2019

Let Me Explain: the Dynamical Alternative

Russell Meyer
University of Wollongong

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Let Me Explain: the Dynamical Alternative

Russell Meyer

Supervisor:
Dr. Patrick McGivern

This thesis is presented as part of the requirements for the conferral of the degree:

Doctor of Philosophy

The University of Wollongong
School of Humanities & Social Inquiry

December, 2019

Declaration

I, Russell Meyer, declare that this thesis submitted in partial fulfilment of the requirements for the conferral of the degree Doctor of Philosophy, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Russell Meyer
4th March 2021

Abstract

Mechanistic explanation is often held to be necessary for providing causal explanations within the special sciences. A countervailing push for non-mechanistic explanations, often appealing to dynamical models, has been met with criticism from mechanists, who claim these dynamical explanations are incomplete unless reduced to mechanisms. This mechanist critique incorporates the widespread view that mechanistic explanations are objective explanations, and hence possess exclusive causal explanatory power for the special sciences that trumps dynamicists' efforts. The mechanist–dynamicist debate has subsequently featured prominently in arguments over the desirability of E-approaches to cognition—such as enactivism—versus traditional cognitivism. While traditional cognitivist explanations describe computational mechanisms, E-approaches tend to explain cognitive phenomena by invoking dynamical models. Yet, if mechanists are right, it follows that dynamical explanations of cognition are incomplete, and the explanatory power of the E-approaches is rendered suspect. My purpose in this thesis is to defend dynamical explanations and argue they are not always sensibly improved via reduction to underlying mechanisms. I also cast doubts on attempts to use mechanism to integrate accounts of explanation and cognition. First, I develop an account of dynamical explanation for cognitive science based on an even-handed application of interventionism. Second, I show how dynamical causes are not always reducible to mechanistic explanations. Third, I discuss problems with recent attempts to use mechanistic explanation to integrate theories of cognition. Fourth, I argue, similarly, that attempts to integrate mechanisms into enactive cognitive science have not been successful. Finally, I argue that mechanistic standards of explanation are not objective, derived from nature, and value-free, as some proponents claim.

Preface

The motive for this thesis is "an agitation of mind over seemingly odd questions", in the words of the film director Werner Herzog. A little over four years ago I did not know what philosophy was. If anything, my time as a psychology undergraduate had exposed me to a general suspicion of philosophy. Physicists, chemists, and biologists do not seem to worry about accidentally dissolving their entire discipline by getting too philosophical—sometimes psychology is a little anxious of the prospect.

Fortunately one psychology undergraduate course grappled with—though I did not realise it at the time—the philosophical side of studying cognition: the *History and Meta-theory of Psychology.* We were asked to think seriously and critically about the ingrained assumptions of our field. I wrote, with great passion but little art, an essay on the mutually reinforcing nature of (what I felt were the deeply wrong) assumptions that the mind is modular, computational, and mechanistic.

Re–reading it recently, two things jumped out. First, it wasn't very good. Secondly, and more importantly, I saw the roots of the agitation of mind that resulted in the present thesis.

Encountering Tony Chemero's (2009) book on *Radical Embodied Cognitive Science* some time later (and in retrospect, only understanding about a quarter of it at the time), the agitation grew. Here was a story about cognition I could get behind—one that made sense of our lives as embodied agents.

It may seem strange, then, that this thesis instead focusses chiefly on mechanistic explanations, and less so on cognition. However, this thesis represents my attempts to *get underneath* questions about cognition by asking how our picture of scientific explanation, and our picture of cognition, frame one another. The explanatory questions we ask about cognition—and the answers we consider acceptable—are my targets, and perhaps the real source of this ongoing agitation.

This has consequently been a steep and immense learning experience. I am still climbing it. I hope what I have left behind here will be of interest.

A small note: some of these papers have been submitted to or published in different journals, and so I apologise for any lingering inconsistencies in style or tone that have resulted from this process.

Acknowledgments

When I turned up at my former lecturer Nigel Mackay's office asking whether a Masters degree in "theoretical psychology" existed, essentially walking in off the street as a total stranger, he very generously decided to help me. We eventually figured out the proper place for my interests was philosophy. After a campaign of turning up to talks, bending people's ears, and auditing some philosophy courses, I was eventually allowed to start my PhD in 2016. None of this would have happened without Nigel's kindness. My first thanks go to him.

While I remain a workmanlike philosopher, I could not have been more fortunate in the colleagues and friends I found here at the University of Wollongong. These people have been instrumental in my development both as an academic and a person. I cannot help but think that without such a community, this would have been a grim few years—instead it has been some of my best. Even in moments of calamity, I have loved just about every minute of it—because of all of you.

So my eternal gratitude goes to Patrick McGivern, who rescues drowning postgraduates as a hobby. Patrick is a supervisor and philosopher—and more importantly, person—par excellence. For more than one of us, Patrick saved the day. This thesis would have been impossible without him.

This next set of acknowledgements feel like goodbyes; and a few of them are for now, or will be soon enough. They go to my fellow postgraduates. Our office became a small community of close friends that has enriched my life and been an invaluable source of support. We have been through the trenches together! As waves of disasters—almost comedic in their clockwork regularity and frequency—swept through our small gang of philosophers, you all stood firm.

To Miguel Segundo-Ortin. Wherever Miguel goes, he accumulates friendships with an ease that speaks to his nature. There is no better friend to have. Thank you for making me welcome, and being my partner in crime around campus.

To Alan Jürgens. Thank you for being a good friend. Alan never flinches from his principles, which takes a special kind of strength not many people have. I'm proud to have spent this time with you. Don't let the bastards grind you down.

To Anco Peeters. Behind a thick layer of relentless, Dutch incivility lies someone who (at the risk of offering him a compliment) lives up, in rare fashion, to the ideals of our discipline—he is rigorous, knowledgeable and endlessly charitable. He's not a bad bloke either. I wish you, Lies and Nora all the best in your new home together.

To Cameron Lutman. Thank you for our (perhaps counter–productively lengthy) conversations about philosophy and *Dark Souls*. Let's catch up more often.

To Nick Brancazio, my former nemesis. It is impossible to know Nick and doubt her strength of character or her moral clarity; her excellence elevates everyone. She can be found wherever there is a friend in need, or a fight that needs fighting. I have learned more than I ever bargained for from Nick. To say we have been birds of a feather would immodestly compare me to you—so instead I simply say, in earnest, that I could not have asked for a better best pal.

Thank you also to friends of all stripes who each did their bit to help me through this trial, or just shared the occasional coffee or beer: Jono and Zoeya, John, Liz, Ding, Vern, Ana, Lies, Nat, Bec & Lysh, and anyone else who I've neglected to mention.

How to thank my family? I originally entered psychology, and thereby philosophy, mostly by the wise intervention of my mum Narelle. The night before the cut-off date for university admissions, with me still undecided, she pointed out the relatively new *Bachelor of Psychology* being offered at the University of Wollongong. It seemed like something I'd be interested in, she thought. So if anyone is to blame, it's her.

My parents Jim and Narelle, and my sister Vickie, have supported me with a limitless faith in my strange enthusiasms. It is a great comfort to know that no matter what I decide to do in life, they are convinced I will succeed in it. Each of them has always set an example of a calm moral courage that sets the bar high. We are, I am pleased to say, a stubborn, argumentative and proud group of people: what better preparation for a career in philosophy?

Acknowledgments are a way of giving out thanks. But there are other important kinds of acknowledgement. I acknowledge the traditional custodians of the land on which this university is built. I pay my respects to Elders past, present and emerging, and acknowledge Aboriginal and Torres Strait Islanders as the first people of Australia. They have never ceded sovereignty, and remain strong in their enduring connection to land and culture.

Finally, and with a heavy heart, I give a different kind of acknowledgement to the damage done by the founding of the so–called "Western Civilisation" degree here at the University of Wollongong, funded by the Ramsay Centre. I never expected that this university would serve elitism and Western supremacy—yet it has. It has torn out the guts of our small department, and ruined many friendships.

Ramsay board member Tony Abbott claims that schooling is unacceptably "...pervaded by Asian, indigenous and sustainability perspectives" despite Western traditions making up almost the entirety of humanities curricula across Australia.

A university should not be a private service hired to salve the fragile sensibilities of Ramsay board members, nor to hold up a parasol to shade the old champions of the Empire from criticism. While the Australian National University and the University of Sydney rejected Ramsay, citing academic integrity concerns, and the supremacist overtones of the degree, UOW—to its shame—did not. It will forever be a black mark upon it. To those philosophers who have forgotten self–examination and uncritically thrown in their lot with Ramsay: shame.

Contents

Introduction

The question this thesis grapples with is this: how are accounts of explanation shaped by the needs of theory, and how are the needs of theory informative of our accounts of explanation? I see both questions at the heart of the discussions that will form the bulk of this thesis, and I also see them as mutually informative—yet critically under–discussed.

How we ought to explain phenomena, and what these explanations ought to look like, is a key question in the philosophy of science. Often this discussion involves general concerns over the "justificatory step" (Bokulich 2017) that marks the turn from a mere description into a bona fide explanation. Divisions between broad perspectives on explanation often hinge on where precisely this comes about. For instance, is an explanation an observer-independent accomplishment, rooted in the ontic causal structure of the world, or is it an epistemic achievement? (Salmon 1984)

Contemporary discussion has turned towards scientific practice, and in particular the role of models in science (Potochnik 2015, Bokulich 2017). Many kinds of models are deployed in science, and many seem to grant us insight into nature in one way or the other. Yet how this insight should be characterised, justified, and distinguished from failure, remains difficult to chart in a single, uniform fashion. Scientific successes—description, prediction, explanation, understanding—seem to take many forms, and require just as many stories to make sense of them.

One of the more successful accounts of scientific explanation—and my target in this thesis—has broad, sometimes unificatory ambitions. The mechanistic resurgence, beginning in earnest with Machamer, Darden & Craver's (2000) influential paper *Thinking With Mechanisms*, proposes that explanations ought to reveal the underlying causal processes that produce phenomena. Along with the mechanist touchstones of Bechtel & Richardson (1993) and Glennan (1996), this Mechanism proposes a far-reaching account of scientific explanation. Mechanistic explanation aims to decompose and localise mechanisms into their component parts and activities and show how they interact to produce phenomena. Craver (2007) influentially combines the equally influential interventionist (Woodward 2003) and mechanistic stories together into an overarching framework for mechanistic explanation that aims to be satisfactory both in terms of its relevance to actually–existing science, and broad spectrum application across different domains of science.

I see mechanistic explanation as holding to a few broad claims about its nature, often only indirectly stated but key to the endeavour:

> (i) Mechanism—rooted in interventionism—is the exclusive mode of explanation for its target phenomena.

Identifying strongly with the ontic conception of scientific explanation (Salmon 1984), some strains of mechanism—as in its classic, Cartesian form—claim to explain by revealing the real ontic causal structure that underlies the phenomena we observe in nature (Craver 2007, 2014; Craver & Kaplan 2018). One consequence of this view seems to be a belief in the exclusivity of mechanistic explanatory power within its target domains: since mechanism gets at the causal structure of nature it is authoritative and exhaustive.

This view is expressed as a rejection of non-mechanistic, and non-causal approaches to explanation in the cognitive sciences, on the basis that these sorts of accounts do not reveal causal structure. Some pushback on this front has concerned discussions of non–mechanistic explanations in computational neuroscience (Chirimuuta 2014, 2017) as as dynamical systems theory–based explanations (Issad & Malaterre 2015; Ross 2015).

Chapter 1 of this thesis makes a similar move, but this time on the home turf of mechanism: I show how an even–handed application of interventionism, the foundation of Craver's (2007) mechanistic account, reveals that dynamical models are equally capable of providing causal explanations. This commences a running theme of the thesis—that interventionism does not necessarily act as a guarantee of exclusivity for Mechanism.

Chapter 2 similarly argues for a causal reading of dynamical models, only this time extending its reach to an example from systems biology—cell fates. The cell fate literature reveals how scientists describe and explain the differentiation of animal cells into distinctive stable phenotypes by reference to dynamical models of complex gene regulatory networks (GRNs). I also argue, using Woodward's (2010, 2018) criteria of specificity & proportionality, that the dynamical causal story is here preferable to the possible mechanistic one. Once again, interventionism does not necessarily support a mechanistic reading of phenomena.

> (ii) Mechanism, as theory-neutral, works to unify and integrate other modes of explanation, as well as disparate theories.

At around the same moment a wellspring the mechanistic philosophy began to bubble over, so too did a split in cognitive science over how cognition should be conceived of, and how, therefore (or so I will argue) it should be explained.

While analytic philosophy had come to think of cognition as best understood as the operations of a neurally–realised computer—the brain—manipulating contentful symbols or representations (Fodor 1983; Fodor & Pylyshyn 1988) several discrete lines of investigation in the sciences and philosophy dissent.

Arguments from ecological psychology (Gibson 1979), phenomenology (Merleau-Ponty 1965), and robotics (i.e. Brooks 1991; Beer 1995; Chiel & Beer 1997) among others have argued that basic cognition is better thought of as fundamentally extensive—extended outside the central nervous system—embodied, and rooted in the direct perception–action capacities of organisms rather than indirect computational and symbolic ones.

This extended–embedded–embodied–enactive perspective (sometimes called E–cognition) ranged from cognitivism–friendly over to the more radical embodied and enactive perspectives (Varela et al 1991; Chemero 2009) which mostly or entirely eschew computational explanations. These radicals have found themselves in need of a new account of explanation—after all, the sorts of explanations favoured by mainstream cognitivism (i.e. Marr 1982) do not seem viable if one is committed to a strictly non–computational, non–representational position on cognition.

Dynamical systems theory (DST) is supposed to provide a solution. According to the dynamical hypothesis (Van Gelder 1995, 1998) instead of proposing functional, representational and computational accounts of cognition, E-cognitive science ought to appeal to dynamical models of cognition (Kelso 1995) for a descriptive and explanatory resource. In dynamical models, cognitive systems are characterised by differential and difference equations that describe and predict with great accuracy the dynamical unfolding of a system's behaviour over time.

The natural alliance between DST and embodied cognition is clear: E-cognition wants to explain extensive cognitive phenomena without appealing to internal neural mechanisms, and dynamical models provide accurate descriptions and predictions of phenomena without appealing to mechanisms. The role of dynamical models in this new school of cognitive science is seen as both a necessity, and a source of great insight through these descriptions and predictions (Chemero & Silberstein 2008; Stepp et al 2011; Silberstein & Chemero 2013).

Mechanism re-enters the story here as a sharp critic of this dynamical hypothesis. While mechanists agree on the descriptive and predictive power of dynamical models, they do not see them as standalone explanations. There are several reasons for this (explored in depth in Chapter 1), but the biggest one is that explanations in cognitive science, in the view of mechanists, ought to speak to causes (Craver 2007). Dynamical models do not specify the causal-mechanical workings underneath the dynamics they predict and describe, and therefore do not provide explanations. Only by mapping onto a mechanistic model which does provide this causal detail can dynamical models explain (Kaplan 2011; Kaplan & Craver 2011) providing dynamical mechanistic explanations (Bechtel & Abrahamsen 2010, 2013).

Mechanism is portrayed here as a neutral arbiter of theoretical claims—by testing claims via interventions, these differences of opinion about the boundaries of cognition can be resolved (Kaplan 2012). Nevertheless, many E-cognitive accounts reject Mechanism, possibly in small part because of the existing allegiance between Mechanism and cognitivism (Miłkowski 2013) but also more substantially because in their view cognition is not decomposable, localisable, or explicable in terms of its constituent parts (Chemero & Silberstein 2008). Mechanistic models, therefore, do not provide the sorts of answers radical embodied accounts are looking for in an explanation.

In Chapter 3 I examine Miłkowski et al's (2018) attempts to dissolve both the dynamical–mechanistic debate, and the enactive-cognitivist divide at the same time using mechanism as a neutral arbiter. This account claims that mechanistic explanation is capable of taking over the role of unifier in place of competing theories, integrating E-cognition with traditional cognitive science, while also subsuming dynamical explanations. I argue that Miłkowski et al's (2018) arguments however do not smooth the substantial and empirically significant differences between E- and cognitivist accounts, and hence do not genuinely integrate them. I also suggest their proposal to broad the definition of a mechanism in order to incorporate dynamical explanations does more harm to mechanistic explanation than good. In a similar vein, Chapter 4 assesses two attempts to integrate mechanistic explanation into enactivism, by Abramova & Slors (2019) and Walmsley (2019). I find that as a result of the differing explaantory goals of enactive cogntive science versus

> (iii) Mechanism gains its explanatory bona fides from nature: mechanistic explanations are objective explanations

A continuous theme in these discussions, and in the literature, is that mechanisms are seen to occupy a more metaphysically robust position compared to other kinds of explanantia. Craver (2014) for instance argues that mechanistic explanations are objective explanations; the explanatory standards of mechanism are derived from nature. Mechanistic models explain when they completely, and accurately, describe the ontic causal structure of the world. Consequently mechanistic standards of explanation lie outside time and place—they are independent of the explanatory goals, background assumptions or values of investigators. The explanatory standards of mechanism are guaranteed and closed off via their objective character.

In Chapter 5 I address this mechanistic objectivity, and the assumptions of ontic mechanism. While mechanistic explanation is (in its ontic variant) supposed to be objective because it reveals the real ontic causal structure of nature, I show how there is little argumentative force behind the claim other than the assumption that nature is mechanistic. An appeal to interventionism (Woodward 2003) is also shown to be insufficient to provide objectivity. I argue that the explanatory standards of mechanism are not derived from nature, but instead driven by the *explanatory taste* of mechanist investigators. This taste reflects their goals, values and background assumptions. I make the further point that objectivity is better thought of as the consequence of inter-subjective criticism (Longino 1990).

Chapter 1

The Non–mechanistic Option: Defending Dynamical Explanations

This chapter demonstrates that non–mechanistic, dynamical explanations are a viable approach to explanation in the special sciences. The claim that dynamical models can be explanatory without reference to mechanisms has previously been met with three lines of criticism from mechanists: the causal relevance concern, the genuine laws concern, and the charge of predictivism. I argue, however, that these mechanist criticisms fail to defeat non–mechanistic, dynamical explanation. Using the examples of Haken et al's (1985) model of bimanual coordination, and Thelen et al's (2001) dynamical field model of infant perseverative reaching, I show how each mechanist criticism fails once the standards of Woodward's (2003) interventionist framework are applied to dynamical models. An even–handed application of Woodwardian interventionism reveals that dynamical models are capable of producing genuine explanations without appealing to underlying mechanistic details.

