it as behaviour or not can be addressed.

The second problem to behaviour attribution, I argue, is the assumption that behaviour is binary. Indeed, definitions found in the literature—such as the ones from Tinbergen, Davis or Beck and colleagues mentioned above—apparently take for granted that 'behaviour' is to be understood as a categorical binary concept, insofar as they list necessary and sufficient conditions for classification. Researchers furthermore seem to understand certain actions as being behaviour or not, with no researcher ever referring to a case as being “almost” behaviour, “nearly” behaviour, or any such intermediate qualification. All surveyed definitions by Levitis et al. (2009), including their own, assume that a phenomenon can only be a behaviour or a non-behaviour, despite the fact that their survey results hint that such a division is untenable; their survey as well works with that assumption, asking respondents to determine whether a given phenomenon is or is not a behaviour, with no other choices offered. Such a binary assumption could stem from a wish to cleanly sort phenomena for pragmatic or explanatory reasons. Yet if no such clear-cut divide truly exists, then it would not be surprising that there is disagreement about the classifications, leading to the problems mentioned above, namely the fact that a faulty or confused classification could lead to erroneous conclusions regarding the perceived legitimacy of certain types of explanations. I elaborate on the non-binarity of ‘behaviour’ in section 6.3.

6.2 Recognizing instances of ‘behaviour’

Despite the disagreements and problems associated with defining behaviour, biologists do classify phenomena as behaviour and non-behaviour. What is it that biologists recognize when applying the labels? And does it work?

A first thing to note regarding definitions of behaviour is that many of them specify that ‘behaviour’ is attributed only to certain *types of entities*. According to these views, the decision to classify a phenomenon as behaviour depends at least in part on the properties of the entity to which it will be attributed. For instance, Tinbergen (1976) defines behaviour as “the total movements of the intact animal”, limiting behaviour attribution to animals. Others, such as Davis (1966, pp. 4–5) have included plants in their definition, while still others stipulate only that it must be attributed to an organism (Raven & Johnson, 1989; Wallace et al., 1991).

A look at the uses of the term ‘behaviour’ shows that such restrictions do not reflect current scientific practice. Indeed, looking at the literature in biology reveals that virtually all types of organisms and all levels of organisation can be said to exhibit behaviour. Behaviour is of course attributed to ‘medium-sized’ organisms such as humans, mammals, birds, reptiles, and invertebrates, with entire disciplines created around these issues, such as evolutionary psychology, ethology, or behavioural ecology, to name only a few (as evidenced by any textbook in those disciplines). Yet behaviour is also attributed to plants (Belter & Cahill, 2015; Cvrčková et al., 2016; Karban, 2008; Silvertown, 1998; Trewavas, 2009), single-celled organisms such as bacteria (Ben-Jacob et al., 1994; Dussutour et al., 2010) or amoebae (Hansell, 2007), viruses (Miyashita et al., 2015; Quignon et al., 1997), and even proteins (Royer, 2002). Though the use of 'behaviour' to characterize the movements of some of these entities may be more akin to the term as applied to inanimate objects (sense 2 above), it is for now sufficient to point out that given the variety of entities to which behaviour is attributed, the identification of the subject of the phenomenon to be evaluated will (at best) not play an important role in the attribution of behaviour.

A second element found in definitions of behaviour relates to its *intentionality*, or goal-directedness. This is intuitively a central aspect of behaviour, the idea being that what might differentiate behaviour from mere movement or physiological processes is that while the former is purposive, the latter are not. It is interesting to note however that intentionality or goal directedness is seldom, if ever, referred to explicitly in definitions of behaviour within the biological literature. Intentionality as it refers to mental processes has long been defended—and criticized—as an explanatory tool in philosophy and cognitive science. However, in biology, reference to beliefs, desires, etc. has generally been regarded with suspicion, with most researchers opting to avoid such an approach. For instance, Kornblith (2002) and Weber (2012) both suggest that though there may be value in adopting the intentional stance when it comes to the initial observation and classification of behaviour, it ought to be dropped for explanations. Indeed, as Longino has pointed out, the folk psychological concepts that are the hallmark of intentional explanations "are not well suited to scientific investigation, being vague and inflected with the social values that make the behaviours they designate salient in the first place" (2013, pp. 151–152). One attempt at avoiding these problems is to naturalize the notion of intentionality, as Millikan (1993) has done through the use of biological function. According to her definition, for an action to be considered a behaviour, it needs to have been selected for, otherwise it is merely movement. In this way, Millikan naturalizes goal-directedness by equating the function of the adaptive behaviour with the goal of the behaviour. However, this approach excludes exaptations and other non-adaptive behaviours from being labelled behaviour, despite the fact that—as Millikan herself concedes—biologists will tend to call them behaviours. Thus the main issue with Millikan's definition of behaviour is that it explicitly runs counter to many of the intuitions and uses of the term by biologists, which is precisely what my proposed analysis is attempting to capture; while Millikan's approach has a normative aspect to it, mine remains descriptive.

Third, certain definitions will focus on the *output* of a phenomenon as the mark of behaviour. In particular, some will specify sets of criteria that outputs must meet, such as Grier & Burk (1992, p. 4) who define behaviour as “all observable or otherwise measurable muscular and secretory responses (or lack thereof in some cases) and related phenomena such as changes in blood flow and surface pigments in response to changes in an animal’s internal and external environment.” Dawkins (1976) as well emphasizes particularities of the output, characterizing behaviour as “the trick of rapid movement,” the idea being that animals—as opposed to plants—evolved to become efficient vehicles for their genes, able to quickly navigate their environment. Another property of the output which is mentioned by Dawkins (1976), and elaborated on by Sih (2004) is that behaviour is easily reversible, opposing it to developmental changes which, even if plastic, are typically irreversible. Sih (2004) furthermore relates this approach to the idea that behaviour is understood to be labile, allowing for a variety of outputs over the course of a lifetime. Other definitions will be broader and require only that the output be observable (e.g. Beck et al., 1991; Wallace et al., 1991), presumably in order to exclude internal physiological processes (e.g. Millikan, 1993, p. 137).

However, somewhat counterintuitively, evaluating what the ‘outputs’ of the phenomenon are does not always help in its classification as behaviour. There are numerous cases when the outputs can range from just about anything to nothing at all. Indeed, consider an animal playing dead, or a human choosing to do nothing. In both cases, there is no movement to be observed. Yet they would typically be considered behaviours. As for speed being a determining factor, this too is problematic for similar reasons. How are we to evaluate the speed of inaction? And if we go along with many plant biologists (Belter & Cahill, 2015; Cvrčková et al., 2016; Silvertown, 1998; Trewavas, 2009) and agree that plants exhibit behaviour, then it would seem that speed, though a common feature of much of prototypical behaviour, is not a necessary feature.[20] The reversibility and lability aspects of behaviour also face counter-examples, such as the migration patterns of salmon, which happen only once in the lifetime of a given fish, or mate choice in species which form long-term mating bonds (see Sih, 2004, p. 114 for further examples). These examples are nevertheless quite clearly considered behaviour by biologists. In light of these difficulties, it becomes

[20] We furthermore considered another interpretation of ‘speed’: rather than absolute speed, perhaps the appropriate metric would be the ‘execution time relative to life-cycle.’ Our idea was that such a conceptualisation could allow for our definition to adumbrate plant behaviour, under the assumption that plants such as trees have longer life-cycles, and therefore slower movement could still be proportionately seen as ‘fast’. However, the difficulty with such an approach is that we are left with an unwieldy metric which inadequately deals with the measurement of ‘execution time,’ not to mention issues with short-lived plants and other organisms.

difficult, if not impossible, to find common aspects specific to the outputs of phenomena considered behaviour.

A fourth and final element found in many definitions of behaviour is the notion that behaviour is a *response* to stimuli, whether external or internal. This is the case for instance with Wallace et al.'s definition: "observable activity of an organism; anything an organism does that involves action and/or r*esponse* to stimulation" (1991, glossary, our italics), or Raven & Johnson's definition as "the way an organism *responds* to stimulation" (1989, p. 1119, our italics). Though this does seem to capture something important for a biological definition of behaviour, without further precisions regarding the notion of what constitutes a response, it remains somewhat vague. Since a 'response' is merely a change, or movement, which results from a stimulus, these criteria can make it difficult to differentiate examples of paradigmatic behaviour from non-behaviour. For instance, a tree burning or growing both involve changes following stimuli, though few would want to consider them behaviours. Nevertheless, if the notion of response alone is insufficient to make the needed distinctions, more precise definitions may help.

Certain definitions specify which of the organisms' underlying mechanisms need to be recruited in order to classify the response as one which results in behaviour. For instance, Beck et al. emphasize the presence of "a *coordinated pattern of sensory, motor and associated neural activity* [that] responds to changing external or internal conditions" as the mark of behaviour (1991, glossary, my italics). Levitis et al. define behaviour as "the *internally coordinated* responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, *excluding responses more easily understood as developmental changes*" (2009, p.108, my italics). Both these definitions point out that mechanisms of a certain sort must be present in order to classify a phenomenon as a behaviour, which allows for a more precise understanding of 'response' and how this makes the phenomenon a behaviour.

I believe that the emphasis on the underlying mechanisms is most fruitful, allowing the differentiation of superficially similar phenomena which are nevertheless quite clearly at opposite ends of the spectrum of behaviour and non-behaviour. Compare for instance the two following examples: (a) "A dog bites my hand when I touch it"; (b) "A rose bush punctures my skin when I touch it". On a superficial level, both are similar: they involve biological entities to which behaviour can be attributed, both entities seem in some sense to have been biologically disposed to hurt me, and the output is similar insofar as both result in pain to my hand. But the differences become evident when looking at how the phenomenon came about, and what the resulting explanation is. In the case of the dog, there is a perceptual apparatus which feeds signals to a neural mechanism for processing, then coordinated control of bodily movements through muscular activity, which leads to the dog biting. The explanans thereby involves a sophisticated mechanism (perception, processing, coordinated bodily movements etc.), with specific types of entities playing specific roles (visual cells, neurons, muscles etc.), and fed by numerous inputs (through vision, smell, proprioception, etc.). In the case of the plant, the only mechanism recruited is the presence of thorns on the stems, which then come into accidental contact with my skin, leading to a puncture wound. This explanans calls upon virtually no mechanism whatsoever and involves very few inputs. The perceived difference between these two phenomena therefore does not relate to the properties of the organism or phenomenon itself**,** but rather to what is thought to produce the phenomenon. Akin to the other definitions mentioned earlier, I argue that what plays a key role in behaviour attribution are the characteristics of what explains the phenomenon to be classified (more about this below).

Obviously, detailed knowledge about what produces the phenomenon need not be complete in order to feed the intuitions. The knowledge can range from full-fledged theories to tentative schematic explanations, and even to merely assumed mechanisms. In the case of the dog's biting, for instance, one can either conjecture that there is some sensory input, processing and muscular coordination, or one can know the minute details of every part of the whole range of recruited mechanisms. I furthermore suspect that when the mechanism is unknown, the assumptions regarding the underlying mechanisms will often stem from anthropomorphising: when seeing an entity act in a way which somehow resembles human behaviour, the tendency will be to assume that the underlying mechanisms must be similar to a human's, and therefore that the phenomenon will also be a case of behaviour. In sum, the knowledge and details of the mechanisms furthermore need not be explicit—on the contrary they are quite often implicit.

6.3 Characteristics of the explanans as reasons for labeling a phenomenon 'behaviour'

Explanations which rely on explicating the underlying mechanisms through which phenomena are brought about have led to a growing literature in philosophy of biology and other sciences. I take these "mechanistic explanations" to be the fuel behind the intuitions relating to classifications of phenomena as behaviour. It goes without saying that other explanations of behaviour exist, such as evolutionary explanations, or statistical explanations—my claim is merely that the mechanistic explanations are those which play a role in classification. These types of explanations explain in virtue of uncovering the underlying mechanisms which reliably produce the phenomenon that is in need of explanation (Bechtel &

Richardson, 2010; Craver, 2007; Glennan, 2002; Machamer et al., 2000). Mechanistic explanations explain by pointing to a number of specific entities—organized in a particular way and which carry out a number of specific activities—which, when affected by certain input conditions, are susceptible to produce a certain phenomenon. Behaviour as well can be explained through mechanisms. These can come in various types: some have numerous entities, others fewer; some are more complex, others less, etc. I propose that three specific characteristics of the mechanistic explanans play a crucial role in attributing the label of 'behaviour' to a given phenomenon taken as explanandum. As I show below, each one of these three characteristics is best captured as being laid out along a gradient (and therefore not in a binary way).

6.3.1 First characteristic: mechanism complexity

The first of these characteristics is the complexity of the underlying mechanism: the more a mechanism is complex, the more biologists will tend to classify the resulting phenomenon as a behaviour. If we return to the comparison between the dog biting my hand and the rose thorn puncturing my hand, it is intuitively easy to see how the mechanisms underlying the former are quite complex, whereas the latter are relatively simple.

Of course, 'complexity' is a notoriously difficult term to define and to quantify. Many definitions and conceptualizations exist, including Mitchell's three-part distinction outlined in chapter 1. I do not presume to resolve those difficulties here: rather than attempt to give a generalizable definition applicable to all contexts, I am content to propose a fairly simple metric, evaluable through a mechanistic framework: what I have in mind with this characteristic is the number of entities and relations posited in the mechanism which accounts for the phenomenon. The more entities, activities and relations, the more complex a mechanism will be.

Obviously, the evaluation of complexity will change depending on the bottoming-out entities that are assumed to be relevant for a mechanism to be explanatory in a given (disciplinary) context. For instance, in a discipline interested in physiological aspects, such as botany, one would explain how the rose plant punctured my skin just in terms of the presence of thorns, which then come into contact with my skin, leading to a wound. But for other disciplines, such as plant molecular biology, one could go to a finer granularity, for instance by mentioning the cellular entities involved in the growth of thorns. The implication is that measuring the complexity of a mechanism for the sake of evaluating whether or not it is a behaviour should be done comparatively with other phenomena, ideally at something resembling the same level of explanation. This comparative aspect applies to all three characteristics, and is addressed in greater detail in section 6.4.

When trying to assess the complexity of a mechanism, it may not always be possible or easy to evaluate the number of entities and activities that figure in the mechanism, especially when the mechanism is not known or detailed. I propose three rules of thumb relating to judging the complexity. (a) If the current state of science is such that the mechanism is not yet fully understood, it is likely that the mechanism is quite complex.[21] Of course, this is not a firm rule, since knowledge of the mechanisms will depend on contingent factors relating to research, such as financing, availability of the research objects, ethical concerns, etc. (b) I take for granted that phenomena which stem from conscious and deliberate decision-making—or that could be labelled as intentional if one wishes—are among the most complex, because of the neural mechanisms involved, which are often so complex that we understand them only very partially (this relates to point (a) above). (c) When comparing phenomena, it can be useful to wonder whether one phenomenon's mechanism is completely or partially 'included' in the other. For instance, reflex movement is less complex than deliberate movement, since much of the mechanism included in reflex movement will be solicited by deliberate movement, as well as many others.

To clarify my understanding of the valuation of complexity, the following are examples of phenomena which score high on complexity. As alluded to earlier, "A dog bites my hand when I touch it" involves a very complex mechanism, recruiting various senses, deliberate muscle control and other neurological mechanisms. A similar story can be said about other phenomena such as "A dog salivates in anticipation of feeding time", or "A salamander performs autotomy to escape a predator".[22] Complexity is also high for "A salamander regenerates a lost limb", though it recruits complex developmental processes rather than the whole gamut of neurological mechanisms involved in the examples above.

Phenomena which score low are ones such as "A rose bush punctures my skin when I touch it". The mechanistic explanation used to describe the phenomenon is extremely simple compared to the previous examples, limited as it is to the presence of thorns on the stem. Similarly, "A spider falls to the ground

[21] If, on the other hand, the mechanism is not known by the individual researcher but it is known by other biologists, this can lead to mistaken assessments of the complexity of the mechanism (and indeed all characteristics of the mechanism). This issue is addressed in section 6.4.

[22] Autotomy is when an organism deliberately casts off a part of its body when under threat.

along with its orb web; both were caught by the wind" involves essentially no mechanism since the spider is a passive victim in this accident.

There are, of course, intermediate cases as well. For instance, the mechanistic explanations for "A sunflower turns its head to follow the sun" exhibit a certain complexity (Vandenbrink et al., 2014), calling as they do on many entities and processes—such as mechanisms for the directional perception of a light source and differential growth of stem segments—though certainly not of the same order as the deliberate decision-making found in cases which score high. "A human's leg jerks in a reflex response to an impact on the knee" is also considered intermediate, since using the third rule of thumb described earlier, it calls upon muscular and nervous mechanisms, but omits the deliberate decision-making processes involved in high-scoring complex phenomena.

6.3.2 Second characteristic: entities stability

Another strong intuition which fuels our assessment of whether or not a phenomenon ought to be classified as a behaviour is how ontologically stable the mechanism is. This is particularly important in how we differentiate behaviour-like phenomena from development-like phenomena: in the former case the entities constituting the explanatory mechanism stay as-is (i.e. do not cease to exist or do not create new entities), whereas in the latter case, entities tend to be created, modified, or destroyed (for example in instances of growth or cellular differentiation). The second important characteristic of the explanans that I propose to single out is thus the entities' stability in the mechanistic explanation. The more stable the entities of a mechanism are during the production of the phenomenon of interest, the more likely the resulting phenomenon will be labelled as 'behaviour'. Conversely, the more change there is in the entities (for instance, with new entities coming into existence or entities disappearing), the more likely the phenomenon will not be an instance of behaviour—but rather typically an instance of development, aging, or degradation/decomposition.

Of course, change affects all entities of a mechanism through the action of their activities and according to their respective relationships: change is what produces the phenomenon the mechanism is supposed to explain. Yet change may affect the entities of the mechanism in two extreme ways. On the one hand, it may only concern the spatio-temporal position of the entities: this is typically what happens with the gears of a mechanical clock (they turn but remain intact—except for wear, which is neglected in the regular functioning of the clock mechanism). On the other hand, change may concern the very existence of the entities: most of the entities that constitute a firework get destroyed during the functioning of the firework; and conversely, numerous novel entities are created during the course of the development of an embryo. The second of these two characterizations of 'change' is the relevant one for the sake of categorization of phenomena as behaviour.

"A dog bites my hand when I touch it" is a paradigmatic case of a phenomenon scoring high on entities stability: all the entities recruited in the mechanistic explanation remain essentially as they were at the outset. Other cases, such as "A spider builds its web", involve a very small degree of entities production (in this case the silk), but the great majority of the entities remain stable throughout the process (all entities of the spider's sensory mechanisms, of its muscle control and other neuro-physiological mechanisms involved in building a web). The same can be said of "A rose bush punctures my skin when I touch it", since the thorn remains in place.

Phenomena which almost exclusively rely on growth and developmental processes will score low on this characteristic. For instance, "A plant adjusts its root placement in response to neighbours" (Belter & Cahill, 2015) is explained through the mechanisms of root growth, involving the production of a vast amount of new constitutive entities. In the same way, "A salamander regenerates a lost limb" is a phenomenon relying exclusively on developmental processes. The loss of entities is also a reason to score low on stability, as in the case of "A tree burns in a forest fire", where the entities decay and are lost in the fire.

Intermediate cases include "A sunflower turns its head to follow the sun". Though this type of growth-mediated heliotropism does rely on irreversible cell expansion, that change is localized and therefore involves only a small part of the plant (Vandenbrink et al., 2014). "A salamander performs autotomy to escape a predator" is another intermediate example: many entities are lost in the process, in this case the salamander's limb, though the salamander itself and the mechanism through which the autotomy is produced remain mostly intact (if the escape is successful!).

It is interesting to note here that many phenomena which can be labelled behaviour in plants rely on developmental processes and therefore will tend to score low or intermediate on this dimension (though not all: "A mimosa plant shuts its leaves when touched"[23] scores high on stability since the mechanisms do not involve developmental processes). This is one of the reasons plant behaviour is often considered a marginal case of behaviour (see section 6.4).

6.3.3 Third characteristic: quantity and significance of difference-making inputs

The third characteristic of the explanatory mechanism which, I argue, plays a role in differentiating behaviour and non-behaviour is the inputs that trigger the mechanism in the first place. Take the dog and rose bush examples. The dog's biting of my hand involves the perception of stimuli in the form of visual and olfactory cues, and perhaps internal stimuli which dispose the dog to be aggressive (or prone to defend itself). A mechanistic explanation the dog's biting will include all these as inputs of the mechanism. On the other hand, the plant's only input in the explanation of the puncture wound is that my hand came in contact with it. The dog example therefore calls upon multiple inputs, the rose bush example only one; as such, the dog's biting will be considered more of a behaviour than the plant's hurting my hand.