This chapter has been published as Meyer, R. (2020). The non-mechanistic option: Defending dynamical explanations. *British Journal for the Philosophy of Science.*

1.1 Introduction

This chapter demonstrates that mechanist objections to non–mechanistic, dynamical explanations are unsuccessful, as they do not hinder dynamical models from being genuinely explanatory. The justification for this defence comes from the even–handed application of interventionism (Woodward 2003) which will be shown to be a viable framework for both mechanistic and non–mechanistic accounts of explanation.[1] By applying the interventionist framework (specifically the notions of ideal interventions, invariance, and counterfactual explanation) in an even–handed way to dynamical models, it is shown that non–mechanistic, dynamical explanations are not deflated by these critiques. Moreover, Woodwardian interventionism provides the foundations for a fully–fledged account of non–mechanistic, dynamical explanation.

Some critics, mechanist or otherwise, may argue that there is something wrong with the interventionist account offered by Woodward (or simply take a neutral stance towards it) and may be unswayed by my arguments since the whole enterprise of interventionism seems suspect or unconvincing. My goal here is however not to offer a broad defence of Woodward. Interventionism is not necessarily the hill I intend for non–mechanistic explanation to die on. The aim of this paper is rather to make the case that for mechanists who do agree with interventionism (and there are many), and who rely heavily on it for their own accounts (and many do), they must accept that non–mechanistic models like dynamical models can also explain.

The scope and application of mechanistic explanation is a major contemporary topic of discussion in philosophy of science (Boone & Piccinini 2016; Bechtel 2017; Chirimuuta 2017; Matthiessen 2017; Craver 2017). Since the earlier and more restrictive iteration of mechanism laid out by Machamer, Darden & Craver (2000) mechanism has been increasingly extended in scope. Its proponents claim that mechanistic explanations are applicable across a broad range of domains, including systems and cognitive neuroscience (Zednik 2014; Boone & Piccinini 2016), systems biology (Matthiessen 2017), cognitive science (Bechtel & Abrahamsen 2010, 2013) and psychology (Piccinini & Craver 2011). This encroachment on new domains has in turn resulted in disagreements over how appropriate mechanism is to these 'special sciences', with some critics preferring to advance varieties of non–mechanistic explanation (Weiskopf 2011; Dupré 2013; Brigandt et al 2018).

While critics differ in their rationale for preferring a non–mechanistic approach to explanation,[2] there is a common feeling among these dissenters that mechanistic

[1] Related arguments have been presented by Gervais & Weber (2011), Dupré (2013), Silberstein & Chemero (2013) and Woodward (2013). In these cases, the specifics of how interventionism would be applied to dynamical models is left unsaid. Here I have attempted to greatly expand on these sketches and show in detail how interventionism can be integrated into an account of dynamical explanation.

[2] While this chapter focusses on dynamical modelling and dynamical explanation, the term

explanations are not suited to the kinds of nondecomposable, nonlinear or otherwise complex phenomena that are frequently investigated by biologists, cognitive scientists and psychological scientists.

One particularly vocal source of non–mechanistic criticism of mechanism comes from proponents of dynamical modelling, some of whom claim that dynamical models are better suited to some phenomena in cognitive science, psychology and related domains than mechanisms (Chemero & Silberstein 2008; Stepp Chemero & Turvey 2011; Silberstein & Chemero 2013; Lamb & Chemero 2014). It is this dynamicist tendency that I will focus on in this chapter.

Contra the dominant mechanist trend, these dynamicists argue that dynamical models can explain by abstracting away from the mechanistic details of a system and describing highly predictive mathematical models of a systems' behaviour. One influential branch of dynamicism advocates for a revival of covering–law explanation (Walmsley 2008; Gervais & Weber 2011; Stepp et al 2011). This attempt at building a dynamical, covering–law mode of explanation has become widespread enough to be deemed the "received view" (Zednik 2011, pg. 239) amongst dynamicists about how dynamical models could explain. The attempted revival of covering–law explanation has subsequently been criticized heavily by mechanists (Bechtel 2011; Craver & Kaplan 2011; Kaplan & Bechtel 2011; Kaplan & Craver 2011) who consider it to be a flawed approach to explanation, and one that retains the existing flaws in law–based explanation.

This paper is structured as follows: Section 2 outlines mechanistic explanation and Woodwardian explanation, and Section 3 will similarly outline dynamical modelling and the covering–law mode of explanation using the example of the HKB model of bimanual coordination. Sections 4, 5 and 6 then identify and respond to the criticisms made by mechanist, interventionist philosophers and directed towards proponents of dynamical covering–law explanation. Three lines of criticism will be investigated—the genuine laws concern, the causal relevance concern, and the error of predictivism on the part of dynamicists. In Section 7 I address Woodward's criticisms of the HKB model in particular as explanatory within an interventionist framework. Section 8 puts these solutions together into a coherent whole, showing how an account of dynamical explanation would work.[3]

"nonmechanistic explanation" refers to a diverse range of explanatory strategies including topological explanation (Huneman 2010; Sporns 2011; Stepp et al 2011), structural explanation (Hughes 1989a; Bokulich 2011) and functional explanation (Weiskopf 2011).

[3] Concerns about decomposition and localisation strategies may be conspicuous by their absence in this paper. The same dynamicist and mechanist camps are engaged in an ongoing (and closely related) debate about the necessity of these heuristics in explanation. Silberstein & Chemero (2013) for instance claim that some nonlinear and complex phenomena are best explained without low-level mechanistic detail, a view echoed by Woodward (2013). Some mechanists (Bechtel & Abrahamsen 2010, 2013; Craver & Kaplan 2011; Kaplan & Craver 2011; Kaplan 2015) by contrast argue that decomposition and localisation are vital for genuine explanation, and that dynamical

1.2 Interventionism and Mechanistic Explanation

I will outline a few key aspects of Woodward's account of explanation (1997, 2003, 2002, 2008, 2013) namely ideal interventions, the notion of invariance, and counterfactual explanation. In addition, I will show how mechanists appeal to Woodwardian interventionism, and incorporate it into mechanistic explanation. The foundation of much mechanist thought is Craver's (2007) influential account of mechanistic explanation, which in turn heavily integrates Woodward's notions of ideal interventions, invariance and counterfactual explanation. Craver's account has in turn been integrated by a wide range of mechanist philosophers (Bechtel 2008; Darden 2008; Piccinini & Craver 2011; Zednik 2011, 2014; Bechtel & Abrahamsen 2010, 2013; Bechtel & Richardson 1993; Kaplan 2015). When speaking about mechanists or mechanistic explanation in this section, I am referring to this contingent of mechanist thinkers and their approach to explanation.

1.2.1 Causal relevance and ideal interventions

The issue of causal relevance has caused difficulties for several accounts of explanation.[4] If an explanation is unable to distinguish which facts are causally relevant (and should be included in the explanation) and which are causally irrelevant (and should be excluded) then it will fail to properly explain. Explanation requires a description of the causal structure (Salmon 1984) that resulted in the explanandum phenomenon—and describing the causal structure requires a clear picture of what features of the world are causally relevant.

Woodward differentiates causally relevant features of an explanation from causally irrelevant ones by employing an interventionist framework. Woodward's claim is that "causal (as opposed to merely correlational) relationships are relationships that are potentially exploitable for purposes of manipulation and control." (Woodward 2008, pg. 219). Those relationships that can be possibly intervened on are causal relationships, and relevant from an explanatory standpoint.

This process is formalized by Woodward as:

> (M) X causes Y if and only if there are background circumstances B such that if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y would change. (Woodward 2008,, pg. 222).

models only explain when they incorporate mechanistic details. I bracket this issue here, since my focus is on the interplay of interventionism and dynamical explanation, not the necessity of mechanistic detail for explanation.

[4] For an overview of the recurring problems with several other (now historical) accounts of causal relevance, see Craver (2007).

Establishing causation therefore requires finding relationships that can be manipulated. Here manipulation is cast as an ideal intervention. An ideal intervention is a manipulation made on the value of a variable, and it is ideal because the intervention needs to be possible in principal, but not necessarily in practice. The target might be too big, too small, too far away, or in some other condition that makes an actual intervention impossible. Woodward makes this point to avoid an anthropocentric concept of causation and intervention—causation would still be a feature of nature without humans around, and interventions occur all the time without the involvement of humans. An intervention on X ought also to isolate its effect on Y and eliminate confounding variables.

Mechanistic explanations make heavy use of Woodward's notion of an ideal intervention for establishing causal relevance:

> To say that one item (activity, entity or property) is relevant to another, is to say, at least in part, that one has the ability to manipulate one item by intervening to change another. (Craver 2007, pg. 93–94)

The use of ideal interventions is a crucial aspect of mechanistic explanation: Craver claims that Woodward's account of causal relevance provides "an essential normative component to previous counts of mechanistic explanation" (Craver 2007, pg. 105) marking it as an improvement over the previous mechanist accounts produced by Glennan (1996), Machamer et al (2000) and Bechtel & Richardson (1993).

1.2.2 Invariance

The next component of Woodward's account is invariance, a measure of both the stability and causal potency of a generalisation, and that bears some resemblance to the notion of a law of nature. While Woodward (2003) does not consider laws of nature to be crucial to explanation, he is however keen to distil out something useful from the notion of laws. What laws do for explanation, Woodward claims, is to provide stable generalisations that hold between a collection of variables. The problem of distinguishing a generalisation from a bona fide law of nature is a moot point, since Woodward argues that lawfulness is not really the useful or interesting property of a given generalisation. The truly important distinction is between generalisations that are accidental, and those that are invariant.

Accidental generalisations (everyone in this room is currently sitting down; every coin in my pocket is silver) are unstable, do not expose causal relationships and are hence not particularly explanatorily useful. Distinguishing merely accidental generalisations from more stable and explanatory useful generalisations is, for Woodward, conducted by establishing the invariance of these generalisations:

Because invariance is the key to explanatoriness, we don't need to decide whether a generalization counts as a law (and hence we don't need to find a sharp dividing line between laws and nonlaws) to distinguish between the explanatory and the nonexplanatory. (Woodward 2003, pg. 183–4)

> ...it follows that whether or not a generalization can be used to explain has to do with whether it is invariant rather than with whether it is lawful. (Woodward 2000, pg.2)

How can invariance be established?

> A generalization is invariant if (i) it is... change–relating and (ii) it is stable or robust in the sense that it would continue to hold under a special sort of change called an intervention. (Woodward 2000, pg. 198)

Establishing invariance necessarily requires the use of interventions. Both (i) and (ii) involve the application of interventions. In the case of (i) the goal is to establish that this set of relationships between variables depicted in the generalisation is "change–relating" (also known as "difference–making" - both are synonyms of causally relevant).

This criterion eliminates the kinds of accidental cases discussed earlier. For instance, the case where being present in this room (P) is related to sitting down (S), the truth of P relates to the value of S. However, this generalisation breaks down if we intervene on P and S and observe their effects on one another—someone could stand up while remaining in the room, or be outside the room but sitting. This relationship between P and S is merely accidental, and we cannot explain the value of P or S by citing this relationship.

For criterion (ii), Woodward also states that this relationship needs to demonstrate stability, and show that it holds under a "range of interventions". The more stable the relationship, the more explanatory "depth" is established, meaning a broader range of scenarios under which the generalisation continues to provide explanatory power. It is important to note that neither stability (and hence invariance) is a binary condition, but rather stability "admits of degrees" (Woodward 1997, pg. 34). A generalisation that remains stable over a great range of interventions is explanatory over a greater range of scenarios, but this does not exclude less stable generalisations from explaining within a narrower range of conditions.

Invariance is a key part of mechanistic explanation, and there is a broad acceptance by mechanists that "interactions in a mechanism should be characterized in terms of invariant change relating generalizations" (Kaiser & Craver 2013, pg. 25). Craver claims that mechanistic explanation 'relies closely on James Woodward's account of the role of invariance in explanation' (pg. 94), and integrates the criteria of change–relating and stability as important for explanatory power:

> ...causal relations need not be universal to be explanatory, nor need they be unrestricted in scope, nor need they lack any reference to particulars. All that matters is that there is some stable set of circumstances under which the variables specified in the relation exhibit the kind of manipulable relationship sketched above. (Craver 2007, pg. 100).

1.2.3 Explanation

So far, we have seen how interventions establish the causal relationships between variables, and that these kinds of causal relations can be captured and described systematically in the form of invariant generalisations. In addition, mechanistic explanations make use of both ideal interventions and invariance in order to provide explanatory accounts.

Following from this, Woodward claims that "explanation is a matter of exhibiting systematic patterns of counterfactual dependence" (Woodward 2003, pg. 191). Once we know the "systematic patterns" (the invariant generalisations) at work in producing a phenomenon, we can also investigate counterfactual scenarios. Knowing what happened, and what would have happened if circumstances had been different (counterfactual dependence) is the key to explanatory power.[5] A model that provides a combination of a description of the invariant generalisations at work, and predictions of what would happen in counterfactual scenarios based on the description of those generalisations, is explanatory. Explanations in the Woodwardian sense can tell us why things obtain, and what would have happened had circumstances been different.

Mechanistic explanations also utilize counterfactual notions of explanation. A mechanism shows the counterfactual dependencies operating between its components (a collection of invariant generalisations between components), and in this way, provides an explanation of why a certain phenomenon occurred, and how things would have occurred in different conditions.[6]

[5] Those who do not subscribe to counterfactual explanation or interventionism may already have different views on nonmechanistic explanation. However, my targets are specifically those mechanists who do uphold Woodwardian interventionism.

[6] Craver (2007) extends this counterfactual account by also specifying that a mutually manipulable, inter-level constitutive relationship must hold between the mechanism and the explanandum phenomenon. The scope of this article, however, only covers the causal and counterfactual side of mechanistic explanation and its relationship with interventionism.

1.3 Covering–Laws and Dynamical Explanation

1.3.1 Dynamical models

Dynamical models are a kind of descriptive model used to investigate systems by way of the mathematical tools of dynamical systems theory (DST). The techniques of DST developed out of the field of synergetics (Haken 1983; Kelso 1995), and have come to be applied across a wide variety of domains including circadian rhythms (Bechtel & Abrahamsen 2009, 2010, 2013); infant locomotion (Thelen & Smith 1994) and reaching behaviour (Thelen et al 2001; Smith & Thelen 2003); robotics and artificial intelligence (Beer 1995); human coordination (Haken et al 1985; Mechsner et al 2001); and cellular genetics (Huang et al 2005). At the heart of these models are differential equations which model the often nonlinear and highly complex relationships between those variables.

The Haken–Kelso–Bunz (HKB) (Haken et al 1985) model of bimanual coordination is a frequently invoked example of a dynamical model (Kelso 1995; Stepp et al 2011; Chemero 2011; Lamb & Chemero 2014; Kaplan 2015). The HKB model is a dynamical model that attempts to describe the phenomenon of bimanual coordination. Bimanual oscillations (“wagging” the index fingers of both hands at the same time) can be conducted in either in–phase or anti–phase conditions.

The differential equation used in the HKB model:

$$\frac{d\phi}{dt} = -a \sin \phi - 2b \sin 2\, \phi$$

Here ϕ is relative phase, having a value of either 0 degrees or 180 degrees (representing in– and anti–phase conditions respectively) and b/a is the coupling ratio inversely related to the frequency of oscillations.

The variables in the HKB equation track abstract mathematical features of a system, and show how the system would behave over time given certain input values. In practice, the HKB model and other dynamical models are often highly accurate descriptors and predictors of the modelled system’s behaviour (Chemero & Silberstein 2008; Chemero 2011; Silberstein & Chemero 2013).

Using the HKB model as an exemplary case, this section will examine two properties ascribed to dynamical models: that they provide covering–law explanations, and that their predictive power is indicative of explanatory power.

1.3.2 Covering–law explanation

Many dynamicists have claimed that dynamical models like the HKB model can be more than merely descriptive, and can provide genuinely explanatory accounts of phenomena (Kelso 1995; Van Gelder 1995, 1998; Bressler & Kelso 2001; Chemero & Silberstein 2008; Walmsley 2008; Stepp et al 2011; Silberstein & Chemero 2013; Lamb & Chemero 2014).[7]

One common claim is that dynamical models produce something like lawful statements about systems:

> ...the brain is fundamentally a pattern forming self–organized system governed by potentially discoverable, nonlinear dynamical laws. (Kelso 1995, pg. 257)

The same sentiment is echoed in Bressler & Kelso (2001), who argue that the HKB model "exemplifies a law of coordination that has been found to be independent of the specifics of system structure" (pg. 28).

These dynamicists are invoking a long–standing conception of explanation as being based on laws of nature. An influential account of how laws explain comes in the form of the covering–law model of explanation (Hempel & Oppenheim 1948; Hempel 1962a; 1965). The covering–law model suggests that:

An explanatory account may be regarded as an argument to the effect that the event to be explained...was to be expected by reason of certain explanatory facts. These may be divided into two groups: (i) particular facts and (ii) uniformities expressed by general laws. (Hempel 1962a, pg. 10; quoted in Salmon 1984, pg. 19).

Hence a covering–law explanation ought to provide a set of facts governed by a law, and show how a logical necessity holds between these facts and the governing law. For instance, the orbit of a planet is governed by Kepler's first law:

L1. The orbit of a planet is an ellipse with the Sun at one of the two foci.

We can combine the antecedent facts (the location, velocity and acceleration of the planet as well as the location of the Sun) with Kepler's first law and show that the planet's orbit was to be expected based on these facts. This set of facts combined with this law would qualify as an explanation of the planet's orbit under the D–N model.

Walmsley (2008) attempts to fit dynamical accounts into this covering–law mode of explanation:

> ...the explanatory goal of dynamical cognitive scientists is to provide covering–law explanations whereby a cognitive phenomenon is explained by way of citing the laws (qua differential equations) that govern the system that produces it. (pg. 344).

[7] Though the nature of the distinction between description and explanation is contentious, the existence of a distinction is generally accepted. Advocates of both mechanistic (Craver & Kaplan 2011; Kaplan & Bechtel 2011; Kaplan & Craver 2011) and dynamical approaches (Chemero & Silberstein 2008; Silberstein & Chemero 2013) agree that there is a difference between mere description and explanation of a phenomenon.

The central claim from Walmsley is that the equations used in dynamical modelling can become explanatory within the context of a covering–law explanation. Dynamical equations such as those used above can act as governing laws for the system they represent, and they describe a logical necessity operating between the antecedent conditions and the equation much as laws of nature do in Hempel's account.

> Depending on our interests, then, we can insert the values we know into the equation, and solve the equation in order to find the values we do not know. Finding the values for a, b, and ϕ when d/dt is 0 would constitute to an explanation of why d/dt takes the value it does...(Walmsley 2008, pg. 341).

An explanation produced using this method would have explanatory power because it can tell us how, based on the logical necessity operating between the values of the variables b/a and ϕ that the result was to be expected.

1.3.3 Prediction

In addition to claims about covering–law explanation and dynamical models, some dynamicists have claimed a direct connection exists between the predictive powers of dynamical models like the HKB model and their capacity to explain phenomena. The HKB model accurately predicts several important outcomes of bimanual coordination, namely that there will exist two stable basins of attraction at low frequencies (in– and anti–phase), as well as phase switching at certain critical values of k (Kelso 1995; Walmsley 2008). Chemero & Silberstein (2008) argue that this predictive power is an indicator of a deeper explanatory power:

> If models are accurate enough to describe observed phenomena and to predict what would have happened had circumstances been different, they are sufficient as explanations. (pg. 12).

This claim also appears in Walmsley's account, where he attributes explanatory power to predictive models like the HKB model. This is justified by a quirk of the covering–law account:

> ...the only difference between a prediction and a covering–law explanation is whether or not the state of affairs described in the explanandum is known to have obtained. It is therefore solely a pragmatic difference-prediction and explanation have an identical logical structure, but differ in terms of what one knows and what one wants to know. (Walmsley 2008, pg. 340).

In the same vein, Stepp et al (2011) also claim that "...dynamical explanations show that particular phenomena could have been predicted, given local conditions and some law–like general principles..." (pg. 432). They also note the significance of the counter–factual supporting nature of the model, since "we can use the mathematical model to make predictions of the activity of the slave system with so–far–unobserved activity in the master system." (pg. 432). In their view, this combination of prediction and counterfactual support is explanatory within the context of a covering–law explanation.

What follows in Sections 4, 5 and 6 are three lines of criticism presented by mechanist philosophers who are critical of the possibility of dynamical explanations that appeal to the covering–law model.

1.4 Causal Relevance

1.4.1 The causal relevance concern

The causal relevance concern regards the apparent inability for dynamical explanations to establish causal relevance, and is a development of long standing criticism of covering–law explanations. While an interventionist approach allows mechanistic explanations to determine causal relevance, covering–law explanations have no comparable technique. Covering–law explanations are therefore at risk of excluding causally relevant variables, and including causally irrelevant variables—there is effectively no way of knowing which variables are relevant to the explanation and which are not. They cannot tell us about the causal structure that produced a phenomenon (Salmon 1984).

Without an account of causal relevance, we cannot give a causal account of why a phenomenon occurred, since "...the line that demarcates explanations from merely empirically adequate models seems to correspond to whether the model describes the relevant causal structures that produce, underlie, or maintain the explanandum phenomenon." (Craver & Kaplan 2011, pg. 602). Dynamical models do not by themselves tell us anything about causal structure, and so according to this critique they remain in the category of descriptive models, without explanatory power.

Craver & Kaplan (2011) claim that this flaw in covering–law explanation is generally accepted and long standing, stemming from critics of the covering–law approach like Salmon (1984, 1989). Their rejoinder to dynamicists on causal relevance is a "reminder from the past 6 decades of philosophical work on scientific explanation" (pg. 602) that explanations need to provide an account of the causes of a phenomenon, not just describe it. Since causal relevance remains a problem for covering–law explanation:

> ...there is no currently available and philosophically tenable sense of 'explanation' according to which such models explain even when they fail to reveal the causal structures that produce, underlie, or maintain the explanandum phenomenon. (Craver & Kaplan 2011, pg. 602)

Hence dynamical explanations would be at best merely descriptive pseudo–explanations, which fail to identify the causal relationships among variables.

1.4.2 Intervening on dynamical models

I argue contrary to Craver & Kaplan (2011) that dynamical models can in fact provide facts about causal relationships, and ultimately explain. There are no grounds on which to assume that dynamical models cannot also utilize Woodwardian interventionism to establish causal relevance. While mechanist philosophers consider mechanisms to be the best subject for the interventionist framework, Woodward's account is shown here to also be useful to a non–mechanistic and dynamical account of explanation.

For mechanists who are proponents of Woodwardian interventionism, causal relationships are those relationships that can be exposed via ideal interventions. Mechanisms are essentially bundles of causal relationships, and experimental interventions are necessary to show how a component is causally relevant to the mechanism, and thereby reveal those causal relationships. Describing these relationships is the basis of explanatory power.

By applying this same Woodwardian notion of ideal interventions to dynamical models, the causal relevance concern can be countered. If we can intervene on the values of variables in a dynamical model, and see changes in the value of another variable, then (on the interventionist account) we have exposed a causal relationship. I am not arguing for any modification to Woodwardian interventionism—rather, I am arguing for its application outside of just mechanistic models and explanations, something which Woodward's account is perfectly capable of doing.[8]

[8] Another related question is the possibility of macro-to-micro level causation (Baumgartner 2009, 2010; Baumgartner & Gebharter 2015). In the case of HKB, this would mean that b/a and ϕ being macro-level variables of the dynamical system that they supervene on, are not capable of causation independent of their micro-level supervenience base. If one cannot intervene on ϕ without also simultaneously intervening on the micro details which form the supervenience base for ϕ then supposedly there cannot be causation from the macro-level down to the micro-level. Woodward (2015) offers a solution by adding a clause to his requirements for interventions specifying that non-causal relations (like constitutive or supervenience relations) do not need to be controlled for during interventions on macro variables, which means macro variables are not prone to systematic overdetermination of effects or becoming epiphenomena. While Baumgartner & Gebharter (2015) argue that this adjustment undermines Cravers account of constitutive relationships (and take this as a motive to find an alternative) my purpose here is not to defend mechanistic explanation, and hence whatever violence is done to mechanistic constitution relations is not for me a good reason to reject Woodwards amendments.

1.4.3 Test case I: The HKB model

In order to test out this argument, I will consider empirical studies of the HKB model and determine whether they represent successful interventions that exposed causal relationships within the model. In order to establish these causal relationships, there needs to be an intervention on the value of the variables concerned, in this case relative phase (ϕ and frequency (b/a).

For b/a to have a causal relationship with ϕ it would need to be shown that:

> (M) b/a causes ϕ if and only if there are background circumstances B such that if some (single) intervention that changes the value of b/a (and no other variable) were to occur in B, then ϕ would change.

Scholz & Kelso (1989) increased and decreased the frequency of queues which subjects were instructed to match the frequency of their oscillations to. This represents an intervention on to an intervention on b/a. The results were in line with the HKB model's predictions, showing that when oscillation frequency is slow (b/a >0.25) there are two stable attractors in both in–phase and anti–phase conditions. As frequency increases (b/a $=<0.25$) the anti–phase attractor disappears, and the system is liable to fall towards the in–phase attractor.

This satisfies (M) so far. The coupling ratio, b/a, is intervened on by increasing or decreasing the frequency of oscillations. Interventions on the value of b/a result in regular changes in relative phase, ϕ. Hence, the relationship from b/a to ϕ is a causal relationship. Under the scheme of Woodwardian interventionism, the relationship described by the HKB model between b/a to ϕ appears to meet the criteria for being "difference–making" or causal. This example shows how the interventionist solution to the problem of causal relevance can be applied to non–mechanistic, dynamical models just as well as it can to mechanistic models.

1.4.4 Test case II: Dynamical field model

My second test case is the dynamical field model of infant perseverative reaching developed by Thelen et al (2001), and discussed also by Zednik (2011). Thelen et al (2001) model the reaching behaviours of infants in an A–not–B error task. The A–not–B error occurs when infants are induced to repeatedly reach for a desirable object (a toy) which is hidden in one of two locations—location A or location B. When the infant witnesses the toy being repeatedly hidden at A, they are prone to erroneously continue reaching for A even if in subsequent trials the toy is clearly shown being hidden under B (hence it is called the A–not–B error).

The model investigates the variables that can influence this tendency to reach for either A or B. The dynamical field model relates the values of variables at each point (x) on a field overlaid onto the task environment—for instance, xA and xB will be the locations on the field where the hiding spots A and B are located. Each location has an activation value (u) and when u reaches a critical value, reaching is likely to occur towards that location. Several variables play a role in determining how activation changes over time, as the model demonstrates:

$$\tau\dot{u}(x,t) = -u(x,t) + S(x,t) + g[u(x'); x']$$

Where activation level ($\dot{u}$) of every point (x) on the field changes over time (t) as a function of the field's previous activation (u), an input vector (S), a cooperativity parameter (g), and a temporal decay constant (τ). I will investigate the effect of S on u. S has, according to the model, some kind of effect on the value of u—what I aim to establish is whether or not interventions on S do in fact reveal a causal relationship with u.

According to Thelen et al (2001) there is considerable experimental evidence for the causal relationship from S to u. S represents several concurrent inputs—task, specific, and memory. Interventions on the task input such as changing the distinctiveness or desirability of the toy increases or decreases activation at the chosen location (xA or xB) (Diedrich et al 2001). Specific input can be intervened on by giving cues to reach (the experimenter tapping or gesturing to a particular location) which similarly influences u at that xA and xB (Smith & Thelen 2003). The effect of memory input is that of these previous inputs, which continue to influence u.

Like the HKB model, I think that the dynamical field model of infant perseverative reaching is therefore also a good example of a dynamical model that can be treated in interventionist terms to establish causal relationships. The input variable is a cause of changes in the value of activation, and is a difference–maker to the outcomes of the system. Whether or not the infant reaches for A or B is controlled by inputs to S, and this relationship satisfies Woodward's requirement (M).[9]

[9] (M) also requires that the relationship be between b/a to $\phi$and excluding other confounding variables. This is a sensible requirement since we would like to know that b/a is the cause of ϕ and not some other hidden variable or a combination of b/a and some other variable. Much like for non-dynamical or mechanistic models, identifying and controlling for confounding variables is done on a case-by-case basis, with a view to (at least) conceptually disentangling confounding variables and controlling for them (Woodward 2003). In the case of the dynamical field model, it would be important, per Woodward (2003), to correct for the influence of (for example) the cooperativity input g when determining the causal role of S on u. While in practice it may be impossible to set up a scenario where g does not also influence u, we only need to be able to imagine an ideal scenario where the influence of g was removed.

1.5 Genuine Laws

1.5.1 The genuine laws concern

The genuine laws concern regards the use of dynamical equations as laws of nature in a covering–law explanation, as proposed by Walmsley (2008). The core of this concern is the uncertainty around what constitutes a genuine law of nature. There is considerable disagreement over what makes a law a law, a criticism that has been directed at law–based explanations for decades (see for example Salmon 1984, 1989). The nature of laws continues to be a live topic of debate and it suffices to say that there is no uncontroversial position on laws. As such, this makes it hard to see how dynamicists like Walmsley (2008) can state with any confidence what makes a dynamical equation a law, given the lack of reliable criteria to judge them by. There is simply no widely agreed upon framework for determining what kinds of generalisations should be accepted as laws (Kaplan 2015).

Despite this disagreement, a common intuition within the literature on laws is that they ought to be exceptionless and universal in scope (Woodward 2003; Kaplan 2015), neither of which seem to apply to dynamical models. Dynamical models might apply to a wide range of phenomena, but they are not exceptionless—for instance, the HKB model does not apply to all examples of coordination such as gait switching from walking to running in humans.

In addition, it is debateable whether laws of nature are applicable within certain scientific domains. For instance, while laws of nature are not uncommon within physics and chemistry, they rarely figure in the explanations produced by biologists, neuroscientists and cognitive scientists (Woodward 2003; Craver 2007; Bechtel 2011). This is coupled with an absence of laws in these domains: "uncontroversial examples of laws are less easy to find in sciences like biology and geology and harder still to find in the social and behavioural sciences." (Woodward 2003, pg. 183). Appealing to laws, then, might not be suitable outside the "hard sciences" of chemistry and physics.

These uncertainties create a problem for advocates of covering–law explanations since they require nomic expectability (Salmon 1984) – that the law can be expected to reliably describe a set of relations between variables. Both the D–N and I–S models which comprise the covering–law mode of explanation depend upon either determinate or highly probable relationships between variables, upon which deductive or inductive arguments can be made. In the case of accidental, non–lawful generalisations there can be no nomic expectability since there is no way of knowing whether the relations between variables it describes will continue to hold in that way. Certainty that a set of relationships is lawful (and as a result dependable and stable) is therefore very important for covering–law explanation.

The follow–up of the genuine laws concern regards the avenues dynamicists can possibly take to remedy the situation, granted the uncertainty surrounding laws. Kaplan (2015) claims that there are two equally undesirable paths open to dynamicists—either they can claim that dynamical equations are ceteris paribus laws, or they can attempt to produce some set of criteria for what a law of nature needs to be. The ceteris paribus option leads into another thicket of uncertainty (Woodward 2002), and the latter option could be a very difficult and long–term task, in light of ongoing debate surrounding such a set of criteria for laws. According to Kaplan (2015), dynamicists are better off conceding that covering–law explanation is not viable at present.

1.5.2 Using invariance in place of laws

When mechanists claim that laws of nature are rarely encountered or applicable outside of physics and chemistry, I agree entirely. However, laws are not the only kind of explanation–worthy generalisation available for a dynamical mode of explanation. Woodward himself argues that lawfulness is the wrong criterion to gauge the explanatory value of a generalisation by. Instead of being concerned about how exceptionless, universal in scope (or some other criteria) a generalisation is, we should look for how invariant the generalisation is in showing how different variables causally relate to one another:

> ...generalizations in the special sciences can be used to explain, as long as they are invariant in the right way, whether or not they are regarded as laws. (Woodward 2003, pg. 183).

Whether a set of relationships are covered by a law of nature is not important. What matters is how invariant they are, and as a result how reliable they are for the purposes of describing stable and robust patterns of causal and counterfactual relations. Non–lawful generalisations directed towards psychological, biological, and cognitive phenomena (and any other phenomena within a "special science") are not precluded from being explanatory so long as they are shown to be invariant.

At this point I am diverging from the covering–law view proposed by Walmsley (2008), derived directly from Hempel (1965), and at which much mechanist ire has been directed. Instead I propose that Walmsley, while generally correct in his assessment of the explanatory power of dynamical models, need not rely on a covering–law account (a direction of development he himself acknowledges). The covering–law account does, however, point us in the right direction in some respects:

> In its emphasis on the role played by generalizations, including those taking a mathematical form, in explanation and causal analysis, the in-

> terventionist account has some affinities with the DN model. (Woodward 2017, pg. 7)

So, the intuitions of dynamicists like Walmsley are more or less correct. Dynamical models provide useful generalisations, and show dependencies between antecedent facts and this generalisation—a valid approach under both the covering–law and interventionist approaches. Woodward however provides a less problematic alternative to laws of nature in the form of invariant generalisations. Instead of talking about laws of nature dynamicists can talk about invariant generalisations. The general form of dynamical explanation, in shifting from a covering–law to an interventionist approach, changes surprisingly little.

While Walmsley (2008) specifies a kind of explanation derived from antecedent facts and laws of nature, using dynamical equations in the place of laws, here I suggest that a more robust account of dynamical explanation would replace laws with invariant generalisations, and thereby entirely avoid the genuine laws concern. My goal is not to rule out the possibility of the existence of laws, or to argue that they have no role in providing explanations. Instead I am bracketing the issue entirely, and showing how according to the interventionist framework dynamical models can figure in explanatory accounts.

1.5.3 Test case I: The HKB model

We can start to form a picture of a mode of non–mechanistic explanation where dynamical equations are explanatorily useful based on their invariance, using the HKB model. The critical question is: can the HKB equation meet the requirements of invariance? An invariant generalisation fulfils two important criteria:

(i). The relationships described in the generalisation are causally efficacious or "difference–making".

(ii). The generalisation is stable under a range of interventions.

The arguments presented in Section 4 already provide an answer to (i): the relationships described in the HKB model can be intervened upon to show how they are causally efficacious, and the model therefore meets this criterion.

An answer to (ii) requires us to determine how stable, under a range of interventions, these causal relationships are. If the model remains descriptive of bimanual coordination under a narrow range of conditions, then its explanatory "depth" will be very limited. This latter criterion (ii) is a sliding scale, compared to the binary condition (i). How invariant a generalisation is depends upon the interests of the investigator—invariance is greater when the generalisation continues to hold over a greater range of values of variables we are interested in.

In the case of the HKB model, if our interests are in explaining bimanual coordination in humans, then the model ought to cover the range of values that humans are capable of. A sufficiently invariant generalisation ought to cover all the scenarios we will encounter when trying to explain this phenomenon, while a less invariant generalisation would only hold across a portion of them. For instance, if the HKB model only held in conditions of low frequency, where b/a >0.25, then it would have a limited invariance.

In view of the experimental results, the HKB model appears to be stable across a sufficiently broad range of interventions—there is no observed frequency, high or low, at which the model ceases to describe the relationship between b/a and ϕ. For the purposes of empirical research on human subjects (i.e. Scholtz & Kelso 1987) the model is stable. We could imagine circumstances where the model might break down—for instance if the frequency became too rapid for human subjects to maintain. However, invariance does not require universal scope, and admits of degrees. For the purposes of explaining bimanual coordination in humans, the HKB model appears to be sufficiently stable and hence invariant (Kaplan 2015) also discusses this scenario at length).

The dynamical equation at the heart of the HKB model qualifies as an invariant generalisation, fit for the purposes of explanation, filling a similar role to laws of nature in the covering–law mode of explanation. The relationship between b/a and ϕ described in the HKB model meets Woodward's criteria for an invariant generalisation, and it is therefore explanatorily useful.

1.5.4 Test case II: Dynamical field model

Like the HKB model, the dynamical field model equation meets the criterion (i) for invariance—the target variable S, when intervened upon, reveals causal relationships between itself and the effect variable, u. The model describes the underlying causal structure of infant perseverative reaching and the A–not–B error.

The model can also meet criterion (ii), demonstrating sufficient stability to be an invariant generalisation and fit for explanation. The model is stable enough that constructing experimental scenarios based on the model across a wide variety of circumstances—even in extreme scenarios, like where no toy or cue from the experimenter is presented at all, representing a null value for the task and specific inputs elements of S (Thelen et al 2001)—the model continues to accurately describe and predict the infant reaching behaviour (Smith & Thelen 2003).

Under these kinds of circumstances that we might want to investigate in order to explain infant perseverative reaching, the dynamical field model remains stable under intervention. Hence, I think the dynamical field model is sufficiently stable to meet criterion (ii), and subsequently qualifies as an invariant generalisation.

1.6 Prediction

1.6.1 Predictivism

The third criticism of dynamical explanation is what some mechanists have called predictivism (Kaplan & Craver 2011), the position attributed to dynamicists who hold that prediction is sufficient as explanation. As we have seen, dynamicists like Chemero & Silberstein (2008) consider the predictive powers of dynamical modelling to be indicative of some deeper, explanatory power on the part of those dynamical models.