This characteristic relates to the inputs which are relevant to the mechanism explaining the phenomenon. The more relevant inputs there are, the more we will tend to call the phenomenon a behaviour. This can intuitively be understood as a form of reactivity or response to inputs in general, whether environmental or otherwise: if the mechanism changes its operation, end-state or outputs according to a broad range of inputs, it will be more reactive, and closer to paradigmatic cases of behaviour.

Of course, for each mechanism there are a potentially infinite number of inputs if one includes (for instance) all necessary circumstances. For the sake of classification of behaviour, and indeed for scientific explanation in general, what is relevant is the identification of 'difference-making' inputs in the sense highlighted by Waters (2007) in the context of causation. According to Waters, an actual difference-making cause is one which accounts for an actual difference in outcome, this difference being measured for a given population (in other words, a set of instances) that have been determined by our epistemic interests (2007, p. 569). Similarly, we can define a difference-making input to any given mechanism as one which accounts for an actual difference in the phenomenon that is produced by that mechanism, for a given set of instances of that mechanism. What this implies for our purposes is that it does not suffice to have many

[23] *Mimosa pudica* plants are well known for their capacity for seismonastic reaction to touch, meaning that they can 'shut' their leaves by folding them close to the stem. The mechanisms involve electrophysiological processes, as opposed to developmental ones (R. D. Allen, 1969).

inputs; only the inputs which account for some difference will be relevant. For instance, although the growth of a tree's limb requires a great number of inputs—sunlight, water, nutrients and minerals (Poorter et al., 2013)—these inputs may not be actual difference-makers if the relevant population of growing trees is affected by the same inputs. In the context of behaviour, if the relevant population is a copse of trees which are all exposed to sensibly the same inputs, then these will not be difference-making inputs between the trees that compose that copse (population). If, on the other hand, only one of the trees grows much faster or slower due to particular inputs, then those particular inputs will be considered difference-making and therefore relevant to behaviour attribution.[24]

Consider the following examples which score high on difference-making inputs: "A dog bites my hand when I touch it" and "A salamander performs autotomy to escape a predator". In both cases the underlying mechanism integrates a great number of inputs over time in order to arrive at the end state, whether it is the biting or the autotomy. Conversely, the explanatory mechanism for "A rose bush punctures my skin when I touch it" recruits very few inputs, simply the presence of a thorn on the rose bush and the trajectory of my hand relative to the bush. A more difficult case could be "A salamander regenerates a lost limb": the mechanism underlying this phenomenon is likely to be a quite complex growth process with many entities at the cellular and molecular level; yet this complex growth process is simply triggered by a single variable: the sectioning of the limb. In this respect, it too can be considered to score low on difference-making inputs.

Intermediate cases are those which integrate a certain number of inputs over the unfolding of the mechanism, but fewer than the paradigmatic cases. For instance, "A dog salivates in anticipation of feeding time" involves the dog integrating the information regarding the time of day and the implications thereof (whether consciously or not), but that remains fewer inputs than those involved in the example of the biting of a person's hand above. "A plant adjusts its root placement in response to neighbours" is an example of a plant mechanism integrating a certain number of difference-making inputs, namely its neighbours' location, species and even relatedness (Belter & Cahill, 2015), and changing the root placement accordingly.

[24] Note that Waters (2007) adds in footnote 24 that this analysis also applies to singletons when the population is considered to be the individual "before, during, and immediately after" the input is considered. This also applies to our analysis.

6.3.4 Rejected characteristics

There are further characteristics that I considered but ultimately rejected due either to the existence of counter-examples, or because the characteristics I propose already account for those aspects. If it is possible to find an example of a phenomenon which is considered to be a paradigmatic example of behaviour, but which does not satisfy the criterion highlighted, then that criterion is not a crucial feature of behaviour. As such, in our review of the various definitional elements (section 6.2), we have already given good reasons to reject many of the characteristics that we considered as contenders for relevant aspects of behaviour.

This explains why I ignore the type of entity to which the behaviour is attributed, as well as all the ways of characterising the outputs of the mechanisms in question. As mentioned earlier, behaviour is attributed to all manner of entities, and the sheer variety of possible outputs which are labelled 'behaviour' make it impossible to use those elements as determining factors for behaviour classification. For instance, though it is often the case that rapid movement will be considered behaviour, there are nevertheless paradigmatic examples of behaviour which involve no movement whatsoever, such as playing dead, or the freezing behaviour of prey animals. Thus, even in a spectrum framework, entity type and outputs are not factors which make a phenomenon more or less a behaviour, since certain examples are clearly considered behaviours despite the variety of entities and outputs.

The notions of intentionality or goal-directedness may be intuitively appealing, but as previously mentioned, within biology they are typically discarded in favour of a more naturalistic understanding of the mechanisms of decision-making. In that respect, the notion of behaviour being a "response" is captured by the three characteristics we have highlighted, since a response, translated in terms of a mechanistic explanation, will involve a complex mechanism which adequately deals with a number of difference-making inputs, often without significant changes to the constituent parts of that mechanism.

6.4 Behaviour space

Paradigmatic cases of behaviour are legion, and are those phenomena that typically score high on all three dimensions. For instance, the mechanism put forward to explain "A dog bites my hand when I touch it" is complex, stable, and relies on a great many inputs. The same can be said of innumerable other examples, such as a bird building a nest, a wolf hunting its prey, an ant cleaning its antennae, a person playing soccer, etc. Any phenomenon which is uncontroversially considered behaviour will score high on all three characteristics. On the other hand, paradigmatic cases of non-behaviour will score low on at least two of the three dimensions. For instance, “A mouse floats in zero gravity in outer space” (taken from Levitis et al. 2009) is paradigmatically not a behaviour, since the underlying explanatory mechanism scores low on both complexity and inputs. Similarly, “A salamander regenerates a lost limb” scores low on both stability and inputs (as discussed earlier), and is therefore clearly non-behaviour. The three characteristics taken together thus make it possible to differentiate between paradigmatic cases of behaviour and non-behaviour.

But in between these paradigmatic extremes are a broad range of intermediate cases. One of the interests in using the combination of these characteristics is to reveal these intermediate phenomena: those which score neither high nor low on at least one of the characteristics, or which score low on only one of the three. When forced to say whether a given intermediate case is or is not a behaviour, some biologists will say that it is, and others that it is not. However, in my graded view of behaviour, these are simply phenomena which remain intermediates: neither paradigmatic behaviour, nor non-behaviour, but somewhere in between.

Inspired by Godfrey-Smith’s (2013) treatment of Darwinian individuals, I have constructed a three-dimensional space, where each of the characteristics is represented as a dimension, which makes it possible to capture the full range of phenomena and compare them to one another by means of assessing how their underlying explanatory mechanisms fare along each of these characteristics (see Fig. 6.1). The phenomena placed within it are intended to be relatively uncontroversial, with the placements along each axis assumed to be fairly intuitive when remaining as close as possible to the same level of analysis. The illustration shows certain informative contrasts among phenomena, but limits itself to a coarse three-way distinction for each axis, with cases rated on each dimension as low, intermediate or high. As a general rule, the closer a case is to the corner representing a high value for all dimensions, the more it is a behaviour, though as will shortly be discussed, interesting differences exist between disciplines and when looking at finer distinctions for certain axes. I note furthermore that there exist cases which are high on one dimension and low on the others, implying that the dimensions are conceptually independent from one another, the important point being their capacity to capture the diversity of behavior-related phenomena.

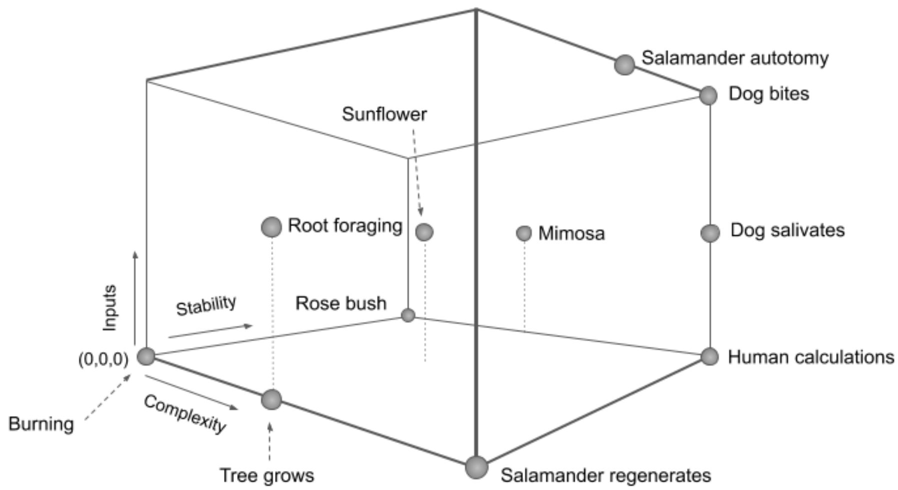

Figure 6.1 - Three-dimensional behaviour space, with positioning of sample phenomena (positioning done by assessing the score of the underlying explanatory mechanism along each one of the three dimensions: complexity, stability and inputs).

Most of the examples present in the cube have been at least partially justified in the discussions for each of the three characteristics (Section 6.3), and I will cover here only a few of the as yet unjustified placements. “A mimosa plant shuts its leaves when touched” has already been highlighted as an unusual example of plant behaviour which scores high on stability. It furthermore falls in the intermediate area for complexity since the underlying mechanism does involve a good amount of entities and activities, including electrophysiological processes which affect turgor pressure (R. D. Allen, 1969). It furthermore scores intermediate for inputs, since there is some sensitivity to touch (though not all touch at all times [J.C. Cahill, personal communication]), though few other inputs will have an effect. “A tree grows a limb over years” obviously scores low on stability due the fact that the whole purpose of the mechanism is to add constitutive entities to the tree in the form of additional biomass. The complexity is intermediate since though the mechanisms involved in growth-ring formation do contain many entities and activities (Rossi et al., 2006), they are not as complex as the neurological mechanisms involved in many of the paradigmatic cases of behaviour. This growth furthermore scores low on inputs since despite the fact that inter-species variability can be great, growth rates tend to be similar for a given species in a given location (Rossi et al., 2006, p. 302), implying that the inputs tend to not be actual difference-makers. “A tree burns in a forest fire” scores low on all dimensions: the explanatory mechanism is simply a set of oxidation reactions of the organic compounds of the tree; because the burning destroys the constitutive entities of the mechanism, the stability is low; and the difference-making inputs are low when considering that the burning requires only the difference-making input of heat. Interestingly, this is one of the few examples I have found of a phenomenon scoring low on all dimensions. I suspect that this is due to the fact that it is rare that biological mechanisms will jointly present low stability and low complexity, since the former often relies on complex developmental mechanisms, or rather passive ones such as in the example just mentioned of the tree burning. Finally, “A human performs mental calculations” is high on complexity and stability since the intricate neural mechanisms it recruits remain essentially unchanged, and low on inputs since very few elements will be difference-making aside from the numbers and formulas to solve.

It is possible to place any phenomenon of interest within these three dimensions—simply by assessing its underlying explanatory mechanism—and hence evaluate just how close it is to being a paradigmatic case of behaviour. Comparing our results to those of Levitis et al.’s survey reveals that our approach captures the paradigmatic cases, as well as the difficult intermediate cases which caused “major divergences among respondents as to whether the phenomena were behaviours.” (2009, p. 107). Indeed, all phenomena which met with widespread approval score high on all three dimensions, those which were generally considered non-behaviour score low or intermediate on at least two dimensions, and intermediate cases fall somewhere in between.[25]

The placement of each of these cases within the cube comes with a few caveats, notably with respect to the valuing or measurement of each of the dimensions. First, I should note that I do not expect there to be precise, numeric values for each dimension; the placement should instead be understood as a way of sketching comparisons between different phenomena. The position of a phenomenon will therefore be done with respect to other phenomena which are believed to be better understood or more easily classified than the phenomenon at hand in terms of their underlying explanatory mechanism. I take for granted that comparing cases involving very different entities as we have done makes it impossible to

[25] For brevity’s sake we have not reviewed each of the cases in Levitis et al. (2009) here, though we encourage curious readers to see for themselves how our approach compares favourably with the survey results. We also are aware that the survey takes for granted a binary approach and therefore does not measure the *degree* to which a phenomenon is a behaviour; as such the results need not—and do not—directly correspond to our valuations. Nevertheless, I believe that it is reasonable to assume that divergences in opinions among biologists are an indication that the cases are intermediate cases and not paradigmatic ones.

discern fine distinctions along each axis, hence the coarse-grained nature of the behaviour-space in Figure 6.1.

Second, it is furthermore necessary, when trying to compare mechanisms with a view to assessing whether a phenomenon is an instance of behaviour, to attempt as best as possible to remain at the same level of analysis. There will be times when comparisons across levels are done, but it will always be very difficult due to the fact that entities and activities present in different mechanisms can sometimes belong to radically different types due to epistemic interests and choice of bottoming-out entities (Machamer et al., 2000). This difficulty may be why our intuitions are so conflicted when we attempt to do so; talk about behaviour at the level of single-celled organisms (e.g. Dussutour et al., 2010; Hansell, 2007), in particular bacteria (Ben-Jacob et al., 1994) or even proteins (Royer, 2002) can be somewhat jarring, but close examination of the phenomena can leave biologists with the feeling that it does indeed look something like behaviour.

Finally, the knowledge that is mobilized to determine where to place a phenomenon can have a great impact on the valuation for each dimension. This is reflected in individual and disciplinary knowledge, as well as the advancement of science in general. In Section 6.2 I mentioned that, on an individual level, knowledge need not be complete, and indeed may be schematic or assumed; though in some circumstances the degree of detail may not change the final classification, it sometimes does. The knowledge-dependent nature of the valuation can also be reflected on a disciplinary level, especially when considering the comparative aspect of the valuations. This is notably the case in phenomena sometimes labeled 'plant behaviour': what could look to the layperson (or non-plant biologist) as a clear example of non-behaviour may be judged otherwise by a plant biologist. On a more general level, this also implies that our understanding of a phenomenon as a behaviour—hence its relative positioning in the cube—is liable to change as our explanations for its underlying mechanisms change or get refined over time.

6.5 Shades of behaviour

There are many cases of phenomena falling somewhere between paradigmatic cases of behaviour and non-behaviour. As touched upon earlier, these intermediate cases pose problems for those who assume that behaviour is a categorical concept: where does one draw the line, both in terms of definitions and classification? My conceptualisation removes the need for such a line. Behaviour is understood not as a definite and circumscribed class of phenomena, but simply a way of understanding certain complex biological processes which are particularly receptive to inputs, and which tend to preserve the entity or mechanism in question as-is. Intermediate cases are therefore phenomena which score lower on (at least some of) these dimensions. My approach explains away the variable intuitions, by pointing out that forcing biologists to draw a line between that which is and is not behaviour leads to unnecessary and difficult decisions on an individual level, and conflicting decisions when comparing with peers. If it is recognized instead that phenomena can simply be classified as more or less behaviour, then these conflicts disappear, and are replaced instead by debates regarding the valuation along each axis.

Because placement in the behaviour-space depends on the knowledge of the classifier, the level of analysis, and importantly, comparison with other cases, different disciplines are likely to place phenomena at somewhat different locations, and even to understand behaviour in a broader or stricter sense. In Figure 6.2 I have identified broad regions understood as behaviour for different classes of entities (often—though not exclusively—correlated with different biological disciplines). The region typically associated with *human* behaviour tends to include at its margins phenomena which score very high on complexity and stability, but that need not score high on the inputs dimension. This is likely due to the fact that we, as humans, can understand or assume the mental processes behind some apparently input-free phenomena. For instance, we appreciate that a person performing mental calculations and who is otherwise immobile is doing something like behaviour. However, when looking at animals, plants, or other entities, we will often assume that immobility is simply the absence of behaviour—except in notable cases such as when animals play dead, in which case we understand that the lack of movement is the result of complex neural processes and the integration of multiple inputs. This explains why the region associated with animals is similar to that associated with humans, except for a higher requirement with respect to the inputs dimension.

When it comes to plants, plant biologists will tend to conceptualize even marginal cases as behaviour. This is due, at least in part, to the comparative valuing of the dimensions mentioned earlier: explanatory mechanisms within plants which seem complex compared to other plant mechanisms may appear less so when compared to certain human or animal neural processes such as those involved in deliberate decision-making. The same applies to the inputs dimension. The stability dimension is also treated quite differently, as evidenced by phenomena such as root foraging, which are sometimes conceptualised as behaviour despite the fact that they clearly rely on developmental processes. In light of these examples, some plant

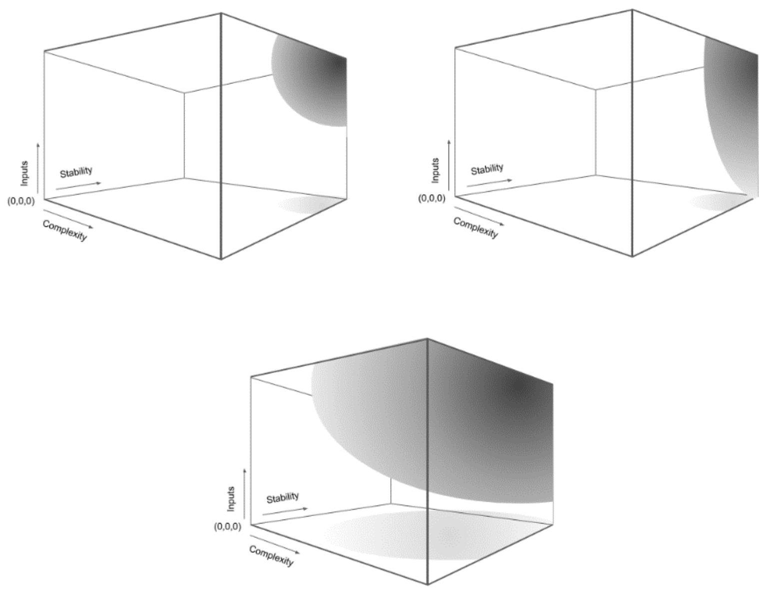

Figure 6.2 - Behaviour-spaces representing the volumes containing phenomena commonly considered behaviour for animals (top left), humans (top right), and plants (bottom).

biologists do not believe that the stability dimension is very important (J.C. Cahill, personal communication; Cvrčková et al., 2016), since in many cases it is the only way that plants *can* exhibit something resembling behaviour. Indeed, for sessile organisms, development is often the most viable option for complex interactions with the environment—though not the only option (e.g. *Mimosa pudica* mentioned above). The understanding of what counts as behaviour for plants is therefore more permissive than for animals or humans.

One last area which we have not highlighted in the behaviour space is the use of the term 'behaviour' as applied to inanimate objects. As previously mentioned (section 6.1), the common definition of behaviour is divided in two definitions, one applying to animate entities (the way an animal or person behaves) and the other to inanimate entities (e.g. "the erratic behaviour of the old car"). I believe that my proposed conceptualisation of behaviour not only captures the use of the term in the biological literature, it also captures its use elsewhere. Thus, a mechanistic explanation of the car's erratic behaviour will likely focus on a few worn-out components—say the steering rack—resulting in relatively low scores in terms of difference-making inputs (compared to well-functioning cars); complexity of the mechanism however could still be found to be average, and stability high. All in all, this will result in placing the old car behavior far from the upper-right cases of paradigmatic behaviour, though not in the lower-left corner either. Thus, when the term is used to refer to inanimate behaviour, it can be understood as cases that are even further from prototypical biological behaviour, though still within the behaviour-space. This also applies to certain uses within biology which, it could be argued, use the inanimate sense of the common definition, for instance when talking about the behaviour of proteins or viruses. By encompassing even these uses, I forgo the need to assume the two different senses of the common definition, once again relying instead on the three dimensions of behaviour proposed.

In light of this, and despite intuitions to the contrary, it is therefore incorrect to say that a given phenomenon will “count” or not as a behaviour. Indeed, abandoning the binary assumption not only removes the problems related to intermediate cases, but it also quite radically changes the conceptualisation of behaviour itself. The debate surrounding the classification of a particular phenomenon changes from one regarding the definition of behaviour to an empirical and comparative investigation into the underlying mechanisms. If a researcher can show that a given phenomenon is in fact produced by a complex, stable and input-laden mechanism, then its placement within the behaviour-space will be uncontroversial, and hence its position relative to paradigmatic examples of behaviour. Classifications of phenomena will therefore be based on empirical grounds, regarding specific characteristics of the explanation (which is, incidentally, what biologists are in fact interested in). Whether or not a given case *truly is* a behaviour is simply not a question which can be answered: rather, one should ask whether one deals with a paradigmatic case of behaviour or a more marginal one.