Mechanists have responded by pointing out the insufficiency of prediction for explanation. One frequently invoked example is that of a barometer (Salmon 1989; Kaplan & Craver 2011; Kaplan 2015). A barometer, which measures air pressure, is excellent at predicting rainy weather. A sudden drop in air pressure is correlated with a change in the barometer's readings, which is in turn correlated with imminent rain. However, it would be false to claim that as a result of these readings, the barometer is the cause of the rain, or that the barometer readings somehow explain the phenomenon of rain.

This criticism alleges that what dynamicists are doing is conflating prediction (which can be purely correlational) with genuine causal explanation. Hence claiming that the predictive powers of dynamical models make them explanatory is misguided, since there is no justification for suggesting that any amount of predictive power will grant explanatory power to a model.

1.6.2 Crude and invariant prediction

In their use of the barometer example, mechanists like Kaplan & Craver (2011) are referring to an example of what I will call crude prediction. Crude prediction refers to merely correlational, non–invariant relations between variables. A correlation exists between the barometer reading and the imminence of rain, where the reading is a useful predictor of rain, but there is no causal relationship at work between these two variables. This kind of crude prediction is rightly to be thought of as accidental and with little to no explanatory value.

I argue that another kind of prediction, which I will call invariant prediction, can provide explanatory power. While predictions made on the basis of correlations (like the barometer) are evidently not indicators of explanatory power, not all predictions are made on this basis. What I am pointing to here is a combination of counterfactual support and prediction, where a model is capable of showing what would happen if circumstances were different on the basis of an invariant generalisation.

Invariant prediction is made on the basis of the counterfactual–supporting abilities of an invariant generalisation – and these predictions are the kind of counterfactual dependencies that Woodward–style explanations are built on. Counterfactual dependency, as a point of clarification, is distinct from mere non–causal description in that it shows how one or more causal relationships connect to one another. It does this in a sense that is more than just establishing what is "barely true" by appealing to certain "truth–makers" like invariant generalisations. (Woodward 2008, pg. 230)

Crude prediction does not result from knowledge of invariant generalisations and as such is merely correlational and accidental. It is not indicative of explanatory power. Invariant prediction is conversely made on the basis of knowledge of the stable causal relations governing that system, and how that system will behave under a range of different counterfactual scenarios. If explanation is a matter of describing counterfactual–supporting invariant generalisations (Woodward 2003), then the claim I am making here about invariant prediction and counterfactual support seem largely in line with this explanatory goal. A model which can provide invariant predictions based on the counterfactual dependencies it describes is an explanatory model.

For instance, the dynamical field model can predict the outcomes of infant perseverative reaching based on knowledge of the causal relationships between variables, within a sufficiently stable (and hence invariant) generalisation. Even without actually observing the outcomes associated with the input variable S holding a particular value, the invariance of the equation means that we can entertain the counterfactual, and predict what the outcome would have been. These predictions, based as they are on a sufficiently invariant generalisation, have the same kind of explanatory import as observations of particular outcomes. Whether or not the predicted result actually obtains or not is not significant: if made on the basis of an invariant generalisation then a prediction has explanatory power.

This approach to prediction and explanation comes close to resembling the covering–law mode of explanation, but without its attendant flaws—the causal relevance concern and the genuine laws concern—which should not trouble the present account since, as we have seen in Sections 4 and 5, it resolves them by integrating Woodward's notions of ideal interventions and invariant generalisations.

1.6.3 Interventionist Criticism of the HKB Model

Recently, Woodward has been critical of the HKB model's capacity to provide explanations. He argues that the HKB model is a "non–starter" (Woodward 2017, pg. 23) for explanation, on the grounds that the model fails to associate itself sufficiently with features of the world, and is therefore not explanatory:

> To the extent that the theory does not specify at all what structures or relations in the world are supposed to correspond to the dependency relationships postulated in the theory, then, according to the interventionist framework, it is not even a candidate for an explanatory theory. (Woodward 2017, pg. 22).

Woodward's requirement is something like a softer version of Kaplan & Craver's (2011) 3M requirement, where there must be a direct mapping of variables in a model to components of the mechanism being modelled. This version is asking (without the explicitly mechanistic requirements) for some sort of association between features of the world and the model in question. The problem in its purest form seems to be this: a model cannot explain anything if the model is not clearly associated with an explanandum (features of the world). I take this to be a fairly uncontroversial claim about explanation—explanations need to be about something, or they are not explanations. If Woodward's criticism hits the mark, then the HKB model is not sufficient as an explanation of bimanual coordination because it does not tell us enough about what features of the world the model is actually modelling, and hence what the explanandum actually consists in. It is potentially "... just a mathematical structure or an entirely uninterpreted set of equations relating certain variables" (Woodward 2017, pg. 22). By extension, this could cause trouble for the explanatory power of other dynamical models.

I think that this claim lacks force though, because Woodward's mapping requirements are indistinct. What counts as sufficient grounding in features of the world seems very open to interpretation. For instance, Woodward thinks that while the HKB model is a non–starter for explanation, the dynamical model of the Hodgkin–Huxley model of the action potential is explanatory, because it associates itself with a physical system—it is about the neuronal action potential. However, I disagree that there is a significant difference between Woodward's chosen case of the HH model, and the HKB model, or that Woodward has so far articulated some principled way to separate them. The HKB model is also about features of the world—it is about the physical system which exhibits bimanual coordination. When experimenters conduct research on the HKB model and bimanual coordination they are surely investigating some features of the world. The fact that experimentally testing the HKB model can be carried out at all seems to eliminate the possibility of the model being so vague about its constituents as to be a non–starter.

Lacking a good account of how to determine whether a model is actually associated with a particular system (outside the account of mechanistic explanation described in this paper, which does for appropriately mechanistic phenomena) is a problem for non–mechanistic accounts of explanation. However, the account I offer in this paper goes some way towards resolving this concern. Using interventionism in concert with dynamical models, we can establish the closeness of fit between a dynamical model and the causal structure of the features of the world it is modelling. The best models (and those that explain) will be those that describe that causal structure accurately according to an interventionist account. Development of the approach to dynamical explanation outlined in this paper may help further assuage this concern.

1.7 Dynamical Explanation

Mechanist criticisms of dynamical explanation are not fatal to the enterprise, and dynamical explanation is viable. By modifying the covering–law account proposed by Walmsley (2008) to instead resemble Woodward's (2003) account of explanation, dynamical explanation can avoid potential problems of causal relevance, concerns about the status of dynamical equations as laws of nature, and the use of prediction as an indicator of predictive power.

So, what should an account of dynamical explanation look like? Firstly, dynamical models can provide causal explanations, so long as an ideal intervention can expose the causal relationships between variables featured in the dynamical equations featured in the model. In addition, the covering–laws proposed by Walmsley can be instead thought of as invariant generalisations, which operate in much the same way as laws of nature but without the requirements of being universal in scope or exceptionless. If dynamical equations meet Woodward's requirements for invariance, they are suitable as invariant generalisations. A combination of dynamical equations and ideal interventions on the variables features in those equations is sufficient for explanation. These dynamical explanations also furnish invariant predictions of counterfactual scenarios, and these predictions contribute to explanatory power.

The HKB model and the dynamic field model provide exemplary cases of dynamical explanation. Firstly, ideal interventions on the variables at work in these models show how relative phase caused by the coupling ratio, and activation is caused by task inputs respectively. Secondly, these equations are useful as invariant generalisations, which possess sufficient stability and difference–making capacity to be used for the purposes of explanation. Finally, the ability of these models to predict counterfactual scenarios contribute to their explanatory power when coupled with the invariance of the equations.

1.8 Conclusion

My claim is that mechanist criticisms of dynamical explanation can be overcome by adopting an interventionist perspective on explanation, and applying it to dynamical models. I have shown how the causal relevance concern, the genuine laws concern and the charge of predictivism do not deflate dynamical explanation, and how dynamical models can meet the requirements of Woodward's interventionist account of explanation. Once the account of dynamical explanation is adapted to rely on invariant generalisations and the broader Woodwardian account of explanation it avoids the problems mechanists have targeted. For mechanists to argue against the explanatory power of dynamical models under an interventionist framework, they would need to show either that dynamical models are for some reason not valid candidates for invariant generalisation, or that interventionism itself is flawed. This chapter shows that the former is not true—dynamical models can meet the criteria for invariant generalisations. The latter approach would be equally damaging for mechanistic explanations, since they also hinge upon interventionism. Through this defence I have also outlined an account of how dynamical models can apply the interventionist framework and demonstrate their explanatory power.

Chapter 2

Dynamical Causes

Mechanistic explanations are often said to explain because they reveal the causal structure of the world. Conversely, dynamical models supposedly lack explanatory power because they do not describe causal structure. I argue instead that dynamical models do reveal causal structure and consequently produce dynamical explanations. Taking the example of cell fates from systems biology, I show how dynamical models, and specifically the attractor landscapes they describe, identify fine-grained causes of cell differentiation. These causes are irreducible and inaccessible to mechanistic models. Dynamical models can therefore provide explanations beyond the reach of mechanisms.

This chapter has been published as Meyer, R. (2020). Dynamical causes. *Biology & Philosophy.*

2.1 Introduction

The concept of causal structure of the world, or just *causal structure*, is a touchstone of modern Mechanism[1] originating from Salmon's (1984) causal-mechanical approach to scientific explanation. The claim that a close relationship exists between causal structure and explaining a phenomenon is outlined by Craver (2007):

> There are perhaps many interesting things to be said about explanatory texts, but one crucial aspect of their adequacy has to do with whether explanatory texts accurately characterize the causal structure of the world. (Craver 2007, pg. 27)

A related and complementary claim made by mechanists is that scientific explanations should describe the causal relationships which comprise this causal structure:

> In many areas of science, explanations are said to be adequate to the extent, and only to the extent, that they describe the causal mechanisms that maintain, produce, or underlie the phenomenon to be explained, the explanandum phenomenon. (Kaplan & Craver 2011, pg. 601)

Putting these ideas together, I take causal structure to refer to the comprehensive web of causal relations that underlie or produce a phenomenon. Something explanatory ought to follow on from having a description of causal structure—if you understand all the relationships driving a phenomenon to occur, you have explained it. In short, explanation is all about describing causal structure.

And as mechanists have argued, mechanisms have a special, if not unique, role to play in this revelatory task:

> Mechanisms explain the diverse aspects of the explanandum phenomenon, and so unify them by relating them to an underlying causal structure(Craver 2007, pg. 49)

> [Mechanistic] models...carry explanatory force to the extent, and only to the extent, that they reveal (however dimly) aspects of the causal structure of a mechanism.(Kaplan & Craver 2011, pg. 602)

This closely pairs the notion of causal structure to explanatory power, and mechanisms link the two because, in many cases, a description of a mechanism is necessary for describing causal structure. Driving the point home, these mechanists argue that explanations necessarily capture the totality of causal relations producing

[1] Following the convention proposed by Glennan & Illari (2018) I distinguish the philosophical stance of Mechanism from the object called a mechanism via a capitalisation added to the former.

a phenomenon, and that this totality—the causal structure—is in large part what differentiates a loose bundle of descriptions from a genuine explanation (Craver 2006, 2007).

But not everyone is satisfied with this mechanistic arrangement. A movement to articulate a mode of non–mechanistic, dynamical explanation based on dynamical models has been afoot for some time (Chemero & Silberstein 2008, Stepp et al 2011). Dynamical models are a kind of mathematical model that employs the tools of dynamical systems theory to capture the unfolding of variables over time using differential and difference equations. These models have a track record of impressive descriptive accuracy and predictive power when applied to various cognitive, neuroscientific and biological phenomena. By their very nature they do not by themselves give much detail about the physical realisers or substrates of the variables they model (van Eck 2018). Proponents consider this advantageous—because of this feature dynamical models can "zoom out" from these fine-grained details and say new and interesting things about the dynamical features of a system.

Breaking with mechanist views on explanation, proponents of dynamicism reject the necessity for mechanistic models as a prerequisite for explanation, and emphasise accuracy in description and prediction as sufficient (Chemero & Silberstein 2008). Dynamicists have mostly conceded that dynamical explanations need not adhere to mechanist standards around explanation, specifically the requirement that they ought to describe causal mechanisms (Stepp et al 2011).

The dynamicist account further departs from the mechanist orthodoxy by suggesting that dynamical explanations "need not respect the underlying causal structures that give rise to system-level dynamics." (Kaplan & Craver 2011, pg. 602). Mechanist philosophers have, as a result, been critical of claims that dynamical models are explanatory. After all, if mechanists are right that causal structure is core to explanation, and mechanisms are crucial to getting at causal structure, then on two counts dynamical models are a non-starter as standalone explanations. This general concern, termed the *causal relevance concern* (Chapter 1) is the biggest hurdle for getting dynamical explanation up and running.

According to mechanists, the solution to the causal relevance concern is straightforward: associating models with a mechanism in line with the model-to-mechanism-mapping (3M) requirement (Kaplan 2015; Kaplan & Craver 2011). The idea behind 3M is that so long as the terms in a dynamical model can be associated with (mapped onto) the mechanistic components underlying the model, then the dynamical model can thereby describe causes:

> (3M) A model of a target phenomenon explains that phenomenon to the extent that (a) the variables in the model correspond to identifiable components, activities, and organizational features of the target

> mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these (perhaps mathematical) variables in the model correspond to causal relations among the components of the target mechanism.(Kaplan 2011, pg. 347)

By being grafted onto a mechanism, dynamical models do say something about the causal structure underlying the phenomenon, namely the temporal and organisational features of the causal relations between mechanistic components, a view also developed and endorsed by Bechtel & Abrahamsen (2010, 2013). Similarly, when Craver & Kaplan (2018) claim that "[n]ot all dynamical models describe causal relations. Explanatory dynamical models do..." it means that properly mechanistic models (with added dynamical details) are explanatory. The status of dynamical models, according to mechanists, is therefore as descriptive tools in service to mechanistic explanations.

On their own, dynamical models do not describe causal (and hence explanatory) relationships,[2] but rather function as a useful tool for describing the temporal organisation of mechanisms (Kaplan 2015). Their utility acknowledged, dynamical models are still in an asymmetrical relationship with mechanisms, "their explanatory value can be seen as clearly depending on the presence of an associated account (however incomplete) of the parts in the mechanism" (Kaplan 2015, pg. 760). There is no causal story a dynamical model can provide that does not, ultimately, boil down to a mechanistic model.

In this paper I will argue against this mechanist interpretation of dynamical models, and provide a novel example of non–mechanistic, dynamical explanation at work. First, I will contest the claim that dynamical models do not describe causal relations, appealing to an even-handed application of interventionist standards (Meyer 2018). This clears the immediate path to allow dynamical models to give descriptions of causal structure. Second, I will show how dynamical models of cell fates, an example borrowed from systems biology, uncover the causal structure underlying this phenomenon. The attractor landscape described by this model, I will argue, reveals the causes of cell differentiation without reference to mechanisms. Thirdly, following from this positive account, I use Woodward's (2010, 2018) and Waters' (2007) notions of specificity and proportionality to further illustrate how attractors are the best candidate difference-makers for cell fate outcomes, and best describe the causal structure of the phenomenon. Finally, I dispute that the causal structure identified by dynamical models of cell fates can or ought to map onto mechanistic models, per the 3M requirement.

[2] It should be mentioned here that there are ongoing discussions regarding the feasibility of non-causal dynamical explanations (e.g. Ross 2015; Chirimuuta 2017). I will however focus specifically on the case for causal dynamical explanations, and bracket the non-causal option.

2.2 Causal Structure and Dynamical Models

Recently several arguments have been made targeting the causal relevance concern generated by the foregoing mechanist picture of explanation (Meyer 2018; van Eck 2018). These authors have claimed that dynamical models can in fact describe causal relations and thereby explain non–mechanistically, while retaining the mechanist's own interventionist framework to make their case.

The main strains of Mechanism all appeal to Woodward's interventionism to support the notion that mechanisms describe causes. However, interventionism itself does not privilege mechanisms as the only possible source of causal relations (Woodward 2003). All that is required to establish causal relations is variables, which dynamical models provide. Therefore, if dynamical models can meet Woodward's criteria for establishing causal relations, then they ought to be considered causal (Meyer 2018).

What are Woodward's criteria for a relation between variables to be considered causal? Woodward supplies (M):

> (M) X causes Y if and only if there are background circumstances B such that if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y would change. (Woodward 2008f, pg. 222).

(M) specifies how ideal interventions can be used to establish causal relevance of variables. These interventions establish the relationship between the value of a variable, X, and the value of a variable Y. Changes in Y which are the direct result of changes in X demonstrate a causal relationship. X is causally relevant to Y if (M) is satisfied. Hence if a variable in a dynamical model can meet the requirements of (M), then it ought to be considered a cause.

Meyer (2018) uses the example of the Haken-Kelso-Bunz (HKB) (Haken et al 1985) model of bimanual coordination to demonstrate how dynamical models can describe causes. Bimanual coordination is the phenomenon whereby synchronised movements on either hand (in this case moving index fingers from side to side) can be coordinated to move either in-phase, or anti-phase (where ϕ =0 and 180 respectively). To accomplish this, the HKB model uses a differential equation to map the system's evolution over time:

$$\frac{d\phi}{dt} = -a \sin \phi - 2b \sin 2\, \phi$$

Where ϕ represents relative phase, ranging between 0 degrees and 180 degrees (in- and anti-phase conditions respectively); and b/a relates the frequency of oscillations.

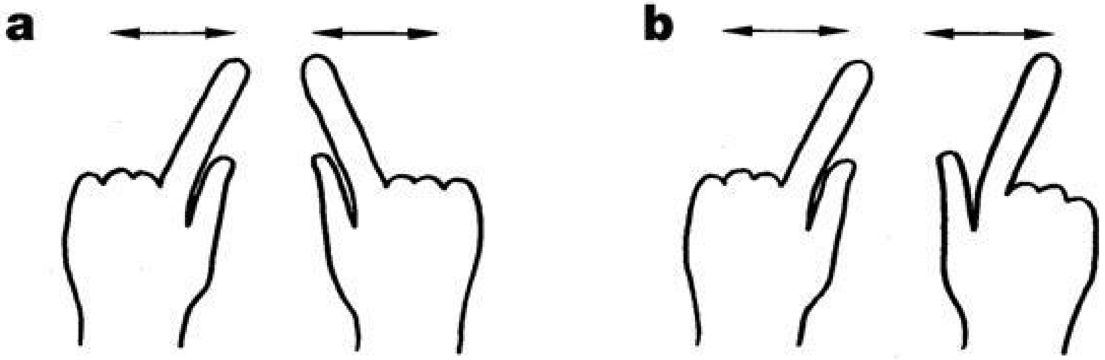

Figure 2.1: An illustration of bimanual coordination, where (a) represents in-phase coordination, and (b) represents anti-phase coordination. Reproduced from Mechsner et al (2001).

In order for the relation from b/a and ϕ to be causal, the following would need to hold:

> (M) b/a causes ϕ if and only if there are background circumstances B such that if some (single) intervention that changes the value of b/a (and noother variable) were to occur in B, then ϕ would change.