If recognizing that behaviour is a spectrum yields interesting consequences, placing the cases within a behaviour-space furthermore has heuristic value. Not only can we identify regions associated to particular entities or disciplines, populating the behaviour-space reveals that certain areas are sparser than others. Some of these areas seem to be due to logical impossibilities, such as high inputs but low stability and complexity (how could a mechanism integrate many inputs and change the entities within it without complexity?). Others may simply be due to our own ignorance, both on a personal level (i.e. I may simply not know about them), and on the grander scale of scientific knowledge. As such, the behaviour-space may steer investigations towards the discovery of heretofore rare or unseen phenomena, in an attempt to better understand the limits of behaviour or biological mechanisms in general.

6.6 Opening the door to integration

I want to suggest that this re-conceptualisation of ‘behaviour’ could facilitate integration for biological explanations of behaviour, and as such, I understand my approach to be a tool for dissolving apparent pluralism. By seeing ‘behaviour’ as a gradient rather than a binary, it becomes easier for researchers to understand why certain biologists would be inclined to exclude certain phenomena from the classification as a behaviour, and also allows for more constructive discussions regarding what is at stake. Rather then simply rejecting a phenomenon based on intuition, researchers can look to the gradient definition to identify what it is specifically about the phenomenon which they consider insufficient to warrant the label of ‘behaviour’, and why others may see it differently. It clarifies what biologists are talking about when they are referring to behaviour in general, and gives tools for understanding in finer detail why they would classify specific phenomena as behaviours or not.

As mentioned earlier, this can have significant repercussions regarding the types of explanations which are understood to be applicable in any given case. This therefore represents a more general strategy for dissolving pluralism employed in chapters 3 and 4: rather than tackle the plurality of explanations themselves, we can instead look at the explanandum phenomenon to evaluate how, or to what extent, various researchers are indeed talking about the same phenomenon. In this case, I am suggesting that phenomena which can seem to some as being different insofar as they are classified as behaviour or not can instead be seen to be part of the same spectrum. This in turn relates to integration, since one of the ways integration can be carried out is through the extension of argument patterns into novel areas (e.g. unificationism; see section 5.2.2); for instance, the use of argument patterns stemming from behavioural ecology to explain certain botanical phenomena (see e.g. Cahill, 2019) is an example of integration which could benefit from the definition of behaviour presented here, showing how plant behaviour is an extension of our understanding of animal behaviour (see section 6.5). This understanding of the multiple dimensions which enter into the definition of behaviour also sheds light on the various mechanisms which could enter into explanations of behaviour, as well as the links between them. This could facilitate communication across approaches which focus on different aspects of behaviour, such as neurological, environmental, or developmental aspects, possibly dissolving explanatory pluralism for any given behaviour.

Faced with the disagreements regarding definitions and classifications of behaviour, some have attempted to propose new definitions (Levitis et al., 2009), and others have taken it as a demonstration of disunity within biology (Longino, 2013, p. 151). My conception suggests that the real problem is in the binary assumption, and that understanding the term as a multi-dimensional spectrum effectively bridges conceptual gaps between disciplines. Many researchers have proposed various integrative strategies which often rely on the alignment of concepts used in different disciplines (see chapter 5). Graded concepts such as mine and Godfrey-Smith's (2013) also contribute to the integration of disciplines and concepts, adding another tool to the integrative toolbox.

CHAPTER 7

TEMPERED INTEGRATIVE MONISM

Up to this point I have criticized pluralist approaches and explored integrative strategies relating to scientific explanations. With these building blocks in hand, I now turn to a defence of my own position, which I call tempered integrative monism.

Section 7.1 describes how pluralists tend to attack monism without ever being able to convincingly point to one in contemporary literature. I propose to remedy that problem by describing my own explicitly monist position. To do so, I begin by pointing out that, contrary to pluralist contentions, *both* pluralists and monists must rest their arguments on a priori grounds regarding the future of science. Section 7.2 discusses how pluralists are scientific realists, and what that entails in terms of commitments regarding the nature of explanations. In section 7.3, I argue that the commitments entailed by scientific realism invariably open the door to commensurability and integration, showing how Longino's insular pluralism is untenable. Section 7.4 tempers the monism I propose by describing certain practical limits to scientific explanations and their integration, and section 7.5 ends the chapter by revisiting the cartography analogy cherished by pluralists, subverting it to defend monism.

7.1 Monism

7.1.1 Who is a monist?

As was described in chapter 1, pluralists do not always agree on what it means to be a pluralist. While some emphasize the variety of *types* of scientific explanations, others contend that having multiple explanations of a given phenomenon is unproblematic, and others still make a case against the unity of science more generally. The one thing that they all agree on is that monism is a wrongheaded position and should be abandoned. Yet quite often, there are few or no references to contemporary philosophers who explicitly defend a monist position (Ruphy, 2013, pp. 160–161 makes a similar point). Some explicit references to monist positions include Mitchell (2009, pp. 11–12), who calls on philosophers from centuries past, such as Mill (1843), Herschel (1830) and Whewell (1840) as the standard-bearers of a unified science. Longino, for her part, criticizes Oyama (2000a, 2000b), Griffiths (Griffiths & Gray, 1994; Griffiths & Stotz, 2006), Stotz (2006), and even Mitchell (2002), for holding positions which are allegedly underpinned by monist convictions. It needs to be pointed out however that none of those authors would agree that they are covert monists, and that they would in fact probably (and sometimes do, explicitly) consider themselves pluralists. Who are these so-called monists and what positions do they defend, whether explicitly or implicitly? In this section, I begin by reviewing some of the monist positions which are set up as foils against which pluralists compare their positions, and explain how each fails to be a compelling competitor to pluralism.

One kind of monism which is often attacked in discussions relating to the biology of behaviour is the genetic reductionist monist. This monist is portrayed as believing that all behaviour can, should, and will be explained solely through the genes. Mitchell, for instance, after having discussed the complex etiology of depression, remarks that "at this point, it appears naïve to believe there is a "gene for depression", which would explain the malady through a traditional reductive strategy." (2009, p. 7) Though genetic determinism has no doubt been historically defended (e.g. Galton, 1883), such a simplistic reductive strategy has faded with the rise of the interactionist consensus. Even Morgan, working in the early twentieth-century on inheritance of eye colour in fruit flies, acknowledged that many genes and many elements along the developmental process could affect the outcome of the phenotype, despite the fact that his experiments were designed to single out a single difference-making gene (Waters, 2007). By now it is considered a truism that complex traits including behaviours are the result of an interaction of nature and nurture (Tabery, 2014; Tabery & Griffiths, 2010). Though reductionists surely will defend the idea that behaviours can be reduced to their constituent parts, this does not imply that genes are the only relevant constituents. In this respect, contemporary approaches of the sort cannot be considered monists of this sort, insofar as they explicitly acknowledge the role of multiple other factors.

Another monist position is understood as believing that if there are multiple approaches, or multiple explanations for a single phenomenon, only one of those explanations will eventually triumph, and falsify the others (Kellert et al., 2006b, p. xv). This is illustrated in situations where a given behaviour can be explained through a behavioural-genetic approach, as well as a socio-environment approach. Longino (2013, pp. 139–141), for instance, highlights the work of Pinker (2002), who portrays his favourite approaches, viz. behavioural genetics and evolutionary psychology, as undervalued disciplines, embattled minorities that are constantly under fire from other approaches. In his view, the import of these approaches needs to be brought to the forefront. The opposing camp, however, sees things in the completely opposite manner: Gottlieb (1995), Baumrind (1993), and Maccoby (2000) defend socio-environmental approaches and represent "the behavioural genetic approach as the dominant perspective whose influence must be curbed" (Longino, 2013, p. 141). According to Longino, the perspective of both these camps are demonstrations of an underlying monism, since the defenders of each approach would like to eliminate the other (see also Mitchell & Dietrich, 2006 for a similar argument). But as Beatty (1997) has pointed out, these debates are more aptly framed as being about the relative significance of various approaches, rather than an attempt at imposing a single theoretical framework. And the polarization of the positions likely concerns socio-political matters such as funding and prestige, more than the denial of all explanatory merit to their opponents. After all, the well-known consensus is interactionist, with all researchers conceding that genes and environment interact in complex ways to produce the observed behaviours. Pinker (2002), for instance, does cite environmental research he agrees with, as Longino herself points out. And Maccoby's (2000) article stresses only that studying genetic causes without taking into account the feedback loops with the environment will yield erroneous conclusions about heritability. Monists are therefore very unlikely to assume that any single existing discipline is in a position to eliminate all others, acknowledging that the various disciplines contribute to our overall understanding.

Finally, fragmentation pluralists such as Mitchell conceive of another brand of monism that would call for the unity of science through an explanation large enough to encompass all that biology has to explain (see chapters 2 and 3). Kellert et al. (2006b) also describe scientific monism in such terms, describing it as the view that

> the ultimate aim of a science is to establish a single, complete, and comprehensive account of the natural world (or the part of the world investigated by the science) based on a single set of fundamental principles [and that] the nature of the world is such that it can, at least in principle, be completely described or explained by such an account (p.*x*)

The difficulty with this conception of monism is not only that no contemporary researcher seems to espouse it, but also that it is difficult to understand precisely what such an explanation would entail. Indeed, what would be an explanation of biological phenomena in general? Or an explanation of all the behaviour-related phenomena? As Ruphy (2013, p. 161) points out, the questions one can ask about any given phenomenon seem to be infinite, meaning that an explanation which satisfies this characterization of monism would need to be able to answer a seemingly infinite number of questions. What Mitchell has in mind is a "grand theoretical unification" (2003, p. 207), presumably akin to certain physicists' ideal of the Theory of Everything, whereby a single theory of biology would be able to explain all biological phenomena. Yet Ruphy argues that such an ideal is contentious even within physics (never mind biology), and that even hard reductionists such as Oppenheim and Putnam (1958) did not espouse such a view. Though this idea is not illogical per se, no defender (contemporary or otherwise) is put forward by Mitchell, and indeed, none is apparently to be found.

Faced with this hall-of-mirrors of monists, it goes without saying that pluralism seems like the most reasonable approach. According to these portrayals, if you do not want to endorse genetic determinism, or the idea that behaviours can be explained *only* through behavioural genetics—or *only* through socio-environmental approaches—then pluralism is right for you. And if you believe, just as Mitchell does, that biology will not be completely explained by a single explanation or theory, then surely you are a pluralist. Yet, to my knowledge, none of these positions are actually defended by any researcher, so we are left wondering if there really exists an alternative to pluralism.

I propose to remedy the situation by elaborating an explicitly monist approach to science, more specifically to the biological explanations of behaviour. Though what I propose is novel, it must be said that it builds on previous proposals which either have grand unifying views for specific disciplines (Craver, 2007), or for science in general (Neurath, 1937; Potochnik, 2011). I was also inspired by others who have highlighted ways in which explanations or theories can be unified (Darden & Maull, 1977; Faucher, 2014; Grantham, 2004), or how different perspectives on a given phenomenon can or cannot be compatible (Giere, 2006; Massimi, 2022). But it is significant that none of these authors has come out as being a monist, and indeed, virtually all call themselves explicitly pluralists. And while some of these authors have been tackled in the previous chapters, I have given most attention to those whom I have considered to best represent fragmentation and insular pluralism specifically within behavioural biology, namely Mitchell and Longino. And while I believe that this brand of monism is likely to be applicable across the board for science in general, the task of demonstrating so will be for another time.

7.1.2 I am a monist

The monism I propose uses integration as the glue to unify scientific explanations of behaviour. Whereas pluralists who espouse integration do so with the caveat that there are important limits to integration, I defend the idea that there are, in principle, no barriers to integration. It is important to emphasize here that what I am defending is “in-principle” integration across the board, and therefore that the many pragmatic limitations to integration—such as considerations related to ethical concerns or the accessibility of the subject matter—are acknowledged but considered to be outside of the purview of this thesis. My justification consequently relies on reasoning concerning scientific explanations, their relation to reality, and how integration is a reasonable, realistic, and desirable goal for scientific research on the biology of behaviour, though tempered by pragmatic limitations.

The objective of this chapter is to demonstrate that it is reasonable to think that there are no barriers to integration. Previous chapters have already criticized pluralist positions, and as such, those criticisms will not be covered here. Instead, I will argue that the way scientific explanations attempt to describe the world as it actually is makes the case for the possibility of integration in biological explanations of behaviour. The constant progress and dynamism of scientific research further contributes to breaking down any possible incommensurability. Rather than focusing exclusively on current scientific knowledge, the monism I propose concerns the very possibilities of scientific explanations, and as such makes a commitment about the state of future science, just as do the pluralists when they claim that explanatory pluralism is here to stay. But it also makes a case for the benefits of integrative research, and how it is not only a possibility, but also a desideratum for scientific advancement.

I will begin by discussing how even pluralists espouse scientific realism, which implies that the value of scientific explanations is at least partially due to the fact that they are about the actual world. I argue that this position brings along with it empirical and ontological commitments about the entities and processes called upon in scientific explanations, and how this lays the groundwork for the possibility of integration. The recognition of these commitments is the starting point for integration in practice, which relies on communication and common ground between researchers, resulting in a dissipation of assumed incommensurability between scientific explanations. I conclude by discussing some of the pragmatic constraints which are the true impediments to integration and the unity of science I propose.

7.1.3 *A priori* arguments for the future state of science

Pluralists often claim that their pluralism is based on empirical considerations, saying that contrary to their monist opponents, they do not make a priori claims about what the future of science should look like. According to pluralists, monists make assumptions about the development of science such that they "misinterpret the phenomena about which they write" (Longino, 2013, p.144), leading them to erroneously conclude that monism is possible. Contrary to monism, pluralism is understood as far more modest, accepting only that whether pluralism or monism is possible "is an open, empirical question." (Kellert et al., 2006b, p. x) In an earlier book, Longino explicitly draws this distinction between monism and pluralism: "the pluralist claim […] must be understood as an empirical claim, not as the expression of a necessary truth about the world. Pluralists in philosophy of science base their arguments on cases in contemporary science, not on a priori arguments." (Longino, 2002, p. 94) Thus by her lights, while the monists must propose a priori arguments which insist on the in-principle possibility of unification of scientific explanations, pluralists are simply looking at contemporary science and stating the facts.

Pluralism is sometimes characterized as a stance to adopt concerning the plurality of explanations (Kellert et al. 2006b), the idea being that it purportedly makes no commitment about the future state of science, and is only a prescriptive stance which encourages open-mindedness. Mitchell, for her part, goes further and sees her integrative pluralism as a "research program" (2009, p.13), appealing to complexity as the basis for plurality, and that this complexity will entail that pluralism is here to stay (2003, p.3). Interestingly, Longino (2013) tempers the position she put forward in previous publications, and more rightly describes *both* pluralism and monism as "attitudes towards the multiplicity of approaches" (p.138). This change of position regarding the distinction between pluralism and monism suggests that Longino may very well agree with me that, in reality, both pluralists and monists base their arguments on contemporary empirical science, and extend their claims into an a priori understanding of what is to come for the future of science.

To clarify what is at stake, Kellert, Waters and Longino propose a distinction between plural*ity* and plural*ism* (2006, p.*ix*). Plurality is taken as a fact of current science: there currently exist multiple scientific approaches which attempt to explain certain phenomena (as seen e.g. in chapter 4). Mitchell as well begins with this understanding: "the fact of pluralism in science is no surprise." (2003, p. 2) The variety of models, explanations and theories is understood to be an uncontroversial starting point for discussions about monism and pluralism. In contrast, plural*ism* is a way of seeing the plurality such that it is understood as ineliminable, and the result of successful science running its course.

What the distinction between plurality and pluralism highlights is precisely the fact that all three of the characterisations of pluralism seen above (as a stance, research program, or attitude) go further than mere observation of empirical science. While fine-grained distinctions could be made regarding the differences between the three, they are in all important respects equivalent insofar as each is underpinned by a commitment about the future state of science: that it will remain pluralistic, and that the plurality is desirable and a mark of success. Kellert, Waters and Longino make these claims quite explicitly:

> we do not believe that the plurality in today's science is necessarily a temporary state of affairs. We think that some phenomena may be such (e.g., so complicated or nebulous) that *there can never be* a single, comprehensive representation of everything worth knowing, or even of everything causal (or fundamental), about the phenomenon. If this is the case, that is, if the nature of the world is such that important phenomena cannot be completely and comprehensively explained on the basis of a single set of fundamental principles, then the aims, methods, and results of the sciences should not be understood or evaluated in reference to the monist quest for the fundamental grail. (Kellert et al., 2006b, p. xi; my emphasis)

These claims are somewhat dissonant when compared with their posture concerning pluralism as a stance; indeed, that prescriptive ideal is clearly based on an understanding of what the future of science *can* look like, and not only what it *should* look like. More precisely, their claim is that the future of science will not resemble what the monists put forward. Mitchell's 'research program' view is also explicit about this commitment, suggesting that scientific research needs to adopt pluralism in order to progress most fruitfully. The 'stance' and 'attitude' views are more implicit, couching their approach in lightly prescriptive language; yet the result is the same: science ought to adopt pluralism, because it is the best way for science to be successful.

In this respect, pluralism, just as monism, rests on in-principle arguments. Whereas monists, such as I, will claim that monism is possible in principle, pluralists will argue that monism is impossible, even in principle. No monist could claim that the current state of science is characterized by monism; indeed, the plural*ity* pointed out by Kellert and colleagues is not disputed—how could it be? What monists claim is rather that in principle, and quite often in practice, the plurality can and will be eventually be eliminated, through some form of unification of theories or explanations, or whatever explanatory units the monist is focused on. In a similar way, pluralists need to claim that in principle, and certainly in practice, the plurality will remain forever in science; for instance, as previously quoted, Kellert et al. (2006b) propose that certain comprehensive explanations will "never" be attainable. Because they are making claims about the future state of science, their arguments cannot be merely empirical, nor purely based on current practice. They cannot remain empirical, because we have no empirical data about the future state of science, only predictions based on in-principle arguments. And the pluralism cannot be purely based on current scientific practice, because it must make claims about how—regardless of the future practice of science—certain explanations will forever be plural. And these in-principle arguments are precisely the kind that Mitchell deploys (as seen in chapters 2 and 3) to propose that reduction will be impossible in certain cases, and that a multiplicity of models will be necessary in certain cases. Longino's arguments as well rest on the fact that the parsing of the causal space and the measures of conformation are such that even in principle, there is no integration possible between approaches, due to incommensurability (as seen in chapter 4).

Pluralists and monists are therefore on the same ground, attempting to show that their approach is the more reasonable one through a priori arguments about the future state of science. With this in mind, my objective in this chapter is to show how my monist position is a more reasonable attitude with respect to expectations regarding the future state of science. But just as pluralists, it is not a position which can be proved beyond the shadow of a doubt, since we will have to wait and see what the future holds; nevertheless, in the meantime, we can make informed predictions about what is in store. Mirroring the long quote from Kellert, Waters and Longino above, I rest my arguments on the fact that "the nature of the world is such that" phenomena can and will be explained completely and comprehensively through integration.

7.2 Scientific realism

The foundation of my argumentation rests on the acceptance of scientific realism. Scientific realism is the position that our best scientific explanations tell us something truthful about the world, and that we should therefore hold an epistemically positive attitude regarding the observable and unobservable entities postulated by those explanations; by and large, we should expect the best scientific explanations to be accurate representations of the world. Another way of characterizing this is in terms of scientific aims: scientific realists will suggest that the objective of scientific explanations is to "produce true descriptions of things in the world" (Chakravartty, 2017). In the opposite camp are antirealists, who suppose that the success of scientific explanations does not rest on their correspondence with reality. There are, of course, many different ways of being an antirealist, most of which adopt some form of 'instrumentalism': theories, or explanations, are successful because they are useful for certain purposes, but we cannot deduce from such usefulness anything about reality itself. Words which refer to unobservable entities postulated by theories are considered to be neither true nor false, but simply have no meaningful referent (Devitt, 1997, p. 128).

Though, as will be shown shortly, pluralists are realists, and it is therefore not necessary to propose arguments in favour of realism for the sake of my thesis, there is an interesting argument to be made regarding explanatory integration and realism which is worth the detour: the possibility of integration is a strong argument in favour of scientific realism. Scientific realism is often defended using the "no miracles" argument, which proposes that realism "is the only philosophy that doesn't make the success of science a miracle" (Putnam, 1975, p. 73). An extension of this argument is the corroboration argument: "if an unobservable entity is putatively capable of being detected by means of a scientific instrument or experimentation", as well as another which is theoretically and practically different, then there is good reason to think that the entity exists (Chakravartty, 2017; see also Hacking, 1983, 1985). This would be the simplest explanation for its detection through various means. However, van Fraassen (in Van Fraassen et al., 1985, pp. 297–298) proposes that just because we could assume that the entity exists does not imply that the explanation is true. My argument from explanatory integration resolves this problem: it is not merely the entities postulated which must exist in order for integration to be possible; the explanations themselves must be at least approximately true in order for integration to be possible.

After all, what would warrant the expectation that integration is a possibility if not for the fact that the multiple explanations to be integrated tell us something (at least approximately) true about the same phenomenon? The instrumentalist's alternative is that each explanation is successful in its own way, by its own lights, and that the success is not predicated on a description of the phenomenon as it actually is, but only on its usefulness for our given purposes. This implies that if integration of multiple explanations works, it is the result of mere happenstance: the explanations happen to be related to one another in appropriate ways but (importantly) not because they are accurately describing the phenomenon. From an instrumentalist's point of view, it is not clear what could explain the successes of integration, since each explanation is successful not by virtue of describing the world as it actually is, but by virtue of satisfying its own internal criteria for 'usefulness', which may be completely different from one explanation to the next. Integration thus can be understood as a strong argument in favour of scientific realism, since it is the fact that explanations are both talking about the same phenomenon and the same entities and processes which allows integration.

In line with this argumentation, what I suggest here is that scientific realism implies that explanations will commit to the fact that elements in their explanations reflect what is actually in the world. This can come in two forms: empirical commitments, and ontological commitments. These commitments must further-more be accepted by pluralists, since they are themselves scientific realists.

7.2.1 Pluralists are realists

As mentioned, my position takes for granted scientific realism. I do not, however, need to argue for it extensively, as pluralists in general tend to be scientific realists themselves. Indeed, Longino states clearly that "realist nonmonism is pluralism" (2002, p. 93), meaning that to be a pluralist is precisely to be a scientific realist which is committed to an opposition to monism. Mitchell as well is explicit, arguing "for a pluralist-realist approach to ontology, which suggests not that there are multiple worlds, but that there are multiple correct ways to parse our world" (2009, p.13). In a similar vein, Longino states that "the pluralism envisioned by theoretical pluralists is a pluralism of theories of a singular world." (2002, p. 94) When speaking of a singular world, what these authors mean is that metaphysically, there is only one world which science attempts to describe, as opposed to a multiplicity of worlds. If there were multiple worlds, then pluralism could be expected since we could have many different and incompatible explanations which each explain their own world. Longino would presumably agree then with Mitchell when she states that "however complex, and however many contributing causes participated, there is only one causal history that, in fact, has generated a phenomenon to be explained" (2003, p. 65). Pluralism is therefore a position which takes for granted that science is in the business of describing the (one and only) world, but that attempts to allow for the multiplication of explanations thereof.

To do so, both Mitchell and Longino attempt to reconcile the fact that science explains the world, with the fact that there are many ways of 'seeing' that world. Longino for instance tempers her realism in the following way:

> "World" can mean the whole of all there is, in which case the pluralist would agree that there is only one of those. But it can also mean the collection of aspects of the world that is salient to those approaching it with a given set of assumptions and strategies for acquiring knowledge, not to mention a given sensory and cognitive apparatus. In this sense there are many worlds (2002, p.94).

In her subsequent book (2013), Longino fleshes out how the parsing of the causal space can be different for different approaches, leading to what she here calls "many worlds". This also harks back to discussions in chapter 1, which highlighted the role that the partiality of representations plays in justifications of explanatory pluralism: because explanations are partial, this is thought to 'leave room' for a multiplicity of correct (but partial) explanations of a single world, or a single phenomenon.

In the next section I will argue that this commitment to scientific realism is incompatible with the two stronger kinds of explanatory pluralism, namely insular pluralism and fragmentation pluralism. (This argument is, however, not related to type pluralism, since scientific realism does not imply anything definitive about the types of explanations it will be possible to put forth.) To pave the ground for this argument, I will first make explicit what sort of commitments scientific realism implies when it comes to explanations.

7.2.2 Empirical commitments

Because scientific explanations are about the world, they entail certain commitments about how the world actually is. The most obvious of these are the empirical commitments, which relate to the observations and measurements that go into researching the explanations, as well as the predictions which result from those explanations. Every scientific explanation is based on empirical data, and through this, the explanations commit to the fact that the data is a true measurement of the world as it actually is. As Longino puts it: “The one standard that is common to any scientific community is empirical adequacy, that is, truth of the observationally determinable portion of theories or models.” (2002, p.185) What is measured is uncontroversially taken to be true about the world. And as Potochnik adds: “evidential relationships do not respect field boundaries” (2011, p. 307). In other words, these truthful data points about facts in the world are usable by any approach, since they represent something about the world which is impossible to contradict.

Of course, as seen in chapter 4, the measurements and predictions come bundled with methodologies for measurement and operationalisation, meaning that care must be taken when comparing or using data across research contexts. But as seen in chapter 5, these challenges to data integration are being tackled through more and more sophisticated methods, facilitating the pooling and communication of data. And the empirical commitment as well comes bundled with the measurement and operationalisation: what can be stated with certainty is that such and such a measurement technique will yield the same results, and that tells us how the world actually is.

Potochnik (2011) mobilises work done by Takahashi and colleagues (2008) on peacock trains (their elaborate feathered tails) to illustrate this point. Darwin (1871) proposed that birds with very showy plumage could be the result of sexual selection. Peacocks are the prime example of such plumage, and their large and colourful trains have long been assumed to be the result of selection for elaborate plumage through female mate selection (e.g. Petrie, 1994; Petrie et al., 1991). Takahashi et al. (2008) used methods from field ecology, observing a peafowl population over many years, with the data showing that characteristics of the male train were not correlated with female mate choice. They furthermore noted evidence from other studies that the male trains will develop in the absence of oestrogen, "and conspecific females generally disregard these traits [under oestrogen control] in mate choice" (p.1210). What's more, they evoke research stemming from molecular phylogeny suggesting that in the distant past, both male and female peacocks had bright tail plumage; thus it is not male peacocks which developed elaborate trains, but rather that female peafowls *lost* their colourful plumage over time. These data points, stemming from multiple disciplines, undermine the hypothesis that peacocks are under sexual selection for elaborate trains.

Of course, Petrie and others whose research were the target of Takahashi et al.'s article responded (Loyau et al., 2008), calling into question the generalizability of the results, their interpretation, as well as whether the measurement configurations covered all the relevant aspects of the male train elaborateness (length, number and density of eyespots, symmetry, etc.). Importantly, what is never called into question is the data itself; it is taken for granted that the observations done by Takahashi and her colleagues represent reality as it is, and need to be incorporated into a more complete explanation of peacock trains. As such, the conflicting interpretations and conclusions are not seen in a positive light, and it is understood that a successful explanation will be able to include all the relevant data. For instance, one suggestion to account for the apparently conflicting data is that the Japanese peacocks studied by Takahashi's group could have peculiarities not shared by other peacocks (Loyau et al., 2008, p.1).

The fact that observations and predictions of scientific explanations represent commitments about how the world is, is practically a truism. Even scientific antirealists are typically ready to grant that entities which are observable do in fact exist. The important point for my position is how reference to observables allows multiple research groups to settle on common data which must be accounted for in their explanations, even in cases where the explanations put forth are at odds.

7.2.3 Ontological commitments

From the point of view of a scientific realist, scientific explanations not only come with empirical commitments, they also involve ontological commitments. These are not as straightforward to identify, as they are not mere observations or predictions; rather, they are commitments about the existence of unobservable entities or processes. And this is typically the real distinction between realists and antirealists: whereas antirealists may be ready to commit to the reality of observables, their claim is that unobservables are merely instrumental in explanations, and do not refer to anything real. Scientific realists, including pluralists, commit to the fact that the unobservables postulated by our best scientific explanations either are likely true, tend towards identifying real entities and processes in the world, or at least have the objective of doing so.

The greatest difficulty with ontological commitments is that they can be held implicitly, and their effect may be obscured when not recognized as such. Hochstein (2019) points out the role of what he calls "metaphysical commitments" in the study of psychological mechanisms. These commitments

> must go beyond what is empirically justified, but are nevertheless necessary in order to set up experimental protocols, determine which variables to manipulate in experimental contexts, and which conclusions to draw from our scientific models and theories. (p.580)

He uses the example of neurological studies on emotion, where researchers must assume that particular emotions exist as entities or processes so that, for instance, 'anger' can be identified in brain scans. Because only the effects of anger and not anger itself can be observed, its existence as a (more or less) discrete emotion must be posited in order to design experiments around it.

Another example more closely related to this thesis is Réale et al.'s (2007) proposal to incorporate the study of individual animals' temperament into behavioural ecology. A good part of their article is spent defining what is meant by 'temperament', in an attempt to circumscribe as precisely as possible what they are researching, and how other research groups could research the same phenomenon. They propose that "temperament, personality and individuality [treated here as synonyms] describe the phenomenon that individual behavioural differences are consistent over time and/or across individuals" (p.294). This was a novel approach at the time, since classical optimality models in behavioural ecology typically took individual variation as 'noise'. Rather than disregarding this variability, Réale and colleagues suggest that an individual organism could have a propensity to certain behaviours, and that that propensity could be a selected trait.

As such, Réale and colleagues propose a series of ontological commitments. The first is that animal personality in fact exists, and the second that it can be the target of natural selection. They furthermore propose ways in which temperament can be operationalized and measured, committing to the fact that these measurements do in fact relate to the entity they are positing, namely animal personality. Indeed, they propose that temperament traits could be divided into 5 categories: shyness-boldness, exploration-avoidance, activity, aggressiveness, and sociability. These categories are meant to facilitate communication across research contexts, but the authors clearly admit that the list may not be exhaustive, and presumably could be revised in light of future research. This highlights the fact that their strongest commitment is not to the measurement methods proposed, but only to the existence and relevance of temperament in individual animals, despite the fact that it is not directly observable. In other words, though they are quite sure that animal personality exists, its existence may not yet be clearly delimited, and its contours will need to be refined through successive operationalisations, measurements, and revisions.

Ontological commitments are an integral part of many scientific explanations (perhaps even all scientific explanations), and represent the fact that explanations rely on the postulated existence of unobservable entities. These entities are assumed in order to make sense of the observable data, and are thus an integral and necessary part of the explanation, as well as being revisable in light of new data. Making these commitments explicit is an important step in fostering cross-research communication, as it enables researchers from various approaches to more effectively communicate about the assumptions that enter into explanations.[26]

7.3 Realism, explanations and (in)commensurability

Longino proposes that explanations can be incommensurable, which is the justification for her insular pluralism. As described in chapter 4, that incommensurability is based on the differential parsing of the causal space by different approaches (2013), as well as her views on the conformation of scientific explanations (2002; 2013). According to her, each approach has its own way of carrying out measurements and of grouping together the possible causes for a phenomenon, as well as their own way of measuring the success of their explanations. That success is furthermore not understood as a binary of truth or falsity;

[26] It is important to note that empirical and ontological commitments are characteristics of explanations qua representations, and are not an underhanded way of introducing an ontic conception of explanation à la Craver (2014). My intent instead is to show that *despite* the epistemic conception of explanation put forward by the pluralists (which I too endorse), if we are realists, then there must be reference relations between the representations and reality, and that these reference relations will be sufficient to lead to integration (as will be argued in 7.3).

instead, it comes in degrees of conformation. According to Longino, these elements explain how two (or more) explanations can be incompatible and yet both correct, as well as why that incompatibility will be made permanent through incommensurability. I propose that though these aspects of scientific explanations are interesting, and indeed, important for a good understanding of scientific explanations, they do not imply definitive limits to commensurability, due to scientific realism.

7.3.1 Points of contact

The first thing to note regarding the possibility of perduring incompatibility is that it may not be logically compatible with incommensurability. As Sankey (1994) points out, if there truly is incommensurability between two explanations or theories, then it is unclear what it means for them to be incompatible, or in conflict: “This stems from the absence of logical conflict between incommensurable theories due to their formulation in different, untranslatable languages.” (p.3) What this means for Longino’s account is that if an approach produces explanations which are impossible to compare or relate to explanations coming from another approach, then there seems to be no way of determining if those explanations really are incompatible. But recall that for Longino’s pluralism, and insular pluralism more generally, *incompatibility is necessary*: for insular pluralism to obtain, we must be faced with multiple incompatible yet correct explanations. If there is no incompatibility, then we are merely faced with multiple correct explanations which are all compatible one with the other, with no trace of insular pluralism. There is thus good reason to believe that the multiple explanations are not *entirely* incommensurable, otherwise there would be no apparent contradiction.

And the reason why incompatibility is apparent is because the various explanations are talking about the same phenomenon. This is also why true incommensurability never arises: because the multiple explanations take on commitments about reality which can be shared or contested, which I call *points of contact*. If we agree with scientific realism, then scientific explanations tell us something approximately true about the world. And if the plurality of explanations are all referring to the same phenomenon and attempting to explain it, then there will necessarily be points of contact among the various commitments taken on by those explanations. The explanations will be referring to certain parts, or elements, of the explanandum phenomenon which are evoked in both explanations. Here again: if they do not refer to *any* parts or elements in common, then there is no possible incompatibility, since they are simply not talking about the same thing(s).

Of course, this does not mean that the multiple explanations will necessarily *agree* about those commitments, and this is why incompatibility is possible. In the case of empirical commitments, a researcher may call into question the validity of the measurements of a rival explanation. At the extreme, researchers may consider the explanation to be wrong because it is based on bad data. And this of course may very well be the case, whether it be through error or fabrication! But this is not what Longino's pluralism is about: it is about having multiple *correct* explanations, not one correct explanation and one erroneous or fabricated one. More in line with Longino's perspective are when rival researchers call for increased accuracy or completeness of measurements. For instance, in the case of peacock trains described above, Loyau et al. (2008) propose that certain crucial measurements may have been left out, which could impact the results discussed in Takahashi et al. (2008). Now, while it is true that these disagreements highlight how the observations from other researchers or other approaches will not automatically or always easily lead to integration, it does show that researchers understand that they are talking about the same phenomenon, and that there therefore is incompatibility. This recognition opens the door to discussions relating to how those measurements could in principle be made to be incorporated into a more complete explanation.

Ontological commitments as well can be contested. This is the case for instance with Réale et al.'s (2007) introduction of animal temperament discussed earlier; it contests the fact that optimality models ignore this factor. As Réale and colleagues remark: "ecologists generally do not perceive temperament as an important addition to our understanding of the ecology and evolution of animals." (p.292) The source of conflict here is therefore whether or not temperament exists and whether it is a significant element in explanations of behaviours. Here again, though the integration of these commitments in rival explanations is not a given, their recognition as such opens the door to communication across approaches. This relates to the behavioural gambit described in chapter 5: research on temperament is one way among many to shed light on the mechanisms underlying behaviour, giving us a more accurate explanation of animal behaviour. Because of the assumptions behind scientific realism, a successful explanation will need to take into account reality as it actually is. And if it turns out that animal temperament truly exists (as the proliferation of publications on the topic suggests), then optimality models will need to eventually take them into account if they are to offer more detailed explanations. (Though as will be explained shortly, more detail is not the only possible desideratum for explanations.)

Thus, the links between the various commitments made by each explanation are both what make the incompatibilities salient, but they are also what open the door to integration. What I am attempting to demonstrate here is not that science, at any given moment, can integrate explanations at will. It is rather that when incompatibilities occur, it is seen as a problem in need of resolution, and that the resolution will come about by focusing on the empirical and ontological commitments made by the various explanations, more specifically where there are incompatibilities. In other words, incompatibility is present specifically *because* the multiple explanations are targeting some of the same elements in the phenomenon at stake. And once that incompatibility is made clear, then researchers will not be content to accept it; they will work towards its resolution. That resolution can come in many forms, be it through the elimination of erroneous explanations or interpretations, through the modification of one or both of the explanations, and/or through the integration of the explanations.

7.3.2 Does conformation save incommensurability?

While the above argument is certainly convincing if we take for granted that scientific explanations give us an accurate picture of the world, Longino would point out that explanations are not such clear and complete representations of reality. Her approach proposes that due to the differential parsing of the causal space, as well as the way each approach has its own measure of conformation, each explanation has its own way of seeing the world, and therefore that explanations can be simultaneously incompatible and correct. These two notions together are understood as divorcing, to some extent, the explanations (qua representations) from the world itself, giving some wiggle room for different incompatible explanations to co-exist.

The first thing to note regarding this objection is that I do not contest either the parsing of the causal space nor conformation as adequate, and indeed, valuable ways of understanding scientific explanations. It is true that different approaches will develop their own measurement setups which will have an impact on their conclusions. I also grant that, to some extent at least, approaches have their own priorities when it comes to evaluating the success of their explanations. And these elements can make it difficult at any given time to compare results and explanations across approaches. After all, if two approaches measure the same, or similar, aspects of a phenomenon in different ways, then their data will not be straightforwardly integratable. And *from within* a given approach, it will not be obvious just how successful explanations from another approach ought to be considered. In this respect, Longino is right to point out that a certain kind of incommensurability (defined as a lack of common measure) is possible since the approaches literally have different measures. However, I contend that any such incommensurability will be temporary. The reason why incommensurability will be temporary is because it will always, in principle, be possible to find points of contact between incompatible explanations, due to the commitments implied by scientific realism.

Let me begin by tackling the issue of the partiality of representations, and conformation. In brief, though conformation does imply a certain distance between the empirical data and the representation meant to explain it, it does not imply imperviousness to empirical findings. This is, indeed, the very nature of 'representation'. As has been shown throughout this thesis, pluralists rely heavily on the fact that scientific explanations are representations in order to justify their positions. More specifically, the *partiality* of representations is taken as an important aspect of explanations which opens the door to the possibility of pluralism. According to pluralists, because explanations never capture a phenomenon in its entirety, there is a certain disconnect between the explanation and the world, and it is that space that they point to between the explanation and the world itself to justify the incommensurability and plurality of scientific explanations. But by its very definition, the nature of partiality also implies that the explanations *do* capture (in part) the world as it actually is. In other words, with the notion of partiality of representations comes the commitment to scientific realism, since in order for the explanation to be related even partially to the world, it must be describing parts of the world as they actually are. And indeed, as was shown earlier, pluralists, including Longino, do portray themselves explicitly as realists. What this implies is that while scientific explanations may indeed be mere partial representations, it remains that they are explanations only insofar as they tell us something about the world as it is.

And this is so irrespective of the way conformation is evaluated. Even though the *measure* of conformation is dependent on pragmatic and discipline-specific interests, the representations must nevertheless correspond in some respect to the world for there to be any degree of conformation. Longino herself concedes that this is the case. In her discussion about map-making, she concludes that in scientific explanations, just as with map-making, once the conventions and intentions are set, then the success of the map or explanation still concerns whether or not it conforms to reality: "this is not a matter of choice. If there is no conformation, no fit, then we will be lost—in the mountains and in the laboratory" (Longino, 2002, p. 120). Success is not merely a question of opinion, or of choice: if there is no conformation, no relation to the world as it actually is, "reality will eventually bite back" (Longino, 2002, p. 119). Though conformation allows for a looser relation to reality than does 'truth', it remains a relation to reality, with all the constraints that imposes.

Conformation and the partiality of representations thus may complicate the comparisons of explanations across approaches, but does not preclude it, since they will nevertheless be referring to the same phenomena, and (at least some of) the same aspects of the phenomena. There will therefore always be points of contact between the explanations, which will be handles with which to begin the process of creating commensurability. If, on the other hand, there are no points of contact, then once again there is no possible incompatibility, and therefore no possible pluralism; all we are left with are multiple explanations (correctly) explaining different aspects of reality. Conformation is an important aspect to take into account when attempting to compare or integrate explanations, and does complicate the picture, but ultimately, either there are points of contact *and* incompatibility, opening the door to further work to resolve the conflict through the search for commensurability, or there are no points of contact, and in that case, there is no incompatibility and therefore no pluralism. The difficulties posed by conformation may therefore stall the resolution of incompatibility, but they cannot avoid it altogether.

7.3.3 Working towards integration

The difficulties posed by the differential parsing of the causal space are answered by much the same argument, though with the addition of considerations relating to the dynamism of approaches, and the desirability of resolving the incompatibilities.