Experimental interventions into this system involve changing the frequency of oscillations (b/a) in order to observe their effect on relative phase (ϕ. Subjects are tasked with attempting to maintain bimanual coordination in either the in- or anti-phase conditions, and also match their frequency of their movements to cues given by the experimenters. This paradigm was used in Scholz & Kelso (1989), who intervened to increase and decrease the frequency of the cues provided to the subjects. Several predictions of the HKB model were validated in the results of these experiments: when oscillation frequency is slow (b/a >0.25) both in-phase and anti-phase patterns of coordination are quite stable—subjects are able to maintain these movements without altering phase. But at higher frequencies (b/a =<0.25) the anti-phase pattern ($\phi = 0$) becomes difficult to maintain, and subjects tend to slip into an in-phase pattern ($\phi = 180$).

The key point here is that the relationship between b/a and ϕ is not merely a correlation. There is a direction established by this experimental intervention from b/a to ϕ from cause to effect. Further these interventions produce regular, function-like changes in the value of ϕ. The variable b/a is the difference-maker to ϕ—while other variables may provide necessary background conditions (B), it is b/a that causes ϕ to change in value. So in this case, (M) should be satisfied as a straightforward example of a causal relation—the variables in the HKB model describe causes. In

Woodwardian terms, they describe difference-makers, the systematic relationships from cause to effect.[3]

2.3 Cell Fates

Having outlined the basic framework for causal, dynamical explanation, I now turn to the novel case of *cell fates*, a topic of significant interest in contemporary systems biology and cell genetics. An almost ubiquitous feature of animal cells is their capacity for differentiating into cell fates—alternate, stable phenotypes expressing new traits. Frequently one kind of primogenitor (undifferentiated) cell can differentiate into several distinct cell fates, each exhibiting a different phenotype. Some stem cells, for instance, are bi- or multi-potent, meaning they have the potential to transition into two or more stable phenotypes respectively.

These fates are interesting for a few reasons. Firstly, they tend to be stable and the transitions into them reliably one-directional under normal circumstances. Once differentiated, cells do not tend to "un-differentiate" backwards into progenitor cells or switch over into a different fate. Secondly, cells tend to transition through a series of phenotypes in between the progenitor phenotype and cell fate phenotype in a very directed fashion—even if external perturbations disrupt this typical course, they still find their way to a stable fate.

One story invoked to explain these features of cell fates is a kind of genetic pre-destination. Each phenotype, on this interpretation, must contain some kind of instruction for how to progress to the next phenotype, and that phenotype to the next, and so on down the line. Alternatively, the cell may receive external signals that help direct and drive these transitions and maintain them.

These conceptions have proven too coarse-grained in many situations (Huang 2012). More recent developments in cell genetics and systems biology suggest that differentiations are driven not in a step-by-step or externally controlled fashion, but in a self-organised process driven by networks of thousands of genes engaged in extremely complex interdependent relationships, with all kinds of endogenous activity determining the transitions between fates, as well as their relative stability.

The need to make sense of these complex relationships is in part responsible for

[3] I acknowledge here the significant debates around higher-level interventions in the mechanist literature, particularly the problem of fat-handedness: intervening on a higher-level variable necessitates simultaneously intervening on its supervenience base, and hence violating the interventionist requirement for isolating a single variable for intervention (see Baumgartner & Gebharter 2017, Krickel 2017). I bracket this substantial discussion by adhering to Woodwards (2015) clarification to (M). Woodward specifies that non-causal supervenience relations between micro- and macro-levels need not be held steady in the same fashion as causal relations, so that "properties that supervene on but that are not identical with realizing properties can be causally efficacious." (Woodward 2015, pg. 303)

the development of gene regulatory networks (GRNs) as a modelling tool. A GRN is a network map of the relationships between the many genes that make up the genotype of a given cell, a series of "layers of molecular regulatory networks and cell-cell communication networks - a web of interactions through which genomic information must percolate to produce the macroscopic phenotype" (Huang 2012, pg. 153). These "interactions" consist of each gene's expression behaviour, namely what proteins it instructs a cell to transcribe, and how this influences the activities of other genes. How these expressions promote, inhibit, and otherwise interfere with the expressions of other genes is what makes up the architecture of a GRN.

Unsurprisingly given the number of interconnected transcription processes involved, what GRNs help illustrate is that there is no fixed set of genetic tracks that determines a cell's transitions. While some interactions can be identified as important in particular transitions, what decides the cell fate of a given cell is a highly complex, high-dimensional network involving thousands of interconnected genes (Huang et al 2005). What GRNs can show is which phenotypes are stable or unstable relative to their neighbours, and how these differences can drive transitions to new phenotypes.

GRNs are the first tool used in developing an explanation of cell fates. The second is Waddington's (1957) notion of epigenetic landscapes. Waddington's metaphor has proven particularly durable and appealing to many biologists concerned with cell fate phenomena, and the metaphor appears frequently in this scientific literature (e.g. Enver et al 2009; Davila-Velderrain et al 2015; Moris et al 2016). Adherents to this line of thinking equate progress through stable and unstable phenotypes to progress through the peaks and valleys of epigenetic landscapes, with the phenotype represented by a ball rolling through this terrain, like in Waddington's famous illustration.

This is where dynamical models enter the picture. In order to put Waddington's ideas about the epigenetic landscape into practice, researchers appeal to dynamical models as a way of capturing the trajectory of a cell through the many possible phenotypes it could express. The resulting models and attractor landscapes bear a striking resemblance to Waddington's landscape, which has as a result has been reconsidered from merely illustrative metaphor to something potentially more revealing about the workings of GRNs (Jaeger & Monk 2014).

Attractor landscapes are a frequently used visualisation of dynamical models, and are virtually ubiquitous in models of cell fates. To produce a 3D model that illustrates a cell's trajectory through different states (phenotypes), modellers reduce the many dimensions (genes) involved, of which there may be thousands, into a plane. Each point on that plane represents a possible state for the system to inhabit, and nearby states represent similar states. The relative height or depth of any given

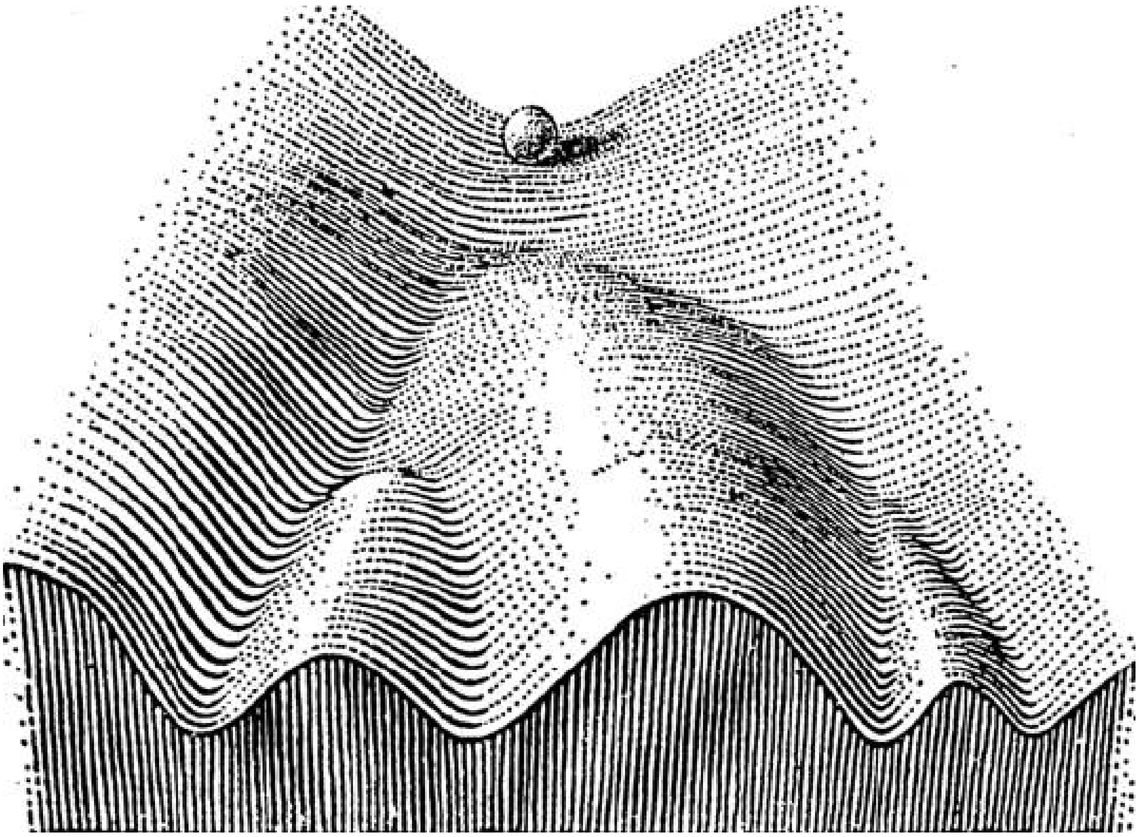

Figure 2.2: Waddingtons illustration of the epigenetic landscape. The ball (phenotype) runs down through the landscape and is canalised into different phenotypic outcomes. Reproduced from Waddington (1957).

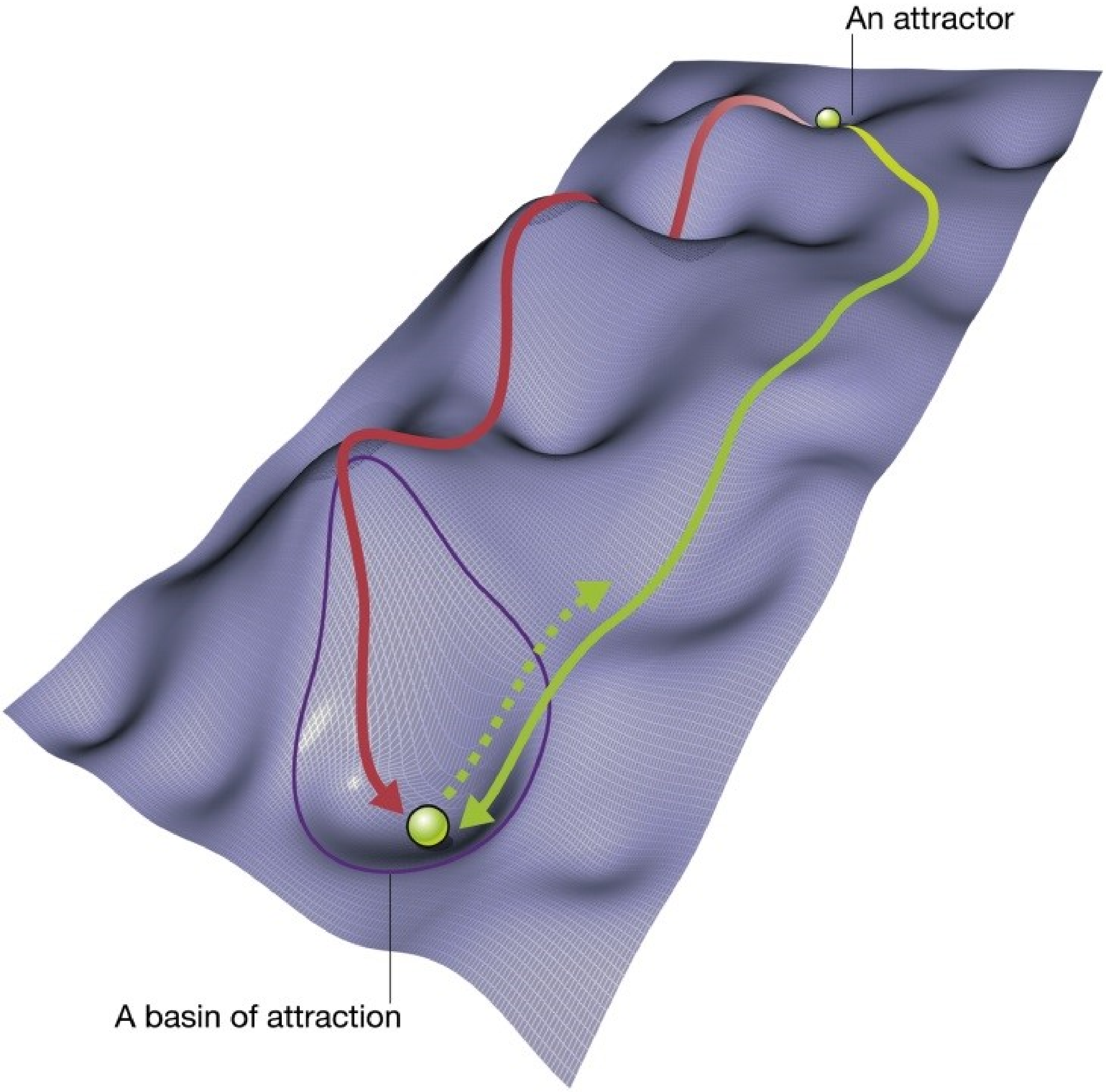

Figure 2.3: An attractor landscape. The green ball represents the phenotype, which has settled into one of several available basins of attraction. From Enver et al (2009).

point indicates its stability or lack thereof.

In dynamical systems theory an attractor is a point in this landscape that represents a stable solution to the equations that make up the model. Over time, the system will converge towards an attractor if it enters its basin of attraction. A basin of attraction is the set of points that "feed into" a given attractor. The convergence on an attractor might be stable, where the system settles right on the attractor point—or it may oscillate around that point ("circling the drain") for some time or even indefinitely.

In the attractor landscape these features are visualised as the troughs and valleys that Waddington's "ball rolling down the hill" follow and settle into. These attractor landscapes will, later in this section, be shown to provide the causal detail needed to describe and explain cell fates.

Biologically, the attractor point itself corresponds to a stable phenotype—a cell fate.

> Extrapolating from Waddington, different cell types may be seen as stable solutions of transcription factor networksor attractors'which occupy the basins of Waddington's landscape.(Graf & Enver 2009, pg. 590)

Figure 2.4 illustrates a toy example characteristic of many GRNs. There are two proteins being transcribed, a and b, each of which inhibits the transcription of the other—*mutual inhibition*—and encourage transcription of themselves in a positive feedback loop—*auto-stimulation*. These processes of mutual inhibition and auto-stimulation are common features of cells that are multi-stable.

There are also three attractors present. The first, a/b, is the stable starting point—the undifferentiated, progenitor phenotype where neither a nor b is transcribed at a high rate, and from which the system is unlikely to budge. If left undisturbed, the effects of mutual inhibition and auto-stimulation will generally ensure that transcription of a and b remains roughly even. Attractors a and b represent two "downhill" cell fates the perturbed system may end up in –if a/b were destabilised, the system will bifurcate, and converge on either a or b.

Graduating from a toy model, I turn now to a real-world example: Huang et al's (2007) model of FDCP-mix cells, a kind of bone-marrow cell. An FDCP-mix cell is capable of differentiating into two distinct cell fates called *erythroids* and *myeloids*. Differentiation into erythroid/myeloid is influenced heavily by two transcription factors, *GATA1* & *PU.1* respectively. GATA1 & PU.1 are both auto-stimulating, and mutually inhibiting. The following model describes the activation and regulation of GATA1 and PU.1:

$$\frac{dx_1}{dt} = a_1 \frac{x_1^n}{\theta_{a_1}^n + x_1^n} + b_1 \frac{\theta_{b_1}^n}{\theta_{b_1}^n + x_2^n} - k_1 x_1$$

$$\frac{dx_2}{dt} = a_2 \frac{x_2^n}{\theta_{a_2}^n + x_2^n} + b_2 \frac{\theta_{b_2}^n}{\theta_{b_2}^n + x_1^n} - k_2 x_2$$

Where x_1 represents GATA1 activity, x_2 represents PU.1 activity, and a_1 & a_2, b_1 & b_2, k_1 & k_2, and $\emptyset$, all represent control parameters. Parameters a_1/a_2 represent the relative strength of auto-stimulation of GATA1 and PU.1 respectively; b_1/b_2 describe the rate of mutual inhibition of GATA1 and PU.1; k_1/k_2 represent the rate of deactivation of GATA1 and PU.1; and $\emptyset$ represents the strength of the regulatory interaction. Much like the previous toy model, this real-world system exhibits tristability—C, the progenitor state, as well as A and B, the differentiated erythroid and myeloid cell fates.

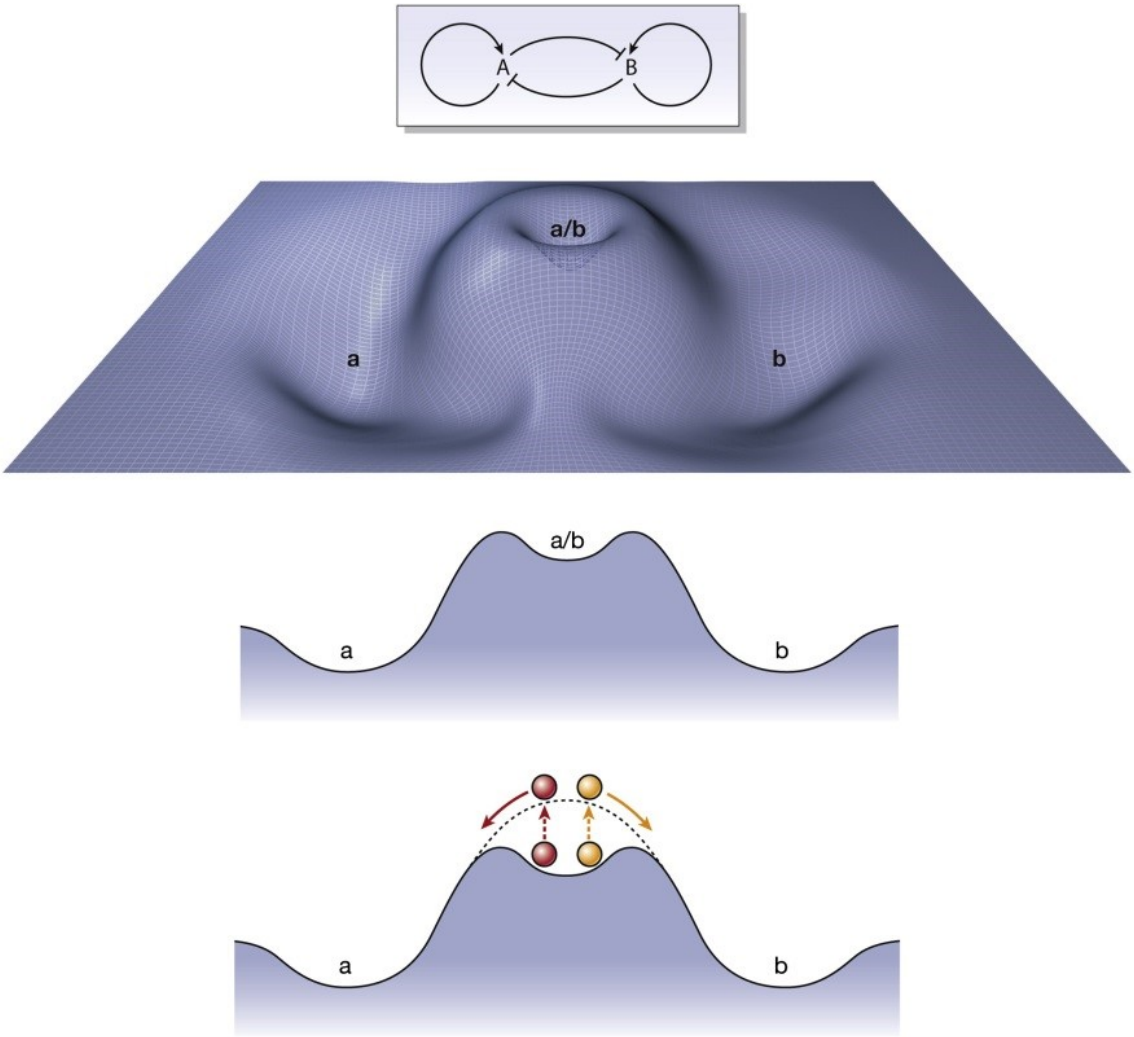

Figure 2.4: The attractor landscape of the toy cell fate system. The system is depicted bifurcating from a/b, the progenitor fate, into either the a or b fate. From Enver et al (2009).