Longino, once again, rightly points out that the fact that different approaches parse the causal space in different ways poses challenges to comparison and integration. Just as with conformation, the parsing of the causal space can make commensurability more difficult, but because it is still a measurement of reality, then as long as there is incompatibility, there will be points of contact. However, due to the different measurement setups, these points of contact do not imply a straightforward comparison. Longino's example described in chapter 4 is a perfect illustration of this point: because the behavioural geneticists consider intrauterine effects as environmental, while the socio-environmental approaches consider them as non-environmental, there is no simple correspondence between the "environmental" data for each. One thing to remember however is just how much overlap there is: everything else labelled 'environmental' could potentially correspond. In this respect, the two approaches are not entirely divorced one from the other: both measurement setups tell us something about the effects of the environment; enough in fact to trigger incompatibilities between the explanations, and therefore open the door to comparisons. And another important point is that the various approaches are not blind to their own measurement setups: good researchers will know and recognize the limits of their measurement setups, what they imply in terms of possibilities of comparisons with other approaches, and how they could be changed to facilitate commensurability. The painting of researchers as 'stuck' in their methods, scopes, and assumptions overly simplifies the practice of scientific research, which is populated by actual people who can recognize those limits, and reach out to colleagues (or even textbooks) from other approaches to make sense of incompatibilities.

Hochstein (2023) makes a similar point with respect to models in cognitive science. Even in the case where the partiality of representations is predicated on differing idealizations or simplifications, the points of contact are such that integration is nevertheless possible. While certain models simplify away different aspects of a phenomenon, making them apparently incompatible and perhaps even incommensurable with other models, Hochstein proposes that integration "only requires understanding how we can draw coherent *inferences* about the same target phenomenon *across* those models" (p.8, emphasis in original). Using his previous work on metaphysical commitments (Hochstein, 2019; see also section 7.2.3) he shows how the idealizations used in modelling are deliberate, and how researchers keep track of those idealizations in order to refine them as research in other approaches progresses. Indeed, "acknowledging and keeping track of these implicit commitments can be the key to understanding how incompatible models connect to one another" (2023, p.9). Hochstein illustrates this with a few different examples, including models explaining the neurons engaged in cognitive maps. Certain computational models focus on the morphological characteristics of the neurons, "to understand how the neuron will respond to particular inputs, and produce particular outputs" (p.10). Dynamical models, on the other hand, look only at "the dynamic electrical properties of the output spike trains themselves" (p.10), while using biologically implausible morphologies for the simulated neurons. Importantly, both computational models and dynamical models draw inferences from the results of the other, to better refine the idealizations, and ensure that the simplifications remain adequate for the modelling objectives. Researchers are therefore well aware that their models or explanations rely on idealizations, and those idealizations are not created arbitrarily: they stem from an understanding of the limitations of their models, and from an understanding of what other models of the same phenomenon can tell them is important to include or exclude.

The vision of these different measurement schemes or idealizations as barring commensurability relies on a static snapshot of approaches at a given time, and omits the dynamism of research over time. Indeed, while it is true that different parsings of the causal space, and the different measurement setups they imply can lead to sets of data which are difficult to compare as-is, it is not the case that researchers will simply throw up their arms and conclude that there is nothing more to say. Rather, research will continue, leading to the accumulation of more and more empirical data, in the form of observations or predictions, and therefore more and more empirical commitments. Explanations as well will be more and more developed, often refining the ontological commitments through additional measurements, revisions, and re-conceptualisations. These additional commitments open the door to additional points of contact between explanations, and therefore to new avenues for integration.

This argument is all the more potent when one considers that the accumulation of empirical data is not a passive matter. Richardson (2008) raises this issue in his defence of consensus as the telos of science. Leaning on the work of Charles Sanders Peirce, he claims that when conflict between explanations arises, scientists do their best to add to the empirical successes of their theories to get rid of the "irritation of doubt". Faced with incompatibilities, researchers will continue to develop and actively try to resolve those incompatibilities in explanations through a better understanding of the methods, scope, and assumptions of both their own approach, and those of the rival approach. As Allchin proposes, "experiment, then, is viewed not as statically supplying observations to assess one alternative or another, but as actively helping to discriminate between potential solutions." (1991, p. vi) Thus, in the case of incompatible explanations, the researchers will actively look to accumulate additional empirical data in order to falsify their opponents' explanations or shore up their own. And this dynamism applies not only to empirical data, but also to the ontological commitments: "It is important to note that theories are formulated in terms of particular ontologies, but are not determined by them." (Laubichler et al., 2018, p. 10) In the face of incompatibilities, of conflict, or of the 'irritation of doubt', researchers will look to their ontological commitments, and can be ready to change them in the face of new evidence. This highlights how Longino's understanding of the scopes, assumptions and methods of the various approaches is too static, and does not accurately describe the dynamic process of research.

Richardson's argument regarding the epistemic itch is even stronger when one considers a second, pragmatic, argument. Richardson does not even need to make a commitment about the inevitability of the resolution of the conflict for the argument to be persuasive. This is an important point, since as long as the resolution of the incompatibility has not happened, it cannot be proved beyond the shadow of a doubt that it is inevitable in practice; this is why arguments for monism and pluralism must be a priori (see section 7.1.3). What this means is that regardless of one's inclination, the necessary approach is to assume that all competition is resolvable, and that the only way forward is to attempt to resolve it. This is because denying outright that it is resolvable limits the possibilities for future research insofar as each explanation is taken completely independently from the other, and therefore hampers possible avenues for solutions. In effect, it amounts to abandoning the idea that these explanations are testable one with respect to the other and encourages research in isolation on each explanation. In contrast, assuming that comparisons and integration of explanations is possible pushes researchers to look for points of contact and ways to break down purported incommensurabilities.

Another way of looking at it is that no matter who you believe has the burden of proof regarding whether or not the conflict is resolvable, the epistemically and pragmatically necessary thing to do is to attempt to resolve it, to see whether or not it is possible. If a researcher wants to show that the resolution is impossible, they need to try to resolve it, to show that it cannot be done. And if they believe it is resolvable, they will of course work at resolving it.

In line with what I am proposing here, Weber (2002) shows how purported incommensurability can be resolved through explicit attempts at resolving incompatibilities between explanations. He illustrates his argument using the oxidative phosphorylation (ox-phos) controversy of the second half of the twentieth century. The question at stake was to explain how mitochondria could generate useful energy. The first mechanism proposed involved a "high-energy intermediate", a chemical compound which allowed energy transfer (p.2). The second proposed mechanism was "chemiosmotic" and involved "an electrochemical gradient of protons across the inner mitochondrial membrane" (p.2). Both these mechanisms were plausible, both accounted for the empirical data at the time, and have been understood as incommensurable due to their radically different way of understanding what the mechanisms could be (Allchin, 1994; Weber, 2002, p. 7). Indeed, they also match Longino's ideas regarding incommensurability, with each approach parsing the causal space in its own way, highlighting different aspects as significant, and each having its own way of claiming that their approach conformed to a greater degree to reality. They were also clearly incompatible because of the differing ontological commitments: it was evident to everyone involved that the energy was created *either* by the high-energy intermediate, *or* by the chemiosmotic mechanism, but there was no way at the time of distinguishing which it was. It was this conflict which pushed the research community to search for ways of resolving the incompatibility; and indeed, this led to the development of experiments made to determine which of the two theories was correct. The typical interpretation of the resolution, adopted by Weber, is that novel experiments showed that the chemiosmotic approach was correct, leading its proponent, Peter Mitchell, to win the 1978 Nobel Prize in chemistry. Allchin (1994), for his part, portrays the resolution as a partitioning of domains of applicability: while the chemiosmotic mechanism was present in the ox-phos phenomenon, the chemical mechanism turned out to be relevant to other phenomena. Regardless, either interpretation concords with the progression that I have laid out above: incompatibility is identified because of the differing commitments; but despite the purported incommensurability, there were sufficient points of contact to recognize the incompatibility, and therefore see it as a problem in need or resolution. That resolution was eventually achieved through the development of new techniques and new experiments for isolating the consequences of the varying commitments carried by each explanation. In this case, the rival explanation was either eliminated as a possible explanation of the phenomenon (according to Weber), or restricted in its domain of application (according to Allchin); but other cases may very well yield other outcomes, such as the modification or integration of explanations.

7.3.4 Realism and integration

A commitment to scientific realism implies that commensurability will always be possible, but also that it will be desirable. It will be possible because of the arguments given above, namely that if scientific explanations tell us something about the world, then incompatibility cannot be sustained, and that connection to the real world—no matter how tenuous it may be at a given point in time—leads to points of contact which open the door to comparison and integration.

But pluralists such as Longino would additionally argue that it is a mistake to *want* to integrate or make commensurable the different explanations: science's ineliminable plurality should be seen as a success, and the desire to eliminate incompatibilities and integrate explanations is a monist relic (Kellert et al., 2006b; Longino, 2013, p. 137). The point is often made that while integration may be able to give us more detailed explanations about a given phenomenon, that level of detail is not always useful (Longino, 2013, p.147; see also Mitchell, 2009, p.116). And this is clearly true: one does not, in every case, want the most detailed explanation possible; quite often, our pragmatic aims are such that it is sufficient to have an approximate explanation of the phenomenon. For instance, if I want an explanation of why a certain dog is aggressive, I may be content with knowing that it is hungry, without needing to go into all the physiological, neurological, developmental, evolutionary, etc. details which could be relevant to such an explanation.

However, scientific realism means that we will not be satisfied with approximate, albeit useful explanations. The objective of scientific explanation is to explain how the world actually is, which will imply the integration of multiple, partial explanations. What this implies is that integration will always be a desideratum of scientific research. But importantly, it is not the *only* desideratum: oftentimes the highly detailed, rich, integrated explanation will be too complex for our needs, and approximate explanations will be better suited.

The fact that integration will always be desirable goes hand in hand with the drive to resolve incompatibilities. Incompatibilities imply that there is a problem with one, some, or all of the rival explanations, since all the explanations are attempting to explain the same phenomenon, the same world. An explanation which is able to resolve the incompatibilities, and integrate within it all the explanations from the various approaches, will be a more complete explanation of the phenomenon at stake. The pluralist's claim that integration is not desirable is incompatible with scientific realism, because realism implies that richer, more detailed descriptions of the world are always desirable (though again: one among many desiderata).

I am, of course, unable to give a detailed solution to problems of integration—whether theoretical or explanatory—for every possible case. Not only would this be long and tedious, it may actually be infinitely long and therefore impossible. Integration does not rely on a single method, nor a single strategy; as evoked in chapter 5, integration is carried out using the integrative toolkit, a collection of integrative strategies and techniques which expands every time researchers find new ways to integrate explanations, and is dependent on the context. We cannot tell a priori what those tools will be, nor trace any kind of limit to how and where those tools could be applied. In other words, the current state of science is such that it does not contain all the tools needed for every possible integration; some of these tools exist, some are yet to be developed. But there is no principled reason to believe that any approach or explanation will be impervious to integration, since every explanation has empirical and ontological commitments, which inevitably tie it in to other explanations about the same phenomenon.

My thesis also does not need to be committed to the reality of all the entities posited by current science, nor even to any subset of these entities. It is only necessary that 'good' scientific explanations are good insofar as they conform as much as possible to reality as it is. In this sense, explanations which posit entities which do not exist will—someday at least—be routed out as bad explanations. And good explanations will be amenable to integration because they will be referring to the world, which is the anchor around which all explanations must gravitate. Because the world is made in only one way (even though we can represent it in many ways), this means that good explanations will capture the world as it actually is—in the (one) way that it actually is—opening the door to commensurability of multiple explanations which are anchored around the same phenomenon.

In sum, though the differential parsing of the causal space and the measures of conformation muddy the waters for commensurability and integration, it never truly renders them opaque. When incompatibility occurs, researchers can look to their own and their rivals' parsing of the causal space and degrees of conformation, and look to the empirical and ontological commitments in order to clear the waters. Finding the points of contact between the explanations allows different researchers to find common threads which can be explored to create more and more detailed and precise integrative explanations about the phenomena of interest.

7.4 The limits to explanation

The fact that integration will always in principle be possible does not, however, mean that there are no limits to what is possible to explain or integrate in practice. Science is not done in a vacuum and is restricted by many practical considerations.

These constraints pose an important practical limit to the kinds of explanations that it will be possible to produce, as well as limiting the possibilities for integration. For if certain explanations are forever out of reach, then that implies that integrations which would benefit from such explanations are also forever out of reach. But importantly, these limitations are not about what is in-principle explainable, but only about what is in practice explainable. And practical limits do not entail in-principle limits.

7.4.1 Ethical

The first of the limits to explanations are ethical limits concerning the experiments which may be carried out. Indeed, many experiments require the manipulation of environmental and/or genetic factors in order to tease apart the various causes of a given behaviour. And many of these manipulations would be completely unethical if applied to humans.

Matthews and Turkheimer (2021) for instance remark that the experiments needed to create mechanistic explanations of heritability in humans are simply impossible for ethical reasons. While gene-knockout experiments, manipulations of the environment, or any such heavy-handed interventions are possible for non-human model organisms, they are (fortunately) unthinkable for humans. As Matthews and Turkheimer point out, “There will not be a day in which we might randomly assign genetically identical humans to different environments, simply for the purposes of disentangling genetic and environmental influences.” (2021, p.2308) And the same goes for any sort of genetic manipulation for the purposes of ascertaining the mechanisms underlying human heritability of traits. And while these restrictions are by now fortunately obviously true for human experimental subjects, these sorts of ethical concerns are also more and more present for experiments on other animals.

So, while it is in principle possible to devise experiments which would tease apart various contributing causes to a behaviour, in practice many of these experiments should not and will not be carried out.

7.4.2 Access

Another important limitation to explanations is related to access to the relevant data or environment. Certain phenomena will seemingly forever be impossible to study directly.

This is the case with events or phenomena which happened in the past, or phenomena which are so far as to be out of reach. Evolutionary explanations, for instance, call on events which are typically far removed in the past, and the observations of the actual mechanisms leading to natural selection are forever lost to time. Though we can glean information indirectly from many sources, the explanations will always have a certain amount of approximation or uncertainty because of the lack of access to the actual phenomena. Similar problems arise when considering explanations about events or phenomena on other planets or other galaxies.

Another limitation for explanations associated to access relates to the resources necessary for investigations. Certain experiments are prohibitively expensive, be that in time, money, or human resources, and as such seem impossible to carry out in practice.

Many other constraints no doubt limit our access, including size, geography, technology, etc. But these remain practical limitations, and not in-principle limits to what can be known.

7.4.3 No end to science

Another important practical limitation when it comes to explanations and integrations is simply the explanations that we have not yet produced. As Ruphy (2013, p.161) remarks, our explanations are answers to the questions that we ask about a given phenomenon, and there is no end in sight to the variety of questions one can ask about a given phenomenon. This potentially infinite number of questions gives us good reason to think that there will always be explanations that we have not yet produced. In this respect there is no hope of a 'complete' explanation for a given phenomenon, far less for a 'complete' science, or even discipline. We will be constantly seeking answers to our new questions, meaning that there is no end to science. As Ruphy points out, despite the fact that pluralists like to paint monist as seekers of a 'complete', unified science, this cannot be the objective for even a monist view of science, since it is an impossibility, perhaps even incoherent.

7.5 Tempered integrative monism

The above arguments are intended to show how a certain form of monism is realistic and desirable. Before going on to a more detailed description of the tempered integrative monism I propose, I will first revisit the cartography analogy so often used by pluralists, to show how it does not support pluralism in the way they believe it does.

7.5.1 The cartography analogy revisited

As described in chapter 5, Longino and other pluralists often use the cartography analogy to illustrate their understanding of the plurality of scientific explanations (Giere, 1999; Kitcher, 2001; Longino, 2002; Mitchell, 2009). In brief, every map is created keeping in mind the intended content and certain conventions. These two building blocks together imply that a given map will always be a partial representation of the terrain, since only certain elements are included, and are represented in ways which could be idiosyncratic. In much the same way, scientific explanations will only ever look at certain elements of a phenomenon, and measure and conceptualise them in their own specific ways (see section 5.4.3 for more details).

Importantly, the analogy is meant to help us understand that increased accuracy and detail is not always desirable, and that one can never have a complete map of the terrain, or a complete explanation of a phenomenon, since they will always leave certain aspects out, and as such, be partial. The justification for this position is that our pragmatic aims will be the guide for the level of detail, and our pragmatic aims do not align with having the most detail possible. Longino claims that “the map with the best fit is not the one with the greatest possible resolution. Because that would duplicate the terrain being mapped, it would be useless. A map must be a partial representation; otherwise it fails to be a map.” (Longino, 2002, p. 116) And, it is claimed, the same goes for explanations: our pragmatic aims will be such that we will not want too much detail; we will always want one which approximates reality and gives us quicker or simpler answers to our questions. Furthermore, we will not want explanations which account for every aspect of a phenomenon, because that would be unwieldy; instead, we should want multiple, partial explanations. Hence the permanency of pluralism: the multiple explanations, just as the multiple maps, are all useful in their own way, and will not be integratable, because they have their own intended content and conventions (in the case of maps), or parsing of the causal space and conformation (in the case of scientific explanations).

The first thing to note is that the idea that a map, or explanation, with too much detail is undesirable full stop is simply wrong. While it is true that pragmatic aims are such that partial maps and explanations can be best in certain cases, it is not the case that a full representation is useless. The claim that a one-to-one mapping of a territory is useless (Kitcher, 2001, p. 60; Longino, 2002, p. 116; Mitchell, 2009, p. 116) is mistaking the map for the territory. After all, one must not forget that it is a *representation*, and not the territory itself; as such, it can have uses that the territory does not, such as representing the territory elsewhere than where it actually is, giving a baseline on which to add information, or highlighting specific aspects of the territory which are most important for our pragmatic aims. Additional detail can clearly be useful in some cases (though—as mentioned above—admittedly not in all cases); one need only think of representations of microscopic entities. In these cases the scale is even greater than one to one! Surely that does not imply that it is even more useless than the one-to-one correspondence. Imagine for instance gaining a fully detailed account of how a certain (up until now mysterious) behaviour is brought about; clearly everyone would agree that this is far better than no account at all, and in some respects better than a partial account. Every level of detail could, in principle, be useful. As Sober (1999) has pointed out regarding other matters, it’s not because it’s not your favorite explanation that it isn’t an explanation at all. In other words, the usefulness of an explanation for any given task is not equivalent to whether or not it is an explanation.

A second point which subverts the pluralists’ cartography analogy are the advances in map-making technology. The development of computer-based geographical information systems (GIS) has completely changed what can be understood from the analogy. First, a map which allows a zooming in to a one-to-one scale is clearly desirable: and the more detail there is, the better it is. This is because not all the data needs to be visible at the same time: through the use of different layers which can be made visible or not, the user can choose what information is best for their current purposes. And second, GIS technology in fact now supports an understanding of scientific explanations as integratable. Indeed, modern mapping software is able to layer different maps, different intentional content, one over the other. To do so, it is necessary to specify the points at which each map 'connects' to the other layers. As long as two or more maps represent the same region, there will always be points which overlap in the maps, since—insofar as they are both maps—they represent the real world and it is therefore possible to have them connect at specific points which relate to the world they represent. For instance, if one wanted to add a map of the Montreal metro onto a topographical map of Montreal, one would need to specify the points on each map where the intentional content overlaps: in this case, the geographical location of each station. The metro map thus gets distorted to fit over the geographical map, and both intentional contents are present in a single representation. Of course, this leads to distortion in one or both maps, perhaps entailing a greater difficulty in parsing the information, but that information nevertheless is present, and in fact mapping software allows for all types of manipulations to make the information more readily understood by the user.

In much the same way, scientific explanations represent the real world, and 'connect' to it through what I have called 'points of contact'. Just as with maps, as long as scientific explanations concern the same phenomenon, and have points of contact, there will always be possible ways to compare, contrast and ultimately tie together differing explanations of a given phenomenon. In other words, there will always be some amount of commensurability for explanations that overlap in the phenomena they purport to explain. And while the explanations, just as maps, may need to be modified or distorted to be made to fit the integration, it is in the name of increased fidelity to the world, and in that respect the integration remains desirable (though again: not the only desideratum for explanations).

The cartography analogy thus subverted shows how increasing the detail of scientific explanations is in fact desirable. And that increased detail will come about among others through the integration of explanations, making integration desirable as well. And that integration will always be possible, because just as maps connect at certain points when they are about the same territory, so do scientific explanations when they are about the same phenomenon.

7.5.2 Unification of science through integration

We now have a complete picture of the monism I propose. It is a monism which relies on integration as the way to make connections between the various explanations created through scientific research. As I have argued, there are no in-principle limits to integration, since explanations tell us something about the world as it actually is, and in that respect, explanations can never be truly incommensurable, nor incompatible one with the other. The ideal end result of this monism is an interconnected web of scientific explanations, related to one another through integration. As described in chapter 5, integration can span multiple levels, as well as connecting explanations at the same level. It can also relate proximal and ultimate explanations, and link together the explanations which answer any and all of Tinbergen's four questions. Because integration is applicable in all directions, it breaks down any possible isolation of explanations from one another, unifying science through innumerable piecemeal and local integrations of explanations.

Thus, this monism does not imply that there will be a unification of explanations under a sparse, elegant, and simple formula or explanatory framework. To assume such an end point is to put the cart before the horse, since we currently have no such formula or explanation. In fact, though that idea may have some traction in physics due to previous unifications, it is not at all clear what that would even mean in terms of the biology of behaviour, or even biology in general. Is a formula which explains all of behaviour in biological terms even conceivable? The sheer complexity and diversity of phenomena and possible causal pathways leading to them seems to preclude its very possibility.

In this respect I tentatively agree with Waters (2017), who proposes that the ontology of the world is such that there is no general structure to the world (at least when it comes to biology), which would explain why no unifying framework has been possible to date. He puts forward a cartography analogy of his own to illustrate his point: the ontology of the world does not look like the neat, structured grid layout of the streets in Calgary, Canada, but instead looks like the messy and haphazard jumble of small, crooked streets of Arles, France. Of course, one might point out that future research could reveal that behind the apparent messiness hide straight lines which appear only through explanations that we have not yet discovered. But regardless of this possibility, the interesting thing to point out for my position is that even if the world is like the streets of Arles, it remains that it is possible to create a map of Arles, to understand which roads connect to which others and in which ways. When it comes to biological explanations of behaviour, this implies that though we may be unable to unify the explanations in some single, elegant way, it remains that the piecemeal explanations of diverse, sometimes overlapping phenomena, could eventually be joined through the integration of explanations, and shown to all be part of a grand web of related explanations, no matter how messy it ends up looking.

Positions analogous to mine have been defended by others, though all seem to posit certain limits to the unity, or at the very least shy away from the monist label. For instance, Potochnik (2011) draws on Neurath's (1937) and to some extent Darden and Maull's (1977) work to defend a view of scientific unity she calls "coordinate unity". Her approach puts forth the "coordination of diverse fields of science, none of which is taken to have privileged status." (2011, p.305) However in her later book (2017), Potochnik seems to backpedal somewhat, accepting certain forms of pluralism. Others will defend certain forms of integration, all the while denying that it could lead to any kind of unity. For instance, the emphasis on the reciprocal relations between fields often associated with integration is also sometimes used to bolster anti-reductionism, such as Grantham's (2004) conceptualization of the unity of science as "interconnection", Craver's mosaic unity (2007), or Faucher's (2012) non-reductionist integrationist account. All these examples either explicitly label themselves as pluralists, or impose certain limits to the unification, either by limiting their perspective to (for instance) interfield theories, or by outright denying the possibility of certain integrations.

In contrast, my position proposes that there is no in-principle limit to integration, for the reasons outlined above. But just as the above proposals emphasize, I also see the relations between approaches, fields, and explanations as reciprocal: there is no a priori hierarchy in explanations. Integration is predicated on the idea that multiple explanations from various approaches can be coordinated to yield a more complete picture of the phenomenon at hand. Whether one explanation is more important than another will be determined on a case-by-case basis, and as such, no approach is fundamentally superior to any other.

Of course, as mentioned, the monism is tempered by practical limitations to the explanations that can be produced, whether that be through ethical limitations or simply due to lack of access to the appropriate phenomenon. Science is furthermore never finished, since there will always be new questions that can be asked about any given phenomenon. This means that in practice, we will never arrive at the ideal end result of a complete interconnected web of explanations; but we can nevertheless work towards it, and there are no in-principle barriers to its advancement.

I believe this integrative monism reflects more accurately than pluralism how science in fact proceeds. Incompatibilities are not acceptable in scientific practice. When they arise, researchers will do their best to break down any apparent incommensurabilities in order to resolve the apparent contradictions. Rather than let the proliferation of incompatible explanations continue untethered, my monism suggests that commensurability will be sought out. And that commensurability can lead to integration because the multiple explanations will be explaining the same phenomenon. Integration will draw connections between explanations, and as such is desirable in order to better understand phenomena under investigation. Integrative strategies are constantly being developed, adding to the integrative toolkit. These are the tools used to resolve apparent incommensurability, and the toolkit is constantly expanding, in new and original ways, suggesting that integration will always be possible.

CONCLUSION

This research began with an interest in the diversity of explanations put forth in the domain of biological explanations of behaviour, and ended up tackling broad questions in philosophy of science about the nature and capacities of scientific explanations. I have proposed that explanatory pluralism—a position defended by many contemporary philosophers of science—is a fundamentally misguided idea. Explanatory pluralists try to reconcile scientific realism with the possibility of having multiple independent, irreconcilable, and perhaps even incompatible, scientific explanations about a given phenomenon. I argue that that project fails, and that scientific realism implies that explanations will always be reconcilable one with the other, through the proliferation of integrative strategies. Because scientific explanations tell us something about how the world actually is, it will always, in principle, be possible to find points of contact between incompatible explanations, which will open the door to the use of tools from the integrative toolkit, or the creation of new tools for the toolkit. Indeed, since scientific integration is contextual and depends on the questions asked and epistemic units involved, new integrative strategies are constantly being developed, adding tools to our capacity to build bridges between explanations, each one chipping away at the purported incommensurabilities necessary to defend certain forms of explanatory pluralism.

As an alternative to explanatory pluralism, I have defended tempered integrative monism. This is a modest form of pluralism, which proposes that scientific research in the biology of behaviour will be unified not through some grand theory or explanation of everything, but through the links between the multiple, local explanations of phenomena, which will joined one with the other through integration. We are left with a web of interconnected explanations, joined together by a multiplicity of integrative strategies, with no explanation being in principle independent from, or incompatible with, any other. The only limits to integration will be the practical limits regarding experiments we should never do due to ethical constraints, those we will never be able to do due to limits in our access to certain phenomena, and those we have not yet done, since there is no apparent end to scientific research.

To defend this monism, I first proposed a new typology for explanatory pluralisms which clarifies what it is about scientific explanations which can be understood to be plural. While most everyone agrees that there can exist many *types* of scientific explanations, others defend stronger forms of explanatory pluralisms. Fragmentation pluralists hold that while certain explanations can be joined together through integration, there will ultimately be barriers to integration which will preclude any unification to scientific research. Insular pluralism is the strongest form of explanatory pluralism, which proposes not only that certain integrations will be impossible, but also that it will be possible to have incompatible explanations of a given phenomenon. And each of these incompatible explanations will nevertheless be considered 'correct' by the researchers putting them forth, with no way of reconciling them due to some form of incommensurability. This typology will allow anyone interested in explanatory pluralism to more clearly grasp what is at stake, and understand the premises and consequences of the various explanatory pluralisms defended in the literature.

I then took a critical look at Sandra Mitchell's (2002, 2009) defense of fragmentation pluralism, which she names "integrative pluralism". She proposes that while certain forms of integration will be possible, there are limits to the possibilities of integration, meaning that there is no possible broad unification to scientific explanations of the biology of behaviour. Her vertical approach to pluralism is to propose that certain explanatory reductions will be impossible, due to the existence of emergent phenomena which are understood as defying any cross-level integration. Yet as I demonstrated, her arguments do not hold up to scrutiny: the reductionism she criticizes contains within it the tools necessary to remedy the problems she contends to have found. And her redefinition of "scientific emergence" is such that the term is deflated of all ontological and even epistemological relevance in the context of cross-level explanations.

Her horizontal approach to fragmentation pluralism rests on the idea that the idealizations involved in scientific models imply that there will always be a multiplicity of possible models to explain any given general phenomenon. I showed, however, that once the identification of a "general phenomenon" has been done, and once we are clear on the explanatory role of models, then the purported pluralism boils down to a list of possible causes for a phenomenon, which is a far cry from the promises made by explanatory pluralism as a novel way of understanding the interactions between scientific explanations. Mitchell's main arguments in defense of fragmentation pluralism have thus been shown to inadequately support her thesis, leaving us to wonder what, if any, limits there could be to integration.

Helen Longino (2002, 2013), for her part, defends a form of insular pluralism, which proposes that incompatible explanations for a given phenomenon can co-exist, and all be considered successful despite the incompatibility. This incompatibility will furthermore remain because of the incommensurability implied in the unique ways in which various approaches parse the causal space of the phenomenon, as well as evaluate the success of their explanations. Against Longino's pluralism, I argued that there will always be ways to integrate explanations, and that incommensurability will not hold.

I surveyed the ways in which integration has been understood to function, showing the various strategies and approaches researchers have used to bring together explanations. Integration is the process of combining epistemic units from various scientific perspectives for the creation of novel epistemic units. The resulting integrations lead to explanations which tend to be more complete and more detailed; though this is not the only desideratum possible for explanations, it certainly is a significant one. Each new integrative strategy is a tool added to the integrative toolkit, giving more and more options to researchers to break down incommensurabilities, and resolving incompatibilities.

To showcase the possibility and value of creating new integrative tools, I proposed a novel conceptualization of the term 'behaviour' as used in biological research. Not only is this a central concept evoked in biological explanations of behaviour, it is also central to this thesis. My proposal is to understand it not as a categorical concept with clear boundaries, but instead as a spectrum concept, graded along three axes: the complexity of the mechanisms involved, the stability of their constitutive components, and the quantity and significance of difference-making inputs. This new understanding of the term should facilitate communication between researchers discussing behaviour in any and all entities, allowing them to focus on the significant characteristics underlying the phenomena, rather than disagreements regarding classifications of phenomena as behaviour or not.

Integration is thus understood as the remedy for incommensurability and incompatibility, for the reasons explained above. This led me to defend tempered integrative monism, which is foremost a descriptive position, arguing that this kind of monism better reflects actual scientific practice, and what it is that scientific explanations about the biology of behaviour can and cannot do. Of note is that biological explanations of behaviour are often understood to be the clearest example of explanatory pluralism due to the complexity of the phenomena, and the number of approaches which can be called on to produce explanations. As such, if this monism holds for biological explanations of behaviour, then there is good reason to believe that it will be applicable to other areas of inquiry which are taken to be less complex, and less diverse in the approaches, though this remains to be demonstrated. This monism is also a prescriptive position, proposing that scientific research will progress better and faster if it is understood that we ought to accept the multiplicity of approaches to research, all the while making efforts to reconcile this plurality into a grand web of integrated explanations, giving us a more complete understanding of the phenomena at stake.

We should embrace the monism made possible by scientific integration, which accepts the diversity of methods and approaches to scientific research, all the while promoting the unity of science through an understanding that all successful scientific explanations can contribute to one another, and can contribute to a greater understanding of ourselves and the world which surrounds us.

REFERENCES

Abney, D. H., Dale, R., Yoshimi, J., Kello, C. T., Tylén, K., & Fusaroli, R. (2014). Joint perceptual decision-making: A case study in explanatory pluralism. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00330

Aizawa, K., & Gillett, C. (2019). Defending pluralism about compositional explanations. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 101202. https://doi.org/10.1016/j.shpsc.2019.101202

Alcock, J. (1987). Ardent adaptationist. *Natural History*, *96*(4).

Alexander, S. (1920). *Space, Time, and Deity*. MacMillan.

Allchin, D. (1991). *Resolving Disagreement in Science: The ox-phos controversy, 1961-1977*. University of Chicago.

Allchin, D. (1994). The Super Bowl and the Ox-Phos Controversy: 'Winner-Take-All' Competition in Philosophy of Science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1994*, 22–33.

Allen, C. (2018, May 11). *A Place for Intentional Explanation?* [Talk]. The Biology of Behaviour: Explanatory Pluralism Across the Life Sciences, Université du Québec à Montréal. https://biobehaviour.wordpress.com/

Allen, C., & Bekoff, M. (1997). *Species of mind: The philosophy and biology of cognitive ethology*. MIT Press.

Allen, R. D. (1969). Mechanism of the seismonastic reaction in mimosa pudica. *Plant Physiology*, *44*(8), 1101–1107. https://doi.org/10.1104/pp.44.8.1101

Andersen, H. K. (2017). Reduction in the biomedical sciences. In M. Solomon, J. R. Simon, & H. Kincaid (Eds.), *The Routledge Companion to Philosophy of Medicine* (pp. 81–89). Routledge.

Ayala, F. J. (1974). Introduction. In F. J. Ayala & T. Dobzhansky (Eds.), *Studies in the Philosophy of Biology: Reduction and Related Problems* (pp. vii–xvi). Macmillian Press.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., & Zdravkovic, V. (2008). Empirical investigation of starling flocks: A benchmark study in collective animal behaviour. *Animal Behaviour*, *76*(1), 201–215. https://doi.org/10.1016/j.anbehav.2008.02.004

Baumrind, D. (1993). The Average Expectable Environment Is Not Good Enough: A Response to Scarr. *Child Development*, *64*(5), 1299–1317. https://doi.org/10.2307/1131536

Beatty, J. (1993). The evolutionary contingency thesis. In G. Wolters & J. G. Lennox (Eds.), *Concepts, Theories and Rationality in the Biological Sciences* (pp. 45–81). University of Pittsburgh Press.

Beatty, J. (1997). Why Do Biologists Argue like They Do? *Philosophy of Science*, *64*, S432–S443. JSTOR.

Bechtel, W. (1986). *Integrating scientific disciplines*. M. Nijhoff ; [distributor] for the United States and Canada, Kluwer Academic Publishers.

Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, *66*(2), 175–207.

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT Press.

Beck, W. S., Liem, K. F., & Simpson, G. G. (1991). *Life, an Introduction to Biology* (3rd ed.). Harper Collins.

Behaviour. (2017). In *OED Online*. Oxford University Press.

Belter, P. R., & Cahill, J. F. (2015). Disentangling root system responses to neighbours: Identification of novel root behavioural strategies. *AoB Plants*, *7*, 1–12. https://doi.org/10.1093/aobpla/plv059

Ben-Jacob, E., Schochet, O., Tannenbaum, A., Cohen, I., Czirok, A., & Vicsek, T. (1994). Generic modelling of cooperative growth patterns in bacterial colonies. *Nature*, *368*, 46–49.

Beshers, S. N., & Fewell, J. H. (2001). Models of division of labor in social insects. *Annual Review of Entomology*, *46*, 413–440. https://doi.org/10.1146/annurev.ento.46.1.413

Bickle, J. (1998). *Psychoneural Reduction: The New Wave*. MIT Press.

Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Springer.

Bliss, T. V. P., Gardner-Medwin, A. R., & Lomo, T. (1973). Synaptic plasticity in the hyppocampal formation. In G. B. Ansell & P. B. Bradley (Eds.), *Macromolecules and Behavior* (pp. 193–203). Macmillan.

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, *97*(3), 303–352. https://doi.org/10.2307/2185445

Boomsma, J. J., Fjerdingstad, E. J., & Frydenberg, J. (1999). Multiple paternity, relatedness and genetic diversity in Acromyrmex leaf–cutter ants. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1416), 249–254. https://doi.org/10.1098/rspb.1999.0629

Borges, J. L. (1975). *A universal history of infamy.* (N. T. Di Giovanni, Trans.). Penguin.

Bouchard, T. J. (2004). Genetic influence on human psychological traits a survey. *Current Directions in Psychological Science*, *13*(4), 148–151.

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of Human Psychological Differences: The Minnesota Study of Twins Reared Apart. *Science*, *250*(4978), 223–228.

Braillard, P.-A., & Malaterre, C. (2015). Explanation in Biology: An Introduction. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in Biology* (Vol. 11, pp. 1–28). Springer Netherlands. https://doi.org/10.1007/978-94-017-9822-8

Brigandt, I. (2010). Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. *Erkenntnis*, *73*(3), 295–311.

Brigandt, I. (2013). Integration in biology: Philosophical perspectives on the dynamics of interdisciplinarity. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *44*(4), 461–465. https://doi.org/10.1016/j.shpsc.2013.09.009

Brigandt, I., & Love, A. (2017). Reductionism in biology. In *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition). Edward N. Zalta. https://plato.stanford.edu/archives/spr2017/entries/reduction-biology/

Bull, J. J., & Wang, I.-N. (2010). REVIEW: Optimality models in the age of experimental evolution and genomics: Experimental evolution and genomics. *Journal of Evolutionary Biology*, *23*(9), 1820–1838. https://doi.org/10.1111/j.1420-9101.2010.02054.x

Cahill, J. F. (2019). The inevitability of plant behavior. *American Journal of Botany*, *106*(7), 903–905. https://doi.org/10.1002/ajb2.1313

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2001). *Self-organization in biological systems*. Princeton University Press.

Campaner, R. (2014). Explanatory pluralism in psychiatry: What are we pluralists about, and why? In M. C. Galavotti, D. Dieks, W. J. Gonzalez, S. Hartmann, T. Uebel, & M. Weber (Eds.), *New Directions in the Philosophy of Science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-04382-1

Campbell, D. T. (1974). 'Downward Causation' in Hierarchically Organised Biological Systems. In F. J. Ayala & T. Dobzhansky (Eds.), *Studies in the Philosophy of Biology* (pp. 179–186). Macmillan Education UK. https://doi.org/10.1007/978-1-349-01892-5_11

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering* (First edition). Oxford University Press.

Cartwright, N. (1983). *How the laws of physics lie*. Clarendon Press ; Oxford University Press.

Cartwright, N. (1999). *The Dappled World: A Study of The Boundaries of Science*. Cambridge University Press.

Cartwright, N. (2009). Evidence-based policy: What's to be done about relevance?: For the 2008 Oberlin Philosophy Colloquium. *Philosophical Studies*, *143*(1), 127–136. https://doi.org/10.1007/s11098-008-9311-4

Chakravartty, A. (2017). Scientific Realism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/

Churchland, P. M. (1981a). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, *78*(2), 67. https://doi.org/10.2307/2025900

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT Press.

Churchland, P. S., & Sejnowski, T. J. (2000). Perspectives on neuroscience. In M. S. Gazzaniga (Ed.), *Cognitive neuroscience: A reader* (pp. 14–24). Blackwell.

Class, B., Kluen, E., & Brommer, J. E. (2014). Evolutionary quantitative genetics of behavioral responses to handling in a wild passerine. *Ecology and Evolution*, *4*(4), 427–440. https://doi.org/10.1002/ece3.945

Collins, A. L., & Sullivan, P. F. (2013). Genome-wide association studies in psychiatry: What have we learned? *The British Journal of Psychiatry*, *202*(1), 1–4. https://doi.org/10.1192/bjp.bp.112.117002

Couzin, I. D., & Krause, J. (2003). Self-Organization and Collective Behavior in Vertebrates. *Advances in the Study of Behavior*, *32*, 1–75.

Crasnow, S. (2013). Feminist philosophy of science: Values and objectivity. *Philosophy Compass*, *8*(4), 413–423.

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376. https://doi.org/10.1007/s11229-006-9097-x

Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Clarendon Press.

Craver, C. F. (2014). The Ontic Account of Scientific Explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the Special Sciences* (Vol. 367, pp. 27–52). Springer Netherlands. https://doi.org/10.1007/978-94-007-7563-3_2

Culp, S., & Kitcher, P. (1989). Theory structure and theory change in contemporary molecular biology. *The British Journal for the Philosophy of Science*, *40*(4), 459–483. https://doi.org/10.1093/bjps/40.4.459

Cusimano, S., & Sterner, B. (2019). Integrative pluralism for biological function. *Biology & Philosophy*, *34*(6), 55. https://doi.org/10.1007/s10539-019-9717-8

Cvrčková, F., Žárský, V., & Markoš, A. (2016). Plant studies may lead us to rethink the concept of behavior. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00622

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science*, *44*(1), 43–64. https://doi.org/10.1086/288723

Darwin, C. (1859). *On the Origin of Species* (1st ed.). John Murray.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. John Murray. https://doi.org/10.5962/bhl.title.2092

Davis, D. E. (1966). *Integral Animal Behavior*. Macmillan.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

Delehanty, M. (2005). Emergent properties and the context objection to reduction. *Biology & Philosophy*, *20*(4), 715–734. https://doi.org/10.1007/s10539-004-2437-7

Deneubourg, J., Goss, J., Pasteels, J., Fresneau, D., & Lachaud, J. (1987). Self-Organization Mechanisms in Ant Societies (II): Learning in Foraging and Division of Labor. *Experientia Supplementum*, *54*(Behavior in Social Insects), 177–196.