By intervening on these parameters, Huang et al (2007) were able to alter the features of the attractor landscape. This is analogous to reshaping Waddington's landscape, where a change in the topography of that landscape—which states are stable and unstable relative to their neighbours—can induce a previously stable system to differentiate, and in this instance bifurcate into new "downhill" cell fates, an arrangement referred to as a "transition via a bifurcation" (Huang et al 2007, pg. 701).

Huang et al (2007) focus on the a, b and k parameters since these collectively represent the various regulatory influences on x_1 and x_2. The effects of mutual inhibition, auto-stimulation and deactivation over time of GATA1 and PU.1 are thought to maintain the stability of the system in its progenitor fate, C. Disturbing these variables, then, has the potential to destabilise C and induce differentiation into either A or B.

The model predicts that reducing the rate of auto-stimulation (a_1 & a_2) and deactivation (k_1 & k_2) will destabilise C and induce differentiation. The model also considers an alternative scenario where—due to different initial values in control parameters—the only attractor present in the model is C, the basin of which covers the entire phase space. In this scenario, reducing the value of b_1 & b_2 destabilises C, and also leads to the appearance of the A and B attractors. In either situation, C is converted from a stable state into an unstable "hill-top", from which any small stochastic variation in GRN activity is enough to induce a differentiation event. The system is compelled to leave C, and converge towards either A or B, the erythroid and myeloid cell fates respectively.

After this initial differentiation event where C can no longer be occupied by the system, the decision to converge on A versus B cell fates becomes available for the cell. Assuming absolute symmetry in the system (A and B are equally accessible from the now destabilised C) then minor variations in initial starting position, and random fluctuations in expression, will be responsible for pushing the system towards A or B. However, asymmetries in the attractor landscape can make one of A or B more accessible than the other and bias the system towards a particular cell fate despite the stochastic nature of the system's trajectory. For instance, a greater value of x_1 versus x_2 will bias the system towards settle into the myeloid state versus the erythroid fate, and vice versa. Huang et al (2007) describe this as "tilting the watershed" in order to "harness and bias the stochastic processes." (pg. 710)

Having considered the phenomenon described by Huang et al's (2007) model (initial differentiation followed by cell fate selection) the question of interest is: how or why do FDCP cells differentiate into erythroid or myeloid cell fates? Put into interventionist speak, we are interested in the *what–if–things–had–been–different question*, or *w–question* about cell fates: counterfactually, under what conditions would we have observed a different result (the selected fate)?

In interpreting this model, I argue that what makes the difference to the outcome, and what answers the w-question, is the attractor landscape. Specifically, I argue the presence or absence of particular attractors is decisive to the outcome for a cell in the process of selecting a fate.

As mentioned earlier, the initial destabilisation event occurs only once C has been intervened upon. If C is present (with sufficiently high ridges to prevent stochastic fluctuations from pushing the cell out of C) then the cell will not undergo a differentiation event. This is ultimately what answers the w-question: a different result would have been obtained depending on the presence or absence of the stable progenitor attractor state C. While the destabilisation of C can be induced via different stimuli (changes to the rate of auto-stimulation, mutual inhibition, and deactivation of GATA1 and PU.1) what makes the difference is C. An investigator intervenes on C through these control parameters. Further, no control parameter stands out as the singular cause of differentiation. In fact, most of the control parameters are, if intervened upon, capable of destabilising the stable attractor C under the right background conditions.

After the initial destabilisation of C, the cell needs to select a new cell fate to differentiate towards from this new unstable position:

> Metaphorically, the destabilization and disappearance of the progenitor attractor can be viewed as S being placed on a “watershed” region in Waddington’s epigenetic landscape (Waddington, 1957) where it can easily be “tipped” into either side to the now easily accessible attractors of the two prospective lineages by small, deterministic perturbations or by random fluctuations in molecular activities, to reliably produce distinct and specific outcomes. This near-symmetric bifurcation model thus is consistent with the ample evidence for the observed stochasticity in fate determination...(Huang et al 2007, pg. 709-710)

Huang et al (2007) are here comparing Waddington’s metaphor with their observed results: a symmetrical destabilisation event leaves the cell equally likely (all other things being equal) to differentiate towards A or B. The general stochastic variation in gene expression is enough to push the cell towards either one, as well as all manner of incidental factors:

> ...the “watershed” metaphor explains the observation that many unspecific (hence, non-instructive) signals, such as solvents or mechanical forces, can cause differentiation in many cell systems: they may do so by “tipping” cells into a predefined program. (Huang et al 2007, pg. 710)

From a position of high instability, it takes very little, up to and including incidental mechanical forces acting on the cell, to trigger differentiation. Small variations in initial conditions also have an influence, according to the model, on whether A or B is ultimately selected.

But what makes the difference outside this inherent stochasticity in the system is, as mentioned earlier, the biasing of these processes by shaping the attractor landscape. By intervening on control parameters, the investigators were able to "tilt the watershed", making A or B occupy more or less of the phase space with their basin of attraction. This is achieved through an asymmetrical intervention on the control parameters such that x_1 is greater than x_2 (or vice versa). This effectively makes one cell fate—myeloid or erythroid—more accessible from the cell's present state than the other.

The takeaway from this part of Huang et al's (2007) discussion is that there are many events than can induce the initial differentiation event (the destabilisation of C) and many events than can determine whether A or B is selected subsequently, ranging from the various control parameters to external forces. Whatever stimulus is involved, what makes the difference to the outcome is the specifics of the attractor landscape. How accessible a fate is from the cell's present state is what makes the outcome more likely amidst a barrage of stochastic fluctuations, and what makes a fate accessible are the dimensions of the associated attractor.

2.4 Dynamical Causes

An objection to the foregoing interpretation would be to question the role of attractors (and really any dynamical feature of a system) as the cause of differentiation in favour of a mechanistic interpretation. If we assume the initial destabilisation event has occurred in a cell, and the system is poised to differentiate to a myeloid or erythroid, perhaps differentiation can simply be explained by the levels of different transcription factors present. For instance, in Huang et al's model, one might prefer a mechanistic interpretation wherein intervening on levels of GATA1 and PU.1 in a cell leads to different outcomes. Higher levels of GATA1 and PU.1 in a cell do indeed predict a higher likelihood of those cells differentiating into erythroids and myeloids respectively (Huang et al 2007). On this reading it seems like the cause of differentiation is decidedly mechanistic. If this were the case, it would of course be fatal to the account being built here.

To air out this criticism, I will consider two possible readings: the mechanist one spelled out above, and the dynamical interpretation I advanced in the previous section, and place them into Woodward's (M) criterion:

> (M1) *GATA1 transcription level* causes *erythroid differentiation* if and only if there are background circumstances *B* such that if some (single) intervention that changes the value of *GATA1 transcription level* (and no other variable) were to occur in *B*, then *erythroid differentiation* would change.

> (M2) *Presence of erythroid attractor* causes erythroid differentiation if and only if there are background circumstances]emphB such that if some (single) intervention that changes the value of *presence of erythroid attractor* (and no other variable) were to occur in *B* then *erythroid differentiation* would change.

(M1) represents a 3M-compliant interpretation, where the causal relations in the dynamical model map to the underlying mechanistic model. (M2) advances the dynamicist argument—that the causal relations identified by the dynamical model are indeed genuine causes. Selecting which of these scenarios is the better interpretation requires some grappling with how exactly we might select between different potential causes.

Fortunately, Woodward (2010, 2018) has elaborated some criteria designed to clarify situations where the role of difference-maker is ambiguous. The ambiguity arises because not all potential causes are created equal—they "...can differ in the extent to which they satisfy other conditions relevant to their use in explanatory theorizing" (Woodward 2018, pg. 1). Two of these conditions developed by Woodward are I think particularly relevant to the current discussion—these are *specificity*, and *proportionality.*

Taking these conditions one at a time, I want to consider (M1) and (M2) first in terms of specificity. Specificity refers to "a kind of fine-grained and specific control" (Woodward 2010, pg. 306) that a cause has over the outcomes for some effect variable. To motivate the importance of this condition, Woodward draws on an example offered by Waters (2007) who discusses the synthesis of RNA molecules by DNA.

In this case, DNA provides genetic information to RNA polymerase, which then produce RNA molecules. The question is over whether DNA or the RNA polymerase is the cause of RNA molecule output:

> DNA is a specific difference maker in the sense that different changes in the sequence of nucleotides in DNA would change the linear sequence in RNA molecules in many different and very specific ways. RNA polymerase does not have this specificity...it is not the case that many different kinds of interventions on RNA polymerase would change the linear sequence in RNA molecules in many different and very specific ways.

> This shows that DNA is a causally specific potential difference maker. The fact that many such differences in DNA do actually exist and these differences actually explain the specific differences among RNA molecules indicates that DNA is the causally specific actual difference maker... (Waters 2007, pg. 574-575)

Of the two possible causes, RNA polymerase is the least specific. It does not provide a fine-grained description of the dependency between different states of the cause, and different states of the effect. Either the polymerase is doing its job, or it is not. This is comparatively coarse-grained when we want to understand the specifics of these dependencies.

On the other hand, DNA seems a far more satisfactory candidate cause. There is a fine-grained dependency between the different states of the DNA (what genetic information it inputs) and what RNA molecule is produced. Hence DNA is more specific and is the better candidate for the genuine cause of RNA molecule output.

Specificity requires that states of a cause map uniquely to states of an effect. The less "overlap" or lack of uniqueness in these mappings, the better. This mapping should describe exploitable counterfactual relationships between cause and effect in this fine-grained way. When selecting between candidate causes, especially in biology, the more specific a cause-effect mapping the more we should be inclined to select it as the genuine cause.

With specificity in mind, let us consider the viability of (M1). Does transcription of GATA1 exercise a fine grained and specific control over the outcome of cell differentiation? GATA1 does increase the probability that a destabilised cell will end up as an erythroid, so it exercises that much control. But that relationship does not tell us why a certain threshold of perturbation is required to initiate differentiation. Nor does it tell us why certain states of the system will differentiate into erythroids, and why others won't. Consequently, there is a considerable amount of overlap between many different values of differentiation and the consequent states of the cell fate. Consequently, the description here is fairly coarse-grained and non-specific.

On the contrary if we accept (M2), then we are delivered far more explanatorily relevant, fine-grained dependencies. The threshold required to initiate differentiation into an erythroid is now explained—it is due to specific stabilities of phenotypes compared to their neighbours, as described by the dynamical model. The fact that some states converge on the erythroid fate and not others is due to the dimensions of the basin of attraction corresponding to that fate.

The second condition to be examined, proportionality, "has to do with the extent to which a causal claim fully captures conditions under which variations in some phenomenon of interest occur." (Woodward 2018, pg. 1). Woodward provides an illustrative example: imagine I train a pigeon to peck at a red stimulus via classical conditioning. I present a new stimulus to the pigeon, and it pecks at it. Two possible causal claims can be introduced to describe what has just happened:

1. The presence of a scarlet stimulus caused the pigeon to peck.
2. The presence of a red stimulus caused the pigeon to peck.

1 is not untrue. Scarlet is a type of red, and on this basis, one could fairly claim that a scarlet stimulus caused the pigeon to peck. However there seems to be something off about this interpretation. Woodward identifies the flaw in that the "scarletness" of the stimulus is not what induced the pigeon's pecking, but rather its "redness". 2 is a more appropriate statement of cause and effect, since it is the redness or non-redness of the stimulus which is, based on the variety of conditions we could submit the pigeon to, the real "difference-maker". There are situations where the stimulus is not scarlet, yet the pigeon does indeed peck—when the stimulus is another shade of red.

Selecting the correct causal claim can be approached by the application of the notion of proportionality: causes should be proportional to their effects, meaning that a statement of a cause should not contain excessive detail, nor omit necessary detail (Woodward is here drawing on Yablo (1992)):

> (P) There is a pattern of systematic counterfactual dependence (with the dependence understood along interventionists lines) between different possible states of the cause and the different possible states of the effect, where this pattern of dependence at least approximates to the following ideal: the dependence (and the associated characterization of the cause) should be such that (a) it explicitly or implicitly conveys accurate information about the conditions under which alternative states of the effect will be realized *and* (b) it conveys only such information—that is, the cause is not characterized in such a way that alternative states of it fail to be associated with changes in the effect.

(Woodward 2010, pg. 298, emphasis in original).

So, the possible states of scarletness (scarlet or non-scarlet) are not depended upon by the possible states of pecking (pecking or not pecking). Conversely, possible states of the pigeon's pecking do depend on the possible states of the stimulus' redness. Example 2 provides the required "accurate information about the conditions under which alternative states of the effect will be realised". It also excludes scarletness since this does not provide said accurate information (also fulfilling (b)). Hence redness fulfils the criterion (P), while scarletness does not. This provides a good guide to the formulation of statements of causal relevance—example 2 is much preferred to 1, since it identifies cause and effect better by eliminating those details which violate (P).

Once again, I argue that (M2) comes out on top over (M1) when it comes to satisfying proportionality. The attractor landscape "*displays* or *exhibits* a pattern of dependence" (Woodward 2018, pg. 3, emphasis in original) between causes and effects in the way that a mechanistic story does not. The dynamical model displays and exhibits how the cell's phenotype depends on the stability of its current state, as well as previous and potential future states. The canalisation and perturbation of this dynamical system is therefore proportionate to changes in the cell's phenotype.

Meanwhile (M1) does not relay us accurate information on why certain states of the proposed cause—GATA1 transcription—lead to certain states of the system. Indeed there are a variety of situations where GATA1 transcription will not induce the erythroid fate and situations where (under the same background conditions) other events will induce the erythroid fate. The perturbation may be insufficient, the ridges surrounding the progenitor state too high, and so on.

On both counts—specificity and proportionality—it seems the best candidate cause is described by (M2). Indeed, this is exactly the kind of interpretation that seems prevalent in the scientific literature and a natural way of speaking about dynamical models of cell fates. For instance, Ferrell (2012) in a review of the role of Waddington's model in cell differentiation emphasises the ultimate role of the attractor landscape in determining fates:

> Although it was natural to assume that the induction stimulus acts by increasing the value of [a transcription factor] I could have alternatively made the stimulus act through any of the other parameters...Would this alter the conclusion that cell-fate commitment occurs as a result of the disappearance of a valley at a saddle-node bifurcation? The answer is no. No matter how I choose to have the inductive stimulus affect the model, the result is the same. (Ferrel 2012, pg. R461)

Whichever parameter one intervenes on—changing rates of auto-stimulation or mutual inhibition, of transcription levels, etc.—the causal import nevertheless lies with the dynamics of the system, with the attractor landscape. Conspicuous by its absence in these scientific discussions is much concern about how to render the causal story described by these dynamical models down into an underlying mechanism. Rather, in theoretical discussions cell fates are equated with attractors, such that changes to these attractors correspond to changes in cell fate phenomena (Enver et al 2009; Davila-Velderrain et al 2015; Huang 2012; Moris et al 2016). My point here is that the interpretation I offer fits comfortably with the way scientists talk about models of cell fates.

This is not to say that scientific talk, which naturally does not always map directly onto how philosophers of science or biology talk, is decisive to how we should formulate a mode of explanation. On the contrary, I wish to pre-empt the claim sometimes advanced by mechanists, apparently originating with Craver (2007), that the mechanistic mode of explanation most closely reflects how scientists think about nature and experiment on it, and hence has some especially pragmatic justification. A dynamicist account evidently aligns just as well with the scientific literature at least as far as cell fates are concerned.

2.5 The 3M Response

The anticipated response from mechanists is that even if attractors meet the interventionist criteria, no such model of cell fates can stand alone as a causal explanation. This is because—according to the 3M requirement—the dynamical model ought to be mapped to an underlying mechanistic model, which is what really provides the causal power and describes the causal structure of cell fates. In other words, 3M is a claim about reduction—causal claims about dynamical models reduce down to causal claims about the underlying mechanistic details. Claims about attractors acting as difference-makers are (if we accept 3M) reducible to claims about difference-makers within the GRN architecture.

However, Kaplan & Craver (2011) stress that 3M is not intended to be rigidly applied, and accordingly set up 3M inclusive of an assumption that the requirement can be defeated, with some caveats:

> Like all default stances, 3M is defeasible. However, those who would defease it must articulate why mechanistic styles of explanation are inappropriate, what non–mechanistic form of explanation is to replace it, and the standards by which such explanations are to be judged.(Kaplan & Craver 2011, pg. 603)

In taking up this challenge, the first point is likely to be the most contentious. Given the breadth of its application and ambitions, putting a hard limit on the reach of Mechanism is a difficult proposition. However, one uncontroversial starting point is this: Mechanism ends where it is unable to provide explanations. As we have seen, comprehensively uncovering the causal structure underlying a phenomenon (or at least, being in the process of doing so) is a requirement for explanation for mechanists. As a consequence, if mechanistic models can't get at causal structure, it follows that this would indicate a situation where mechanistic explanation is inappropriate.