Dennett, D. C. (1998). *The Intentional Stance* (7th printing). MIT Press.

Devitt, M. (1997). *Realism and truth* (2nd ed. with a new afterword). Princeton University Press.

Douglas, H. E. (2007). Rejecting the ideal of value-free science. In H. Kincaid, A. Wylie, & J. Dupré (Eds.), *Value-free science? Ideals and illusions* (pp. 120–141). Oxford University Press.

Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press.

Dowell, J. L. (2006). Formulating the thesis of physicalism: An introduction. *Philosophical Studies*, *131*(1), 1–23. https://doi.org/10.1007/s11098-006-6641-y

Duarte, A., Weissing, F. J., Pen, I., & Keller, L. (2011). An Evolutionary Perspective on Self-Organized Division of Labor in Social Insects. *Annual Review of Ecology, Evolution, and Systematics*, *42*(1), 91–110. https://doi.org/10.1146/annurev-ecolsys-102710-145017

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press.

Dupré, J. (2002). The Lure of the Simplistic. *Philosophy of Science*, *69*(S3), S284–S293. https://doi.org/10.1086/341852

Dupré, J. (2007). Fact and value. In H. Kincaid, J. Dupré, & A. Wylie (Eds.), *Value-free science? Ideals and illusions* (pp. 27–41). Oxford University Press.

Dussutour, A., Latty, T., Beekman, M., & Simpson, S. J. (2010). Amoeboid organism solves complex nutritional challenges. *Proceedings of the National Academy of Sciences*, *107*(10), 4607–4611. https://doi.org/10.1073/pnas.0912198107

Eklund, M. (2015). Intuitions, conceptual engineering, and conceptual fixed points. In C. Daly (Ed.), *The Palgrave handbook of philosophical methods* (pp. 363–385). http://public.eblib.com/choice/publicfullrecord.aspx?p=4329156

Elgin, M., & Sober, E. (2002). Cartwright on Explanation and Idealization. *Erkenntnis*, *57*(3), 441–450. https://doi.org/10.1023/A:1021502932490

Elliott, K. C., & Steel, D. (Eds.). (2017). *Current controversies in values and science*. Routledge, Taylor & Francis Group.

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., Brückner, M. K., Schreiweis, C., Winter, C., Sohr, R., Becker, L., Wiebe, V., Nickel, B., Giger, T., Müller, U., Groszer, M., Adler, T., Aguilar, A., Bolle, I., … Pääbo, S. (2009). A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. *Cell*, *137*(5), 961–971. https://doi.org/10.1016/j.cell.2009.03.041

Enc, B. (1983). In Defense of the Identity Theory. *The Journal of Philosophy*, *80*(5), 279. https://doi.org/10.2307/2026499

Ereshefsky, M. (2001). *The Poverty of The Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge University Press. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=72915

Ereshefsky, M. (2007). Psychological categories as homologies: Lessons from ethology. *Biology & Philosophy*, *22*(5), 659–674. https://doi.org/10.1007/s10539-007-9091-9

Eysenck, H. J. (1994). Meta-analysis and its problems. *BMJ (Clinical Research Ed.)*, *309*(6957), 789–792. https://doi.org/10.1136/bmj.309.6957.789

Faucher, L. (2012). Unity of science and pluralism: Cognitive neurosciences of racial prejudice as a case study. In O. Pombo, J. M. Torres, J. Symons, & S. Rahman (Eds.), *Special Sciences and the Unity of Science*. Springer Netherlands. https://doi.org/10.1007/978-94-007-2030-5

Faucher, L. (2014). Non-Reductive Integration in Social Cognitive Neuroscience: Multiple Systems Model and Situated Concepts. In *Brain Theory* (pp. 217–240). Springer.

Faucher, L., & Poirier, P. (2001). Psychologie évolutionniste et théories interdomaines. *Dialogue*, *40*(3), 453–486. https://doi.org/10.1017/S0012217300018874

Fawcett, T. W., Hamblin, S., & Giraldeau, L.-A. (2013). Exposing the behavioral gambit: The evolution of learning and decision rules. *Behavioral Ecology*, *24*(1), 2–11. https://doi.org/10.1093/beheco/ars085

Feyerabend, P. (1975). *Against method: Outline of an anarchistic theory of knowledge*. NLB ; Humanities Press.

Feyerabend, P. (1981). *Philosophical papers*. Cambridge University Press.

Feyerabend, P. K. (1962). *Explanation, reduction, and empiricism*. http://conservancy.umn.edu/handle/11299/184633

Fodor, J. A. (1974). Special sciences: Or the disunity of science as a working hypothesis. *Synthese*, *28*, 97–115.

Fodor, J. A. (1998). Look! Review of Conscilience: The Unity or Knowledge by E. O. Wilson. *London Review of Books*, *20*(21). https://www.lrb.co.uk/the-paper/v20/n21/jerry-fodor/look

Frigg, R., & Hartmann, S. (2020). Models in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/models-science/

Galison, P., & Stump, D. J. (Eds.). (1996). *The Disunity of science: Boundaries, contexts, and power*. Stanford University Press.

Galton, F. (1883). *Inquiries into Human Faculty and its Development.* MacMillan Co. https://doi.org/10.1037/14178-000

Giere, R. N. (1999). *Science Without Laws*. University of Chicago Press.

Giere, R. N. (2006). Perspectival pluralism. In S. H. Kellert, H. E. Longino, & C. K. Waters (Eds.), *Scientific Pluralism: Vol. XIX* (pp. 26–41). University of Minnesota Press.

Gijsbers, V. (2016). Explanatory pluralism and the (dis)unity of science: The argument from incompatible counterfactual consequences. *Frontiers in Psychiatry*, *7*. https://doi.org/10.3389/fpsyt.2016.00032

Giraldeau, L.-A., & Dubois, F. (2008). Social Foraging and the Study of Exploitative Behavior. In *Advances in the Study of Behavior* (Vol. 38, pp. 59–104). Elsevier. https://doi.org/10.1016/S0065-3454(08)00002-8

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, *69*(S3), S342–S353. https://doi.org/10.1086/341857

Godfrey-Smith, P. (2009). Abstractions, idealizations, and evolutionary biology. In A. Barberousse, M. Morange, & T. Pradeu (Eds.), *Mapping the Future of Biology: Evolving Concepts and Theories* (pp. 47–55). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9636-5

Godfrey-Smith, P. (2013). Darwinian individuals. *From Groups to Individuals: Evolution and Emerging Individuality*, 17–36.

Golani, I. (1992). A mobility gradient in the organization of movement: The perception of movement through symbolic language. *Behavioral and Brain Sciences*, *15*, 249–308.

Goodenough, A. E., Little, N., Carpenter, W. S., & Hart, A. G. (2017). Birds of a feather flock together: Insights into starling murmuration behaviour revealed using citizen science. *PLOS ONE*, *12*(6), e0179277. https://doi.org/10.1371/journal.pone.0179277

Gottlieb, G. (1995). Some Conceptual Deficiencies in 'Developmental' Behavior Genetics. *Human Development*, *38*(3), 131–141. https://doi.org/10.1159/000278306

Gould, S. J. (1980). *The panda's thumb: More reflections in natural history* (1st ed). Norton.

Gould, S. J. (1987). Freudian slip. *Natural History*, *96*(2), 14–21.

Grafen, A. (1984). Natural selection, kin selection and group selection. In J. Krebs & N. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (2nd ed., pp. 62–84). Blackwell Scientific Press.

Grantham, T. A. (2004). Conceptualizing the (Dis)unity of Science. *Philosophy of Science*, *71*(2), 133–155. https://doi.org/10.1086/383008

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart, Lung and Vessels*, *5*(4), 219–225.

Gregory, A. M., Ball, H. A., & Button, T. M. M. (2011). Behavioral Genetics. In P. K. Smith & C. H. Hart (Eds.), *The Wiley-Blackwell handbook of childhood social development* (2nd ed, pp. 27–44). Wiley-Blackwell.

Grier, J. W., & Burk, T. (1992). *Biology of Animal Behavior* (2nd ed.). Times Mirror/Mosby College.

Griffiths, P. E. (2002). What is innateness? *The Monist*, *85*(1), 70–85.

Griffiths, P. E. (2009). In What Sense Does ‘Nothing Make Sense Except in the Light of Evolution’? *Acta Biotheoretica*, *57*(1–2), 11–32. https://doi.org/10.1007/s10441-008-9054-9

Griffiths, P. E., & Gray, R. D. (1994). Developmental Systems and Evolutionary Explanation. *The Journal of Philosophy*, *91*(6), 277–304. https://doi.org/10.2307/2940982

Griffiths, P. E., & Stotz, K. (2006). Genes in the Postgenomic Era. *Theoretical Medicine and Bioethics*, *27*(6), 499–521. https://doi.org/10.1007/s11017-006-9020-y

Guala, F. (2003). Experimental Localism and External Validity. *Philosophy of Science*, *70*(5), 1195–1205. https://doi.org/10.1086/377400

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.

Hacking, I. (1985). Do we see through a microscope? In B. C. Van Fraassen, P. M. Churchland, & C. A. Hooker (Eds.), *Images of science: Essays on realism and empiricism, with a reply from Bas C. van Fraassen* (pp. 132–152). University of Chicago Press.

Hadfield, J. D., Nutall, A., Osorio, D., & Owens, I. P. F. (2007). Testing the phenotypic gambit: Phenotypic, genetic and environmental correlations of colour. *Journal of Evolutionary Biology*, *20*(2), 549–557. https://doi.org/10.1111/j.1420-9101.2006.01262.x

Hammerstein, P. (1996). Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology*, *34*, 511–532.

Hansell, M. H. (2007). *Built by Animals: The Natural History of Animal Architecture*. Oxford University Press.

Hardcastle, V. G. (1992). Reduction, explanatory extension, and the mind/brain sciences. *Philosophy of Science*, *59*(3), 408–428.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. Free Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, *15*(2), 135–175.

Herschel, J. (1830). *Preliminary Discourse on the Study of Natural Philosophy*. Thoemmes Press.

Hildenbrandt, H., Carere, C., & Hemelrijk, C. K. (2010). Self-organized aerial displays of thousands of starlings: A model. *Behavioral Ecology*, *21*(6), 1349–1359. https://doi.org/10.1093/beheco/arq149

Hinde, R. A. (1970). *Animal Behavior: A Synthesis of Ethology and Comparative Psychology* (2nd ed.). McGraw Hill.

Hochstein, E. (2017). Why one model is never enough: A defense of explanatory holism. *Biology & Philosophy*, *32*(6), 1105–1125.

Hochstein, E. (2019). How metaphysical commitments shape the study of psychological mechanisms. *Theory & Psychology*, *29*(5), 579–600. https://doi.org/10.1177/0959354319860591

Hochstein, E. (2023). Integration without integrated models or theories. *Synthese*, *202*(3), 76. https://doi.org/10.1007/s11229-023-04298-w

Hogan, B. G., Hildenbrandt, H., Scott-Samuel, N. E., Cuthill, I. C., & Hemelrijk, C. K. (2017). The confusion effect when attacking simulated three-dimensional starling flocks. *Royal Society Open Science*, *4*(1), 160564. https://doi.org/10.1098/rsos.160564

Horgan, J. (1997). *The end of science: Facing the limits of knowledge in the twilight of the scientific age* (1st Broadway Books trade pbk. ed). Broadway Books.

Horst, S. W. (2016). *Cognitive pluralism*. MIT Press.

Huneman, P. (2014). A Pluralist Framework to Address Challenges to the Modern Synthesis in Evolutionary Theory. *Biological Theory*, *9*(2), 163–177. https://doi.org/10.1007/s13752-014-0174-y

Issad, T., & Malaterre, C. (2015). Are dynamic mechanistic explanations still mechanistic? In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in Biology. An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences* (pp. 265–292). Springer.

James, W. (1890). *The Principles of Psychology*. Henry Holt Company.

Jones, M. R. (2005). Idealization and abstraction: A framework. In M. R. Jones & N. Cartwright (Eds.), *Correcting the model: Idealization and abstraction in the sciences* (pp. 173–218). Rodopi.

Kacelnik, A. (1984). Central Place Foraging in Starlings (Sturnus vulgaris). I. Patch Residence Time. *The Journal of Animal Ecology*, *53*(1), 283. https://doi.org/10.2307/4357

Kacelnik, A. (2012). Putting Mechanisms into Behavioral Ecology. In P. Hammerstein & J. R. Stevens (Eds.), *Evolution and the Mechanisms of Decision Making* (Vol. 11, pp. 21–38). MIT Press.

Kaiser, M. I. (2015). *Reductive Explanation in the Biological Sciences*. Springer International Publishing. https://doi.org/10.1007/978-3-319-25310-7

Karban, R. (2008). Plant behaviour and communication. *Ecology Letters*, *11*(7), 727–739. https://doi.org/10.1111/j.1461-0248.2008.01183.x

Kauffman, S. A. (1971). Articulation of parts explanations in biology and the rational search for them. In Philosophy of Science Association, R. C. Buck, & R. Carnap (Eds.), *PSA 1970: In memory of Rudolf Carnap ; proceedings of the 1970 biennial meeting, Philosophy of Science Association* (pp. 257–272). Reidel.

Kellert, S. H. (2008). *Borrowed knowledge: Chaos theory and the challenge of learning across disciplines*. University of Chicago Press.

Kellert, S. H., Longino, H. E., & Waters, C. K. (Eds.). (2006a). *Scientific Pluralism*. University of Minnesota Press.

Kellert, S. H., Longino, H. E., & Waters, C. K. (2006b). The pluralist stance. In S. H. Kellert, H. E. Longino, & C. K. Waters (Eds.), *Scientific Pluralism* (pp. vii–xxix). University of Minnesota Press.

Kendler, K. S. (2008). Explanatory Models for Psychiatric Illness. *The American Journal of Psychiatry*, *165*(6), 695–702. https://doi.org/10.1176/appi.ajp.2008.07071061

Kendler, K. S. (2012). The dappled nature of causes of psychiatric illness: Replacing the organic–functional/hardware–software dichotomy with empirically based pluralism. *Molecular Psychiatry*, *17*(4), 377–388. https://doi.org/10.1038/mp.2011.182

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2002). Toward a comprehensive developmental model for major depression in women. *The American Journal of Psychiatry*, *159*(7), 1133–1145. https://doi.org/10.1176/appi.ajp.159.7.1133

Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *The American Journal of Psychiatry*, *163*(1), 109–114. https://doi.org/10.1176/appi.ajp.163.1.109

Kendler, K. S., & Myers, J. (2010). The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychological Medicine*, *40*(5), 801–806. https://doi.org/10.1017/S0033291709991140

Kim, J. (1984). Concepts of supervenience. *Philosophy and Phenomenological Research*, *45*(2), 153–176.

Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, *63*(3), 31. https://doi.org/10.2307/3130081

Kim, J. (1992). 'Downward causation' in emergentism and nonreductive physicalism. In A. Beckerman, H. Flohr, & J. Kim (Eds.), *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism* (pp. 119–138). De Gruyter.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *95*(1/2), 3–36.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, *48*(4), 507–531.

Kitcher, P. (1984). Species. *Philosophy of Science*, *51*(2), 308–333.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation: Vol. XIII* (pp. 410–505). University of Minnesota Press.

Kitcher, P. (1990). The Division of Cognitive Labor. *The Journal of Philosophy*, *87*(1), 5. https://doi.org/10.2307/2026796

Kitcher, P. (2001). *Science, Truth, and Democracy*. Oxford University Press.

Kornblith, H. (2002). *Knowledge and its Place in Nature*. Clarendon Press ; Oxford University Press.

Krebs, J. R., & Davies, N. B. (1993). *An Introduction to Behavioural Ecology* (3rd ed.). Blackwell Scientific Publications.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed). University of Chicago Press.

Kutschenko, L. K. (2011). How to Make Sense of Broadly Applied Medical Classification Systems: Introducing Epistemic Hubs. *History and Philosophy of the Life Sciences*, *33*(4), 583–601.

Laland, K. N., & Brown, G. R. (2011). *Sense and Nonsense: Evolutionary Perspectives on Human Behaviour* (2nd ed.). Oxford University Press.

Laland, K. N., Odling-Smee, J., Hoppitt, W., & Uller, T. (2013). More on how and why: Cause and effect in biology revisited. *Biology & Philosophy*, *28*(5), 719–745. https://doi.org/10.1007/s10539-012-9335-1

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and Effect in Biology Revisited: Is Mayr's Proximate-Ultimate Dichotomy Still Useful? *Science*, *334*(6062), 1512–1516. https://doi.org/10.1126/science.1210879

Laland, K. N., Wray, G. A., & Hoekstra, H. E. (2014). Does evolutionary theory need a rethink? *Nature*, *514*(7521), 161.

Laubichler, M. D., Prohaska, S. J., & Stadler, P. F. (2018). Toward a mechanistic explanation of phenotypic evolution: The need for a theory of theory integration. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *330*(1), 5–14. https://doi.org/10.1002/jez.b.22785

Lazzeri, F. (2014). A conceptual difficulty with some definitions of behavior. In J. Conte & C. A. Mortari (Eds.), *Temas em Filosofia Contemporânea* (Vol. 13, pp. 148–155). NEL/UFSC.

Leibold, M. A., Chase, J. M., Shurin, J. B., & Downing, A. L. (1997). Species Turnover and the Regulation of Trophic Structure. *Annual Review of Ecology and Systematics*, *28*(1), 467–494. https://doi.org/10.1146/annurev.ecolsys.28.1.467

Leonelli, S. (2013). Integrating data to acquire new knowledge: Three modes of integration in plant science. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *44*(4), 503–514. https://doi.org/10.1016/j.shpsc.2013.03.020

Levitis, D. A., Lidicker, W. Z., & Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. *Animal Behaviour*, *78*(1), 103–110. https://doi.org/10.1016/j.anbehav.2009.03.018

Lewontin, R. C. (1974). Annotation: The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*(3), 400.

Lloyd, E. A. (1989). A Structural Approach to Defining Units of Selection. *Philosophy of Science*, *56*(3), 395–418. https://doi.org/10.1086/289498

Lloyd, E. A. (2005). Why the Gene Will Not Return. *Philosophy of Science*, *72*(2), 287–310. https://doi.org/10.1086/432425

Lloyd, E. A., Dunn, M., Cianciollo, J., & Mannouris, C. (2005). Pluralism without genic causes? *Philosophy of Science*, *72*(2), 334–341. https://doi.org/10.1086/432451

Longino, H. E. (1990). *Science as Social Knowledge*. Princeton University Press.

Longino, H. E. (2001). What do we measure when we measure aggression? *Studies in History and Philosophy of Science Part A*, *32*(4), 685–704.

Longino, H. E. (2002). *The Fate of Knowledge*. Princeton University Press.

Longino, H. E. (2013). *Studying Human Behaviour: How Scientists Investigate Aggression and Sexuality*. University of Chicago Press.

Love, A. C. (2003). Evolutionary Morphology, Innovation, and the Synthesis of Evolutionary and Developmental Biology. *Biology & Philosophy*, *18*(2), 309–345. https://doi.org/10.1023/A:1023940220348

Love, A. C. (2006). Evolutionary morphology and Evo-devo: Hierarchy and novelty. *Theory in Biosciences*, *124*(3), 317–333. https://doi.org/10.1016/j.thbio.2005.11.006

Love, A. C. (2008). Explaining Evolutionary Innovations and Novelties: Criteria of Explanatory Adequacy and Epistemological Prerequisites. *Philosophy of Science*, *75*(5), 874–886. https://doi.org/10.1086/594531

Love, A. C., & Lugar, G. L. (2013). Dimensions of integration in interdisciplinary explanations of the origin of evolutionary novelty. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *44*(4), 537–550. https://doi.org/10.1016/j.shpsc.2013.09.008

Loyau, A., Petrie, M., Saint Jalme, M., & Sorci, G. (2008). Do peahens not prefer peacocks with more elaborate trains? *Animal Behaviour*, *76*(5), e5–e9. https://doi.org/10.1016/j.anbehav.2008.07.021

Lucas, C., & Ben-Shahar, Y. (2021). The foraging gene as a modulator of division of labour in social insects. *Journal of Neurogenetics*, *35*(3), 168–178. https://doi.org/10.1080/01677063.2021.1940173

Maccoby, E. (2000). Parenting and its Effects on Children: On Reading and Misreading Behavior Genetics. *Annual Review of Psychology*, *51*, 1–27. https://doi.org/10.1146/annurev.psych.51.1.1

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25.

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.

Malaterre, C. (2011). Making sense of downward causation in manipulationism: Illustrations from cancer research. *History and Philosophy of the Life Sciences*, *33*(4), 537–561.

Massimi, M. (2022). *Perspectival Realism* (1st ed.). Oxford University PressNew York. https://doi.org/10.1093/oso/9780197555620.001.0001

Matthews, L. J., & Turkheimer, E. (2021). Across the great divide: Pluralism and the hunt for missing heritability. *Synthese*, *198*(3), 2297–2311. https://doi.org/10.1007/s11229-019-02205-w

Maye, A., Hsieh, C., & Sugihara, G. (2007). Order in Spontaneous Behavior. *PLoS ONE*, *5*, 14.

Mayr, E. (1961). Cause and Effect in Biology. *Science*, *134*(3489), 1501–1506. https://doi.org/10.1126/science.134.3489.1501

McAllister, J. W. (1997). Phenomena and Patterns in Data Sets. *Erkenntnis (1975-)*, *47*(2), 217–228.

McGue, M. (1994). Why developmental psychology should find room for behavioral genetics. In C. N. Alexander (Ed.), *Threats To Optimal Development: Integrating Biological, Psychological, and Social Risk Factors: The Minnesota Symposia on Child Psychology, Volume 27* (1st ed.). Lawrence Erlbaum Associates, Inc. https://doi.org/10.4324/9780203773666

Metelli, S., & Chaimani, A. (2020). Challenges in meta-analyses with observational studies. *Evidence Based Mental Health*, *23*(2), 83–87. https://doi.org/10.1136/ebmental-2019-300129

Mill, J. S. (1843). *A System of Logic*. Longmans, Green and Co.

Millikan, R. G. (1993). *White Queen psychology and other essays for Alice*. MIT Press.

Mitchell, S. D. (1992). On Pluralism and Competition in Evolutionary Explanations. *American Zoologist*, *32*(1), 135–144.

Mitchell, S. D. (2002). Integrative pluralism. *Biology and Philosophy*, *17*(1), 55–70.

Mitchell, S. D. (2003). *Biological Complexity and Integrative Pluralism*. Cambridge University Press.

Mitchell, S. D. (2009). *Unsimple Truths: Science, Complexity, and Policy*. University of Chicago Press.

Mitchell, S. D., & Dietrich, M. R. (2006). Integration without Unification: An Argument for Pluralism in the Biological Sciences. *The American Naturalist*, *168*(S6), S73–S79. https://doi.org/10.1086/509050

Miyashita, S., Ishibashi, K., Kishino, H., & Ishikawa, M. (2015). Viruses roll the dice: The stochastic behavior of viral genome molecules accelerates viral adaptation at the cell and tissue levels. *PLOS Biology*, *13*(3), e1002094. https://doi.org/10.1371/journal.pbio.1002094

Morange, M. (2015). Is there an explanation for...the diversity of explanations in biological studies? In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in Biology* (Vol. 11, pp. 31–46). Springer Netherlands. https://doi.org/10.1007/978-94-017-9822-8

Morgan, C. L. (1923). *Emergent Evolution*. Williams and Norgate.

Morley, E. L., & Robert, D. (2018). Electric fields elicit ballooning in spiders. *Current Biology*, *28*(14), 2324-2330.e2. https://doi.org/10.1016/j.cub.2018.05.057

Morrison, M. (2011). One phenomenon, many models: Inconsistency and complementarity. *Studies in History and Philosophy of Science Part A*, *42*(2), 342–351. https://doi.org/10.1016/j.shpsa.2010.11.042

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Explanation*. Harcourt, Brace & World, Inc.

Nagel, T. (1998). Reductionism and antireductionism. *Novartis Foundation Symposium*, *213*, 3–10; discussion 10-14, 73–75.

Neurath, O. (1937). Unified science and its encyclopedia. In R. S. Cohen & M. Neurath (Eds.), *Philosophical Papers 1913-1946* (Vol. 16). D. Reidel.

Oberheim, E., & Hoyningen-Huene, P. (2018). The Incommensurability of Scientific Theories. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/incommensurability/

O'Connor, T. G. (2014). Developmental Behavioral Genetics. In M. Lewis & K. D. Rudolph (Eds.), *Handbook of Developmental Psychopathology* (pp. 245–263). Springer US. http://link.springer.com/10.1007/978-1-4614-9608-3_13

Okasha, S. (2006). *Evolution and the levels of selection*. Clarendon Press ; Oxford University Press.

O'Malley, M. A., & Soyer, O. S. (2012). The roles of integration in molecular systems biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 58–68. https://doi.org/10.1016/j.shpsc.2011.10.006

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, Theories, and The Mind-body Problem* (pp. 3–36). University of Minnesota Press.

O'Rourke, M., Crowley, S., & Gonnerman, C. (2016). On the nature of cross-disciplinary integration: A philosophical framework. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 56*, 62–70. https://doi.org/10.1016/j.shpsc.2015.10.003

Oyama, S. (2000a). *Evolution's Eye*. Duke University Press.

Oyama, S. (2000b). *The Ontogeny of Information: Developmental Systems and Evolution* (2nd ed.). Duke University Press.

Page, R. E., & Mitchell, S. D. (1998). Self-organization and the evolution of division of labor. *Apidologie*, *29*(1–2), 171–190. https://doi.org/10.1051/apido:19980110

Parmigiani, S., Palanza, P., Rodgers, J., & Ferrari, P. F. (1999). Selection, evolution of behavior and animal models in behavioral neuroscience. *Neuroscience & Biobehavioral Reviews*, *23*(7), 957–970.

Parrott, D. J., & Giancola, P. R. (2007). Addressing "The criterion problem" in the assessment of aggressive behavior: Development of a new taxonomic system. *Aggression and Violent Behavior*, *12*(3), 280–299. https://doi.org/10.1016/j.avb.2006.08.002

Petrie, M. (1994). Improved growth and survival of offspring of peacocks with more elaborate trains. *Nature*, *371*(6498), Article 6498. https://doi.org/10.1038/371598a0

Petrie, M., TR, H., & Sanders, C. (1991). Peahens prefer peacocks with elaborate trains. *Animal Behaviour*, *41*. https://doi.org/10.1016/S0003-3472(05)80484-1

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. Viking.

Plaisance, K. S., & Reydon, T. A. C. (Eds.). (2012). *Philosophy of Behavioral Biology* (Vol. 282). Springer Netherlands. http://link.springer.com/10.1007/978-94-007-1951-4

Plutynski, A. (2004). Review of Biological Complexity and Integrative Pluralism: Sandra Mitchell (2003). *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/news/biological-complexity-and-integrative-pluralism/

Plutynski, A. (2016). Review of Explanatory Pluralism in the Life Sciences: Pierre-Alain Braillard and Christophe Malaterre (eds) (2015). *Science & Education*, *25*(5–6), 681–689. https://doi.org/10.1007/s11191-016-9843-5

Poirier, P. (2000). L'empire contre-attaque: Le retour du réductionnisme. *Philosophiques*, *27*.

Poorter, H., Anten, N. P. R., & Marcelis, L. F. M. (2013). Physiological mechanisms in plant growth models: Do we need a supra-cellular systems biology approach? *Plant, Cell & Environment*, *36*(9), 1673–1690. https://doi.org/10.1111/pce.12123

Potochnik, A. (2011). A Neurathian Conception of the Unity of Science. *Erkenntnis*, *74*(3), 305–319. https://doi.org/10.1007/s10670-010-9228-0

Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226507194.001.0001

Preuss, T. M. (2012). Human brain evolution: From gene discovery to phenotype discovery. *Proceedings of the National Academy of Sciences*, *109*(Supplement_1), 10709–10716. https://doi.org/10.1073/pnas.1201894109

Putnam, H. (1965). Brains and behavior. *Analytical Philosophy*, *2*.

Putnam, H. (1975). *Mathematics, matter, and method*. Cambridge University Press.

Quignon, F., Kiene, L., Levi, Y., Sardin, M., & Schwartzbrod, L. (1997). Virus behaviour within a distribution system. *Water Science and Technology*, *35*(11), 311–318. https://doi.org/10.1016/S0273-1223(97)00278-3

Raven, P. H., & Johnson, G. G. (1989). *Biology* (2nd ed.). Times Mirror/Mosby College.

Réale, D., & Festa-Bianchet, M. (2000). Quantitative genetics of life-history traits in a long-lived wild mammal. *Heredity*, *85*(6), 593. https://doi.org/10.1046/j.1365-2540.2000.00795.x

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology*, *19*(1), 43–62. https://doi.org/10.1080/1350178X.2012.661069

Reiss, J. (2015). *Causation, Evidence, and Inference*. Routledge. https://doi.org/10.4324/9781315771601

Reiss, J. (2019). Against external validity. *Synthese*, *196*(8), 3103–3121. https://doi.org/10.1007/s11229-018-1796-6

Repko, A. F. (2012). *Interdisciplinary research: Process and theory* (2nd ed). SAGE Publications.

Reydon, T. A. C. (2012). How-possibly explanations as genuine explanations and helpful heuristics: A comment on Forber. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 302–310. https://doi.org/10.1016/j.shpsc.2011.10.015

Richardson, A. (2008). Solomon's Science without Conscience, or, On the Coherence of Epistemic Newtonianism. *Perspectives on Science*, *16*(3), 246–252.

Rittschof, C. C., & Robinson, G. E. (2014). Genomics: Moving behavioural ecology beyond the phenotypic gambit. *Animal Behaviour*, *92*, 263–270. https://doi.org/10.1016/j.anbehav.2014.02.028

Ronald, E. M. A., Sipper, M., & Capcarrère, M. S. (1999). Design, observation, surprise! A test of emergence. *Artificial Life*, *5*, 225–239.

Rose, S. (2001). The Poverty of Reductionism. In R. S. Singh, C. B. Krimbas, D. B. Paul, & J. Beatty (Eds.), *Thinking about Evolution: Historical, Philosophical and Political Perspectives* (pp. 415–428).

Rosenberg, A. (2006). *Darwinian reductionism, or, How to stop worrying and love molecular biology*. University of Chicago Press.

Rossi, S., Deslauriers, A., Anfodillo, T., Morin, H., Saracino, A., Motta, R., & Borghetti, M. (2006). Conifers in cold environments synchronize maximum growth rate of tree-ring formation with day length. *New Phytologist*, *170*(2), 301–310. https://doi.org/10.1111/j.1469-8137.2006.01660.x

Royer, C. (2002). *Protein-protein interactions*. Biophysical Society. http://www.biophysics.org/Portals/1/PDFs/Education/croyer.pdf

Ruphy, S. (2013). *Pluralismes scientifiques: Enjeux épistémiques et métaphysiques*. Hermann.

Rutting, L., Post, G., Keestra, M., de Roo, M., Blad, S., & de Greef, L. (2020). *Interdisciplinary integration* (S. Menken & M. Keestra, Eds.). Amsterdam University Press.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Sankey, H. (1994). *The incommensurability thesis*. Avebury ; Ashgate Pub. Co.

Sarkar, S. (1998). *Genetics and Reductionism*. Cambridge University Press.

Scarr, S. (1999). Behavior-genetic and socialization theories of intelligence: Truce and reconciliation. In R. J. Sternberg (Ed.), *Intelligence, heredity, and environment* (Repr, pp. 3–41). Cambridge Univ. Press.

Schaffner, K. F. (1967). Approaches to Reduction. *Philosophy of Science*, *34*(2), 137–147. https://doi.org/10.1086/288137

Schaffner, K. F. (1969). The Watson-Crick Model and Reductionism. *The British Journal for the Philosophy of Science*, *20*(4), 325–348.

Schaffner, K. F. (1974). Reductionism in biology: Prospects and problems. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 613–632.

Schaffner, K. F. (1993). *Discovery and Explanation in Biology and Medicine*. University of Chicago Press.

Schaffner, K. F. (2012). Ernest Nagel and Reduction. *Journal of Philosophy*, *109*(8), 534–565. https://doi.org/10.5840/jphil20121098/926

Schaffner, K. F. (2016). *Behaving: What's genetic, what's not, and why should we care?* (First edition). Oxford University Press.

Seeley, T. D. (1989). Social foraging in honeybees: How nectar foragers assess their colony's nutritional status. *Behavioral Ecology and Sociobiology*, *24*(3), 181–199. https://doi.org/10.1007/BF00292101

Sherman, P. W. (1988). The levels of analysis. *Animal Behavior*, *36*, 616–619.

Sih, A. (2004). A Behavioral Ecological View of Phenotypic Plasticity. In T. J. DeWitt & S. M. Scheiner (Eds.), *Phenotypic Plasticity: Functional and Conceptual Approaches* (p. 14). Oxford University Press.

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*(7), 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Silvertown, J. (1998). Plant phenotypic plasticity and non-cognitive behaviour. *Trends in Ecology & Evolution*, *13*(7), 255–256.

Sober, E. (1990). The Poverty of Pluralism: A Reply to Sterelny and Kitcher. *The Journal of Philosophy*, *87*(3), 151. https://doi.org/10.2307/2026633

Sober, E. (1999). The multiple realizability argument against reductionism. *Philosophy of Science*, *66*(4), 542–564.

Sober, E. (2001). Separating Nature And Nurture.pdf. In D. Wasserman & R. Wachbroit (Eds.), *Genetics and Criminal Behavior* (pp. 47–78). Cambridge University Press.

Sober, E., & Wilson, D. S. (2003). *Unto Others: The Evolution and Psychology of Unselfish Behavior* (4. print). Harvard Univ. Press.

Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford University Press.

Sterelny, K., & Kitcher, P. (1988). The Return of the Gene. *The Journal of Philosophy*, *85*(7), 339. https://doi.org/10.2307/2026953

Stich, S. P. (1981). Dennett on intentional systems. *Philosophical Topics*, *12*(1), 39–62.

Stotz, K. (2006). Molecular Epigenesis: Distributed Specificity as a Break in the Central Dogma. *History and Philosophy of the Life Sciences*, *28*(4), 533–548.

Strevens, M. (2011). *Depth: An account of scientific explanation* (1st Harvard University Press paperback ed., 2011). Harvard Univ. Press.

Sullivan, J. A. (2009). The multiplicity of experimental protocols: A challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese*, *167*(3), 511–539. https://doi.org/10.1007/s11229-008-9389-4

Sullivan, J. A. (2017). Coordinated pluralism as a means to facilitate integrative taxonomies of cognition. *Philosophical Explorations*, *20*(2), 129–145. https://doi.org/10.1080/13869795.2017.1312497

Suris, A., Lind, L., Emmett, G., Borman, P. D., Kashner, M., & Barratt, E. S. (2004). Measures of aggressive behavior: Overview of clinical and research instruments. *Aggression and Violent Behavior*, *9*(2), 165–227. https://doi.org/10.1016/S1359-1789(03)00012-0

Tabery, J. (2014). *Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture*. MIT Press.

Tabery, J., & Griffiths, P. E. (2010). Historical and philosophical perspectives on behavioral genetics and developmental science. In K. E. Hood (Ed.), *Handbook of Developmental Science, Behavior, and Genetics* (pp. 41–60). Wiley-Blackwell.

Takahashi, M., Arita, H., Hiraiwa-Hasegawa, M., & Hasegawa, T. (2008). Peahens do not prefer peacocks with more elaborate trains. *Animal Behaviour*, *75*(4), 1209–1219. https://doi.org/10.1016/j.anbehav.2007.10.004

Thagard, P. (1998). Explaining Disease: Correlations, Causes, and Mechanisms. *Minds and Machines*, *8*, 61–78.

Tinbergen, N. (1963). On aims and methods of ethology. *Animal Biology*, *55*(4), 297–321.

Tinbergen, N. (1976). *The Study of Instinct* (3rd ed.). Oxford University Press.

Tofts, C., & Franks, N. R. (1992). Doing the right thing: Ants, honeybees and naked mole-rats. *Trends in Ecology & Evolution*, *7*(10), 346–349. https://doi.org/10.1016/0169-5347(92)90128-X

Trewavas, A. (2009). What is plant behaviour? *Plant, Cell & Environment*, *32*(6), 606–616. https://doi.org/10.1111/j.1365-3040.2009.01929.x

Turkheimer, E. (2015). Genetic Prediction. *Hastings Center Report*, *45*(S1), S32–S38. https://doi.org/10.1002/hast.496

Van Bouwel, J. (2014). Pluralists About Pluralism? Different Versions of Explanatory Pluralism in Psychiatry. In M. C. Galavotti, D. Dieks, W. J. Gonzalez, S. Hartmann, T. Uebel, & M. Weber (Eds.), *New Directions in the Philosophy of Science* (pp. 105–120). Springer International Publishing. https://doi.org/10.1007/978-3-319-04382-1

Van Fraassen, B. C. (1980). *The scientific image*. Clarendon Press ; Oxford University Press.

Van Fraassen, B. C., Churchland, P. M., & Hooker, C. A. (Eds.). (1985). *Images of science: Essays on realism and empiricism, with a reply from Bas C. van Fraassen*. University of Chicago Press.

van Riel, R. (2011). Nagelian reduction beyond the Nagel model. *Philosophy of Science*, *78*(3), 353–375. https://doi.org/10.1086/660300

Vandenbrink, J. P., Brown, E. A., Harmer, S. L., & Blackman, B. K. (2014). Turning heads: The biology of solar tracking in sunflower. *Plant Science*, *224*, 20–26. https://doi.org/10.1016/j.plantsci.2014.04.006

Verreault-Julien, P. (2019). How could models possibly provide how-possibly explanations? *Studies in History and Philosophy of Science Part A*, *73*, 22–33. https://doi.org/10.1016/j.shpsa.2018.06.008

Wallace, R., Sanders, G. P., & Ferl, R. J. (1991). *Biology: The Science of Life* (3rd ed.). Harper Collins.

Waters, C. K. (1990). Why the Anti-Reductionist Consensus Won't Survive: The Case of Classical Mendelian Genetics. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1990*(1), 125–139. https://doi.org/10.1086/psaprocbienmeetp.1990.1.192698

Waters, C. K. (2005). Why genic and multilevel selection theories are here to stay. *Philosophy of Science*, *72*(2), 311–333. https://doi.org/10.1086/432426

Waters, C. K. (2007). Causes that make a difference. *The Journal of Philosophy*, *104*(11), 551–579.

Waters, C. K. (2017). No general structure. In M. Slater & Z. Yudell (Eds.), *Metaphysics and the Philosophy of Science* (pp. 81–108). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199363209.003.0005

Weber, M. (2002). Incommensurability and theory comparison in experimental biology. *Biology & Philosophy*, *17*(2), 155–169. https://doi.org/10.1023/A:1015286709491

Weber, M. (2005). *Philosophy of experimental biology*. Cambridge University Press.

Weber, M. (2012). Behavioral traits, the intentional stance, and biological functions: What neuroscience explains. In K. S. Plaisance & T. A. C. Reydon (Eds.), *Philosophy of Behavioral Biology* (pp. 317–327). Springer.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199933662.001.0001

Wereha, T. J., & Racine, T. P. (2012). Evolution, Development, and Human Social Cognition. *Review of Philosophy and Psychology*, *3*(4), 559–579. https://doi.org/10.1007/s13164-012-0115-2

Whewell, W. (1840). *The Philosophy of Inductive Sciences, Founded upon their History*. John W. Parker.

Wilson, E. O. (2000). *Sociobiology: The New Synthesis* (25th anniversary ed). Belknap Press of Harvard University Press.

Wimsatt, W. C. (1976a). Reductionism, Levels of Organization, and the Mind-Body Problem. In G. G. Globus, G. Maxwell, & I. Savodnik (Eds.), *Consciousness and the Brain: A Scientific and Philosophical Inquiry* (pp. 205–267). Springer US. https://doi.org/10.1007/978-1-4684-2196-5_9

Wimsatt, W. C. (1976b). Reductive explanation: A functional account. In R. S. Cohen & A. Michalos (Eds.), *Proceedings of the 1974 meeting of the Philosophy of Science Association* (pp. 671–710). D. Reidel.

Wimsatt, W. C. (2000). Emergence as Non-Aggregativity and the Biases of Reductionisms. *Foundations of Science*, *5*, 269–297.

Wimsatt, W. C. (2006). Reductionism and its heuristics: Making methodological reductionism honest. *Synthese, 151*(3), 445–475. https://doi.org/10.1007/s11229-006-9017-0

Winston, M. L. (1991). *The biology of the honey bee* (1. Harvard Univ. Press paperback ed). Harvard Univ. Press.

Wylie, A. (2003). Why standpoint theory matters. In R. Figueroa & S. G. Harding (Eds.), *Science and other cultures: Issues in philosophies of science and technology*. Routledge.