Cell fates are an example of the causal structure underlying a phenomenon being accessible to dynamical models. On the other hand, it is inaccessible to mechanistic models. This is because the difference maker is part of the system's dynamics—an attractor—that can only be described in the context of a dynamical model. For example, the progenitor cell fate which makes the difference to the initiation of differentiation "...is "dynamically" defined, namely, as a metastable state in between two neighboring attractors of the prospective differentiated states..." (Huang et al 2007, pg. 699). Hence a description of the mechanism underlying cell fates (the proteins involved, the GRN architecture, etc.) will not get at the difference maker to the cell fate, and consequently there is no clear mechanistic analogue of this attractor, no entity or collection of entities that corresponds coherently to this dynamical feature.

To entertain the notion, persisting with 3M in this case would presumably involve trying to locate the physical components that underlie abstract dynamical features like attractors. This seems like a difficult proposition, since these features don't appear to have a direct relationship to any component of the system—attractors and basins of attraction result from a network of dynamical activity involving many thousands of working parts. To pursue this option necessitates a description of a brute amalgamation of all the physical components associated with these features of the dynamical model—the thousands of genes, the transcription of proteins, etc.

But in this scenario all of the explanatory work would still be done by the dynamical model, since it provides the relevant details about causal relations/difference makers. Identifying physical features associated with a model doesn't entail those features or components are necessarily providing causal detail. It's hard to see how a model like this could be anything other than a bona fide dynamical explanation, since all the explanatory work would be done by the dynamical model.

If anything, the idea of adding in excess mechanistic detail seems to run directly against mechanist's own standards of explanation, as well as the general rationale for these standards. Craver & Kaplan (2018) for instance discuss the notion of completeness as a benchmark of good causal explanation, where a description includes all the causally and constitutively relevant features of the world. Equally critical is the exclusion of irrelevant detail, keeping out those features of the world that are neither causally nor constitutively relevant to the phenomenon.

It is, on the mechanist account, counter to the purposes of a good causal explanation to include features of the world that do not contribute to the unveiling of causal structure. Hence if the addition of further mechanistic detail contributes nothing to understanding the causal structure of cell fates, then it seems that by mechanists' own standards of explanatory completeness 3M is, in this case, counterproductive.

The second requirement for defeating 3M—articulating a non–mechanistic account of explanation—can be dispensed with here more easily. Both Meyer (2018) and van Eck (2018) provide interventionist-based accounts of dynamical explanation—the former outlined earlier in Section 2—that share many (if not most) of the assumptions about explanation that motivate the mechanist account. This kind of explanation follows the interventionist method of locating difference-makers and developing counterfactual-supporting invariant generalisations. Further details require significant fleshing out, but these accounts at least offer a foundation to build on.

The third point—what kind of standards would dynamical explanation be judged by—provides the most scope for interpretation. I take this reference to standards to mean, roughly, what is required for a mere description to transition into an explanation. Much like mechanistic explanation, the development of counterfactual-supporting invariant generalisations is the basis of providing causal explanations. Similarly, an appeal to the accounts of dynamical explanation offered by Meyer (2018) and van Eck (2018) ought to be considered at least a good starting point for this project.

An important aside here concerns the relationship between mechanistic and dynamical explanations. So far this paper has not grappled with problems of integration or compatibility explicitly. I would argue that the specific characterisation of the phenomenon—the explanatory question being asked—will play a large role in how these modes of explanation hang together. A question specifically about the difference-maker to cell fates appears to require what van Eck (2018) calls a "pure dynamical model", one that is non–mechanistic. We are really only interested in the dynamics causing one outcome to obtain over another, and include mechanistic details (the GRN) only as background conditions. All the relevant causal detail is contained in the dynamical model.

I leave open the question of a combination of the two, which seems a substantial discussion requiring its own treatment. I have only discussed one example from one subgenre of science, and hence it is plausible that some questions require a different model or combination of models to fully encompass the phenomenon.

2.6 Conclusion

In this paper I have argued that non–mechanistic, causal dynamical explanations are viable. I have provided reasons, rooted in Woodward's interventionist account, for thinking that dynamical models do in fact describe the causal structure of the world independently of mechanisms, using the example of cell fates. I have argued that dynamical models can describe more specific and proportional causes than mechanistic models of the same phenomena, and hence are the better descriptors of causal structure in these cases. I have also defended against the 3M criterion, showing how in the case of cell fates, 3M need not be observed.

Chapter 3

Does Mechanism Come to Bury E–Cognition, Or to Integrate It?

This chapter evaluates the possible integration of E–cognition and traditional cognitive science via Mechanism. I focus on Miłkowski et al's (2018) account, which claims that all existing theories of cognition are trapped in insoluble theoretical quagmires. They propose jettisoning theories from cognitive science and reviving them as heuristics to guide mechanistic explanations. Under this *mechanistic–heuristic account*, they argue that enactivist and cognitivist accounts are fully integrated. I address some serious problems with this account. First, this characterisation of theoretical quagmires is not reflected in the cognitive science literature, undermining the supposed benefits of integration. Second, differing theories of cognition demonstrate ongoing empirical significance. Third, Miłkowski et al's (2018) notion of heuristic is ambiguously defined, and unevenly applied. Fourth, their integration fails to address the contradictory commitments of enactivists and cognitivists, instead avoiding or misrepresenting theoretical differences. Finally, I criticise the authors' attempts to head off dynamical explanation via an excessively permissive redefinition of the notion of a mechanism. I conclude that Miłkowski et al's (2018) mechanistic–heuristic account does not genuinely integrate traditional cognitive science and enactivism, better resembling an attempt to bury enactivism alive and clear the way for an unreconstructed cognitivism.

3.1 Introduction

So far in this thesis I have argued that dynamical explanations are genuinely explanatory. The motivation for making this argument, and showing how it applies across the special sciences, comes in part from the consequences this debate about explanation has for the stocks of competing accounts of cognition. Some consider mechanistic explanations to offer some kind of integrative solution to different theoretical commitments: mechanistic explanations, with their toolkit of interventions and mutual manipulability can perhaps neutrally cut through some of the expository differences that are held to by those on either side of various theoretical schisms (Kaplan 2012). Mechanism might be able to get traditional cognitivism and the still–emerging E–cognition accounts to speak the same mechanistic language; to explain by describing mechanisms, regardless of the overarching theory motivating an investigation.

In the eyes of mainstream cognitive science, explanation is the soft underbelly of E–cognition accounts. In this realm, cognitivist accounts are generally seen to occupy the stronger position versus their radically embodied rivals. This is primarily because of the congruence and good working relationship between cognitivism and mechanism (Bechtel 2008). Mechanism is both well–established as a mode of explanation, and is also seen to have a close relationship with a computational and representational account of the mind since "...questions at the computational level can be construed as questions about what a mechanism is doing and *why*...Questions at the algorithmic level...are questions about *how* a mechanism does what it does..." (Zednik 2018, pg. 390, emphases original). The whats, whys and hows of cognitivism are mutually intelligible with those of mechanism, making them natural allies.

Meanwhile enactivists, arguably the least cognitivist–friendly variety of E–cognition, see explanation of cognitive phenomena as often involving an escape from these mechanistic details. In addition, enactivists explicitly deny the necessity of invoking representations and computation when explaining cognition (Varela et al 1991; Chemero 2009; Di Paolo et al 2017). Explanations should appeal to patterns of engagement between organism and environment, with internal, brain–based activity no longer serving as the only locus of many explanations. This proposal can be investigated in a variety of ways—dynamical systems theory is one go–to method, coupled with a non–mechanistic perspective on explanation (Stepp et al 2011; Silberstein & Chemero 2013).

And so scientific explanation has become the location of a proxy argument backed by cognitivist–aligned philosophers of cognitive science on the one side, and various E–cognition advocates and dynamicists on the other. The hope is that by getting purchase on the coveted status of explanation (perhaps even exclusive access), either cognitivism or E–cognition will have a powerful argument in their own favour that will undermine their opponents: our account explains cognition better.

On the other hand, there remains the possibility of integration in place of competition. If cognitivist and E–accounts could be combined into one overarching, unified account of cognition, then the entire disagreement would be resolved. The result would hopefully be an account that does justice to both traditional cognitive science while integrating insights from E–cognition. The task does not seem easy, considering the often diametrically opposed positions some E– theorists and cognitivists hold regarding the explanatory role of representations and computations, and the constitutive role of the body in cognition. Wading into this situation, Miłkowski et al (2018) (from here on referred to as M) nevertheless make a bold proposal for just such an integration. M claim that the problem of duelling theories can indeed be put to bed, not least because in their view "the controversies over these wide perspectives have become outdated" (pg. 1). Overly concerned with squabbles over theory, E–accounts and traditional cognitivism have, according to M, missed a trick: integration via a common mode of explanation. solution is a fully unified cognitive science guided by an overarching explanatory framework that cuts through existing theoretical disagreements.

To this end M recommend that mechanistic explanation take up the mantle of unifier. Through Mechanism, M argue, we can build the one integrated approach to cognition to supersede them all. Their formulation is what I label the *mechanistic–heuristic account*. This solution involves first of all a blanket rejection of all existing theories of cognition. Instead of developing and testing theories, researchers should aim towards integrating elements of E– and cognitivist theories of cognition as helpful *heuristics* for mechanistic explanations. Mechanistic explanation serves as the integrative force, an explanatory framework that will cut down to the real facts of the matter behind the theoretical gloss. Mechanism helps accounts of cognition to produce grander, better explanations in the face of a disunified schism.

However, I will argue in this chapter that M's efforts do not succeed in making the case for integration, nor do they spell the end of an un–integrated E–cognition. Focussing on enactivism as the most illustrative case, I show how M's account its not successful in delivering integration. Their account, I will show, tries to produce integration by variously avoiding or misrepresenting the substantial roadblock to integration: the incompatible commitments of enactive versus cognitivist accounts.

Section 2 outlines M's proposal for a mechanistic–heuristic approach to cognitive science, and its account of integration. In section 3 I take issue with M's characterisation of inter– and intra–theoretical debates as counterproductive, showing how theory in cognitive science heavily influences the formulation of explanations. Section 4 grapples with the notion of heuristic deployed by M comparing it to Bechtel & Richardson's (1993) heuristics of decomposition and localisation, arguing that the authors instead define heuristics such that only cognitivism alone retains its theoretical import. Section 5 evaluates M's vision of integration. I compare Segundo–Ortin et al's (2018) resolution of theoretical disagreements between enactivism and ecological psychology, permitting integration. I argue that M instead misrepresent the commitments of enactivism through their uneven interpretation of heuristics, and therefore only succeed in integrating a straw version of enactivism. Finally, section 6 criticises M's redefinition of the concept of a mechanism in order to head off non–mechanistic, dynamical explanations, showing how it is overly permissive.

3.2 The Mechanistic–Heuristic Account

3.2.1 Theoretical Quagmires

M initially motivate their account by pointing out what they see as a broad and pervasive issue in cognitive science that they aim to resolve: the intractability of questions about cognition that lead to disputes over abstract dichotomies. They have two kinds of dispute in mind, which I will call *inter–theoretic*—between theories of cognition i.e. cognitivism versus enactivism—and *intra–theoretic*—within those theories of cognition. Both, they think, are trouble.

M first list some intra–theoretic disputes that early cognitive psychology has previously been embroiled in: "nature vs. nurture, continuous vs. all–or–none learning, serial vs. parallel processing, analog vs. digital, conscious vs. unconscious, grammars vs. associations for language, etc." (pg. 2) Ultimately these debates, M claim, are a dead–end. M's reasoning here is that questions like "nature vs. nurture" are too abstract to properly guide, or really be decided by empirical observations; if anything, these dichotomies hamper attempts to build unified theories of cognition because they redirect effort to useless investigation: "[g]rand issues in the study of cognition cannot be fruitfully understood in terms of a series of simple dichotomies." (pg. 2) Based on these historical cases, they suppose that similar debates must also be ultimately fruitless at building unity.

They go on to accuse E–theorists of also being stuck in intra–theoretic disputes, and portend a dead–end for E–cognition where "at least some researchers attempt to win the debate by showing that the bodily, the environmental, or the interactive aspect is most essential in cognitive functioning" (pg. 2). Here M attempt to draw a close analogy between historical intra–cognitivist debates, and modern intra–E debates, characterising them both as futile.

Inter–theoretic disputes between cognitivism and enactivism are similarly characterised as theoretical quagmires. Enactivists and cognitivists have long been embroiled in the so–called "representation wars" over the status and necessity of internal, content–bearing representational states for explanations of cognition. This sharp divide between proponents of an enactive anti–representationalism (i.e. Chemero 2009, Hutto & Myin 2013), and a pro–representation cognitivism continues to be contentious. The efforts of enactive cognitive science have often centred on the goal of building models and explanations of cognition without appealing to internal representations and processing thereof (i.e. Di Paolo et al 2017).

But in M's view, while enactivists do propose a significantly different *sounding* theory of cognition, the differences are in their view essentially expository. Whichever way the argument turns out—for instance, the necessity of representations for explanations of cognition—M think it won't change anything meaningful about how we actually investigate cognition. Moreover, the problem of quagmires strikes again: "theoretical presentations of these approaches remain fairly abstract and focused on deciding yes-no questions rather than building unified models of cognitive phenomena." (pg. 2)

Through this framing M suggest a couple of important things about differing theories of cognition. First, they suppose intra–theoretic and inter–theoretic disputes are intractable. Second, they propose that these disputes are not meaningful or worthwhile because they only work to conceal a deeper, more desirable unity of approaches.

3.2.2 Mechanistic explanation to the rescue

The solution to these perceived theoretical doldrums, M argue, is mechanistic explanation (Machamer et al 2000; Craver 2007; Glennan 2017). Mechanistic explanations are capable of unifying cognitive science behind a single mode of explanation and in doing so provide it with a way out of theoretical quagmires, getting beyond theory–squabbling. Embracing mechanism rather than some other approach is desirable for several reasons in M's view.

First, mechanisms span both "wide" and internal components. On this reading of enactivism, the brain–body–environment relationships proposed to be constitutive of cognition by enactivists can be cashed out simply as spatially distributed mechanisms (an avenue also explored by Abramova & Slors (2019) and discussed at length in Chapter 4). They are therefore just as tractable to mechanistic modelling as more traditional mechanistic components. Mechanism thereby manages to straddle cognitivism at one extreme and enactivism at the other—it can produce explanations integrating internal and external components, and hence covers more ground than any of those accounts alone.

Second, the mechanist account is very accommodating to the empirical efforts of both cognitivist and E–approaches, since it is “extremely lean”, leaving it open to describing both wide and narrow mechanisms. The accommodating features of mechanistic explanations allow them to “...offer an integrated view on cognition...” (pg. 2) and work to unify disparate lines of research, like mainstream cognitivism and E–cognition.

Third, mechanism’s framework for identifying causal relationships is useful for avoiding theoretical disputes. Since we can intervene on variables to find those that are causally and constitutively relevant to the phenomenon (Craver 2007), we can just establish what items are causally efficacious rather than speculate. Concerns over what items should figure in our explanations—representations and computations being two salient cases—become irrelevant once we can use interventions to find causal relations.

Having established their case for Mechanism as a solution to inter– and intra–theoretic debates, M now put it to work as a replacement for theories of cognition. M’s strategy is to first smooth over the disputes between the various theories of cognition. To this end they claim that “[d]ifferences between approaches matter only for expository purposes but not really for their practice, which involves mechanistic modeling of cognition” (pg. 4). Perched on the heights of Mechanism, one can, M think, see how cognitivism and enactivism are unified in a more cohesive mechanistic picture of cognition, even when the gloss on the phenomenon might be different from either perspective.

Further, enactivism can’t hope to compete with mechanistic explanation, because individual theories of cognition “...are not poised to be complete and exclusive accounts of cognition. They are not theories in the sense of providing complete predictions or explanations of phenomena in question. For this, they are too abstract.” (pg. 4) M do not specify the boundary of abstract and too abstract, or what sorts of features a spectrum of abstractness might take into account. We can however infer from their statements regarding prediction and explanation, as well as hypothesis generation, that M believe that genuine theories are concrete (versus abstract) enough when they generate testable hypotheses. E–account are all meant to fail in meeting this criterion because they “offer mainly abstract heuristics that cannot do much explanatory work in isolation” (pg. 1).

These claims culminate in the position that mechanistic explanation supersedes the need for theories of cognition. Mechanisms span wide phenomena and are a guide to discovery of causal relations. Empirical success trumps theory, and so mechanism can simultaneously explain broadly across cognitive science and supersede existing theories. Mechanistic explanation supersedes said theories (which fail to even be proper theories) functioning as a unifying or integrating force for explanations of cognition.

3.2.3 Heuristics & Integration

As it stands, there is potentially a large theory–shaped gap in M's account. While mechanistic explanation is meant to fill in this gap, M are rightly concerned about losing the insights gained by traditional cognitive science and E–cognition, despite their perceived inadequacies as standalone theories. Accordingly, M offer an additional component to get their account off the ground: they suggest that bringing theories of cognitive science back as *heuristics* will provide an additional guide for good mechanistic explanation.

But how are heuristics different to theories and what kind of work can they do for cognitive science and explanation? M think the answer is already in front of us in the practices of cognitive science. They claim that in the case of E–accounts, "[r]esearchers appeal to wide factors merely as discovery heuristics." (pg. 13) Likewise over on the cognitivist side, traditional computationalism in the vein of Marr (1982) for example only provides us with:

> ...certain *guiding heuristics*. A proponent of traditional computational modeling would ask what the overall task is and why solving it is appropriate; what the algorithms and representations involved are; and how they are physically implemented.(pg. 4, emphasis added)

Both E–accounts (including enactivism) and cognitivism are already, M think, used merely as heuristics for mechanistic explanations by researchers. This mechanistic–heuristic account is what M think ought to fill in the gap left by theories, and avoid theoretical quagmires. To illustrate their proposal in action, they provide the example of the operation of an aircraft by a pair of pilots taken from Hutchins (1995). The idea is that the operations by the pilots in the cockpit of a large passenger aircraft involve more than just the internal cognitive resources of the individual pilots, but also distributed technological and socio–cultural features of the entire cockpit system (Hutchins 1995); the situation is, for M, ripe for a heuristic division of labour:

> The process is distributed and includes two pilots as its component submechanisms...But the distributed approach may well appeal to heuristics preferred by other wide approaches to the study of cognition...the enactive perspective will stress the importance of dynamic coordination between pilots and consider how the environment is structured in terms of various affordances. At the same time, the distributed approach does not screen off the study of representational devices... (pg. 7)

The idea is that, by their powers combined, the various accounts of cognition—now working as guiding heuristics—will remind researchers of important considerations. Enactivism, for instance, functions within a mechanistic–heuristic account as a rejoinder to scientists not to forget about the interactions between the pilots. It guides the mechanistic explanation more efficiently towards locating all the relevant causal parts. Cognitivism reminds investigators to look for the computational and representational features of the system. Thus, a complementary process: mechanism gives us a framework for finding causally relevant features of the world for the explanandum phenomenon, while heuristics guide investigations towards likely candidate causes. Integration is thereby achieved.

3.3 Theories

Much is made in the above account by M of the inadequacy of theories of cognition for doing the real work of guiding cognitive science. They make three main claims about theories of cognition to this effect: that they are stuck in quagmires which hinder research and undermine unity; that theories merely serve as a gloss that fail to influence research anyway; that mechanistic explanation can take over the role of theories anyway. I will present some criticism of each claim in turn.

3.3.1 Theoretical quagmires?

M claim that theories are trouble: where there are theories, there are intractable debates. One of their targets are inter–theoretic debates between cognitivists and enactivists, though I will argue their characterisation is misleading. I contend that inter–theoretic debates are arguably quite productive, especially in terms of developing alternative theories of cognition that aim to solve ongoing theoretical issues plaguing mainstream cognitive science. Enactive accounts have historically been largely motivated by what they see as serious deficiencies in the mainstream view of cognition:

> Cognitivists struggled to model or explain the flexible, context–sensitive and domain–general intelligence that is characteristic of human cognition. Intuitively basic cognitive capacities like motor control and perceptual recognition seemed particularly resistant to cognitivist efforts.” (Ward et al 2017, pg. 366)

Early autonomous robotics research (Brooks 1990, Beer 1991) demonstrated these difficulties in action, especially the inadequacies of a strict symbol–manipulating intelligence for dealing with the complexities of real–world bodily interactions in coordination with a changing environment. Further, the reliance on computation and representation as core to understanding cognition invites "...the problems of homuncularity, the absence of the body, and the threatening infinite regress of perspectives that are rule–based at the pragmatic level and therefore can never account for meaning–generating processes." (De Jaeger & Di Paolo 2007, pg. 486). These concerns are certainly not new, harking back to criticisms levelled by Dreyfus and Searle, suggesting that even a completed cognitivist account of cognition would not get us beyond a how–possibly model alienated from actual cognition; "how a cognitive problem *should* be solved if it was in the hands of a clever computer programmer." (Di Paolo et al 2017, emphasis original)

Accordingly enactivists frequently present their account in terms of its opposition to the cognitivist mainstream, as "a move from an abstract conception of mental activity toward more concrete and situated, action–based practices and capabilities" (Di Paolo et al 2017), which is intended to be a productive step forward; at the very least a spirited attempt to explore and flesh out theoretical alternatives that escape the above difficulties. Enactivism services the "need to develop theories that can overcome well known problems encountered when attempting to understand and model the fluid and plastic nature of cognition" (Hutto & Myin 2017, pg. 1) In short, enactivism has been geared since its inception towards formulating different questions about cognition in the hope of making the intractable tractable.

Quite the opposite, then, to M's picture of a stagnant and intractable dispute, at least as far as enactivists themselves are concerned. Even some cognitivist–oriented philosophers (or at least those who wish to preserve some version of representationalism) who are critical of the more radical proposals associated with enactive and other E–accounts have been compelled to give credit where it is due. It is now the case that "E is the letter, if not the word, in today's sciences of the mind" and these E–approaches "are now a staple feature of the cognitive science landscape" (Hutto & Myin 2017, pg. 1). This assessment is supported by the recent profusion of E–oriented empirical and theoretical work, and an embodied perspective on cognition is, minimally, seen as worth looking into (Goldman 2012, Aizawa 2015).

It seems at least equally plausible, then, to say that the back–and–forth between mainstream cognitivists and their enactivist adversaries has been productive, rather than a problem for cognitive science. Unless one's goal is the triumph of an unreconstructed cognitivism at all costs, exploring alternative theories of cognition hardly seems threatening. Off the back of these supposedly intractable debates enactivists and other E–cognition theorists have managed to build alternative accounts of cognition that attempt to solve entrenched problems. Since enactive theory has manifestly worked as a catalyst for the development and finessing of theories of cognition, it would be misguided to conflate ongoing disagreement with stagnation.

M also target intra–theoretic debates within E–cognition. However, they provide no examples of this effect within the E–cognition research tradition despite asserting their existence. So, to provide a hard target on M's behalf, I will discuss two debates: first, the intra–E debate over the core notion of embodiment (Kiverstein 2012), and then intra–enactivist debates over the explanatory role of representations. As I will demonstrate, turf wars—M's characterisation—are not the sticking point. Intra–E disputes are almost uniformly about how to move away from the aforementioned intractable cognitivist problems.

Embodiment can be seen as a bone of contention between different strands of E–cognition. Kiverstein (2012) distinguishes between enactive accounts of embodiment, and the more conservative "body functionalism" (corresponding roughly to extended/weak embodied accounts of cognition). The latter, originating with Clark (1997), preserves the central role of computation, representation and similar cognitivist staples in its attempts to explain cognition generally while admitting a functional role for embodiment. In this way it attempts to fix some of the aforementioned problems with cognitivism, getting the body and its meaning–generating capacities into the explanatory picture while also maintaining some of the commitments of an internalist account.

Enactivism, by contrast, is happy to mostly shed these commitments and produce new alternatives without the cognitivist baggage. What differentiates these accounts is not, as M claim, an intractable dichotomy about which slice of the pie chart belongs to body, environment and interaction respectively, or even whether cognition is at all embodied or not. It is about the willingness of E–approaches to reappraise and avoid questions that lead to intractability, and how much is to be salvaged: should we embrace "root–and–branches reform of the cognitive sciences" or "out–and–out revolution"? (Kiverstein 2012, pg. 742) Enactivists generally prefer the latter:

> The alternatives range from conservative models that remain close to cognitivist conceptions of the mind, to more moderate and radical camps that argue we need to rethink our basic assumptions about the way the brain and the mind work. Most recent debates have been focused on the pragmatic and action–oriented perspectives of ecological, enactive and extended conceptions, which either minimize reliance on the notion of representation or eschew it altogether..." (Gallagher 2018, pg. 354)

These are difficult questions to answer, so it is not shocking that different theorists argue for different, often competing, positions. But the upshot is that the different strands of E–cognition are in large part differentiated from one another by how much they distance themselves from high cognitivism. Enactivism sits on the extreme end of the spectrum, rejecting most commitments and arguing for a fresh start, while “weak” E–cognition retains much of the cognitivist framework.

Likewise, when enactivists of different schools of thought come to a disagreement the substantial differences concern a thoroughgoing clean–up of cognitivist artefacts. For instance, some iterations of sensorimotor enactivism (O’Regan & Noë 2001) have appealed to a notion of “knowledge” to explain how agents guide their perception and actions. For analytic enactivists (Hutto 2005) the invocation of knowledge seems too close to a representational conception of perception and agency; an example of “conservative thinking.” (Hutto 2005, pg. 389)

> Although their account is advertised as decidedly skill–based’, on close inspection it shows itself to be riddled with suppositions threatening to reduce it to a rules–and–representations approach. *To remain properly enactivist it must be purged of such commitments.*(Hutto 2005, pg. 389, emphasis added)

Far from being stuck in pointless intra–theoretical quagmires, enactivism is evidently committed to fixing problems by producing genuine alternatives. The charge that these debates are counterproductive somehow therefore doesn’t ring true nor does the analogy drawn between the stagnation of cognitivist and enactive theory. M’s characterisation of these sorts of discussions as turf wars seriously misrepresents them, and their argument relies on drawing a flimsy analogy between cognitivism and enactivism, such that a failure of cognitivism to resolve internal problems equates to a failure of enactivism to do the same.

The second point M make, that a deeper unity is concealed by these debates, I will return to in section 5.

3.3.2 Do theories matter?

So far, the notion of a theory has been bandied about quite a bit. At the core of M’s mechanistic approach is the rejection of all existing theories of cognition *as theories*, cognitivist and enactive alike. They have argued that so–called theories of cognition fail to qualify as such. In addition, theories are meant to be trouble because they just lead to debates that go nowhere. Whatever job they were doing before, mechanistic explanation can safely take over the reins.

Responding to these claims requires some consideration of the role of theories in cognitive science, and whether M’s characterisation really tracks this role. What is meant by theory on the part of M is somewhat vague, and hence why they think a given account succeeds or fails as a theory is equally vague. To provide some clarity on the matter, M’s fellow mechanist Craver (2008) has this to say about scientific theories:

> *A central aim of science is to develop theories* that exhibit patterns in a domain of phenomena. Scientists use theories to control, describe, design, explain, explore, organize, and predict the items in that domain. Mastering a field of science requires understanding its theories, and many contributions to science are evaluated by their implications for constructing, testing, and revising theories. *Understanding scientific theories is prerequisite for understanding science.*(Craver 2008, pg. 55, emphases added)

If we take Craver's view into account, then it seems uncontroversial to say theories are a fairly important feature of good science, cognitive or otherwise. Theories make sense of empirical data and generate further hypotheses that can be put to the test empirically. Hence to develop an idea of what empirical work to do—what variables to intervene on, which items to appeal to in hypothesising and explaining – there need to be theories of cognition.

In contrast to Craver's (2002) conception, M claim that theories of cognition fail to provide more than expository gloss for cognitive science. Bechtel (2016) discusses this exact claim at length. Some critics argue, much like M, that representationalism is a mere gloss within neuroscience—the real business of science can easily do with or without representational nomenclature without losing anything of empirical consequence. This is more or less M's position on all theoretical posits on both sides of the aisle, cognitivist and enactivist alike. Bechtel however begs to differ, arguing that representations are important for the development of theory, advancement of hypotheses and the conduct of research.

To this end Bechtel (2016) discusses the example of place cells, neurons that are hypothesised to play a role in allowing organisms to navigate their environment. Bechtel traces the extensive empirical research that has been investigating this phenomenon for decades, and finds a tight fit between the mutual development of empirical work and the role of representations in theory:

> Characterizing neural processes as representations is not viewed as just a convenient way of talking about brain processes. The research is predicated on these processes being representations; the explanatory tasks they devote themselves to are identifying those neural processes that are representations, figuring out what their content is, and how these representations are then used in controlling behavior." (Bechtel 2016, pg. 1293)

The revision of theory, and the development and proposal of new hypotheses to guide empirical research similarly depends on this understanding of representations as a central explanans of how place cells figure in behaviour:

> ...such additional inquiry is inspired and guided by the initial attributions of representational content and directed at fleshing out the account. The attribution of content is a first step in articulating an account of a mechanism for processing information. Without this initial assignment of representational content, researchers would not be able to formulate the hypotheses that guide subsequent research." (Bechtel 2016, pg. 1291)

Bechtel's (2016) puts paid to M's idea that theory in cognitive science is already defunct. And over in the enactivist camp, there is agreement on the significance of theoretical posits like representations for empirical work and explanation:

> [Bechtel] makes a compelling case that it is only because these scientists conceive of their quarry in representational terms that they have been able to get such a seemingly powerful and impressive grip on what appear to be the working parts of an important mechanism of cognition." (Hutto & Myin 2017, pg. 242)

In fact, the significance of theory for empirical work is written into the very fabric of the cognitivist project. Marr's *Vision* (1982), arguably the cornerstone of the cognitivist outlook on cognition, sets up an overarching theory of visual perception that specifies how empirical work ought to investigate three distinct but interrelated layers or levels of cognition. These are the *computational* level, the level of *representation and algorithm*, and the *implementation* level. The computational level investigates the computational processes underlying cognition, their goals and appropriateness to the task; the representational/algorithmic level concerns how this computation would be implemented to process representational inputs and outputs; and the implementation level concerns how the physical processes of the nervous system actually realise the former levels.

Marr's account importantly lays out the stable of items in the field of cognition that descriptions and explanations ought to appeal to: computations, representations, algorithms, and the neural realisers thereof. Marr, like Bechtel (2016), puts a great emphasis on the central role of representations in producing explanations of cognitive phenomena:

> ...if we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information...The study of vision must therefore include...*an inquiry into the nature of the internal representations by which we capture this information* and thus make it available as a basis for decisions about our thoughts and actions. This...will *profoundly shape our investigation of the particular problems posed by vision.*" (Marr 1982, pg. 3, emphases added)

For Marr, the notion of representation is decisive to how we investigate cognition. Cognitive scientists adhering to Marr's theory (or modern iterations of it) are not poking around in the dark: evidently the explanations cognitive scientists will look for if they subscribe to Marr are characterised by looking for certain items (representations and processing thereof) in certain places (the central nervous system). They will hypothesise about the role of these items and not others in order to explain phenomena.

In fact, Miłkowski (2013) provides just such an instance of Marr's guidance while discussing the course of an explanation for a navigational task performed by a rat:

> We begin by describing in appropriate detail the rat's navigational task (level 1), specify, among other things, that the only source of information available to the rat is its own movements...Yet we also look for constraints at the implementational level (the neural level) to discover the regularities at the level of representations and algorithm. Already the bottom level contains representations, but Marr does not disallow that.(Miłkowski 2013, pg. 116)

We can see how Miłkowski's (2013) characterisation of how the rat goes about completing the task is explicitly inspired by Marr's theory. Aside from the obvious inclusion of Marr's levels of investigation and cognitivist items like representations, what Marr allows and disallows in terms of hypotheses seems important for Miłkowski's (2013) sketch of an explanation. It guides where he looks, and what he looks for. This in turn influences the explanation Miłkowski is ultimately able to provide—naturally, one that adheres to Marr's levels of analysis in providing a story about how computational processes and representations operate in the hardware of the nervous system.

Contrast this with an enactive alternative that eschews these theoretical commitments in place of its own (Di Paolo et al 2017). For one, Di Paolo et al's (2017) sensorimotor enactive view explicitly excludes computation and representation as items relevant to an explanation of cognition. In place of Marrian levels "enactivists see agents as making sense of their environment by coupling precarious processes of self– individuation (at different levels) with environmental dynamics" (pg. 26), such that "...cognitive activity does not depend on vehicles storing information, but on the coordination of dynamic processes at various scales by an autonomous agent." (pg. 27)

Accordingly, enactivists attempt to propose their own set of descriptive and explanatory items that depart from the cognitivist mould in order to investigate this (admittedly sometimes vague) notion that cognition consists in dynamical coupling and coordination between organism and environment. Sensorimotor enactivists propose the notion of sensorimotor contingencies (SMCs) to explain basic capacities of organisms to navigate and recognise features of their environment without appealing to representations at this level of explanation, where SMCs are "...regularities in the sensorimotor field: predictable or "lawful" co– variations of sensory stimulation, neural, and motor activity." (O'Regan & Noë 2001). Di Paolo et al (2017) provide an example:

> ...the projection of a horizontal line onto the retina changes from a straight line to a curved arc as one shifts the eye's fixation point from the line itself to points above or below it. In contrast, if the focus is moved along the line, no such transformation takes place. The geometry of the viewed object, the morphology of the retina, and the particular movement pattern employed all determine these regularities in the sensory stimulation pattern (O'Regan and Noë 2001, p. 941). (Di Paolo et al 2017, pg.34)

These tight relationships between sense and motor behaviour lead enactivists like Di Paolo et al (2017) to consider the constitutive "sensorimotor" whole to be the object of inquiry: in other words, a SMC. SMCs have consequently been used generate hypotheses and guide lines of empirical inquiry that differ in their output to the cognitivist mode of inquiry: they have guided the generation of hypotheses about the nature of habit (Egbert & Barandiaran 2014), perceptual learning (Di Paolo et al 2014), and social cognition and participatory sense–making (De Jaegher et al 2010; Froese et al 2015, 2018).

Hence in line with Bechtel (2016) I argue that theories of cognition are decisive to explanations that research produces. They are no mere expository gloss as M claim but a pervasive influence on how cognitive science is done. Just as representationalism influences how neuroscientists think about explanations of place cells, and how cognitivists try to explain perception and action, I think it is equally plausible that taking a sensorimotor enactive perspective influences the kinds of empirical work to be done, and hence the explanations that are produced. 3.2.

3.3.3 Abandoning theories for mechanisms; or, science with a blindfold

M propose that we should set aside questions of theory and focus instead on building models of cognitive mechanism: we will make a transition from cognitive science guided by theory, in favour of mechanistic explanation as a guide. But is mechanistic explanation equipped to serve this role?

M seem to assume that mechanistic explanations will cut through the thicket of competing theory to find the correct description or explanation of a cognitive phenomenon. In part this seems to arise from the notion that ontic mechanistic explanations, being objective explanations (Craver 2007, 2014) are able to discern correct from incorrect hypotheses by simply revealing the actual causal structure of nature. I discuss this notion of objective explanation at length in Chapter 5, and so I bracket it for now.

In any case, it is hard to see how Mechanism is a fill–in for having a theory of cognition. Theories are required to formulate hypotheses as part of a line of investigation, and to outline the items that can figure in that investigation. By contrast mechanistic explanation is not a theory, but a set of explanatory standards and strategies. It may be closely associated with some theories, like cognitivism, but it is not itself a theory. Mechanism can provide a clear set of requirements on what counts as a description or explanation of a phenomenon. But if it is pushed into taking over theory's role—proposing items of investigation, developing hypotheses, and so on—then it will fail because it is simply the wrong thing for the job. M's rejoinder to cognitive scientists to simply look for the "entities and activities that are jointly responsible for their phenomena of interest" (pg. 4) in pursuit of an integrated cognitive science neglects that investigators still need to develop hypotheses and theories about what kinds of entities and activities they are looking for. Should a given explanation hypothesise and test the role of representations, entelechies, Cartesian animal spirits, or something else? Here is where theories are instructive and decisive, and the mechanistic interest in entities and activities offers little guidance.

Without theories, all cognitive science is left with is the interventionist method (Woodward 2003) provided by Mechanism to take their place. It is difficult to imagine a science so impoverished, but hypothetically it would involve wandering around intervening on bits of the world in the hope that they will be relevant to the phenomenon scientists are interested in. To realistically perform the tasks of science, researchers need theories which can direct and shape their interpretations of data into existing research, and to guide the development and testing of hypotheses. It is not clear from M's account how Mechanism fills this role.

3.4 Heuristics

To buttress their account, M introduce the notion of heuristics as guides for research in cognitive science. They claim that heuristics, when paired up with mechanistic explanation, can take over the job of theory. We have already seen how Mechanism by itself is unable to accomplish this task—here I will argue that the addition of heuristics does not remedy the situation.

3.4.1 Depends what you mean by heuristic

If we accepted M's premise that theories of cognition are no longer viable, they would need to show how heuristics take their place in concert with mechanistic explanations. This proposal seems to hinge almost entirely on what M mean by heuristic—and they are not entirely clear on the point.

For instance, when M claim that a computational heuristic leads investigators to ask "what the algorithms and representations involved are; and how they are physically implemented" (pg.4) the line between heuristic and theory in this statement is made blurry, by contrast with their earlier characterisation of heuristics as gentle "reminders". The proposal of a stable of entities and activities—algorithms and representations implemented in the brain—seems a fairly substantial proposal more in line with a theory, a cataloguing of the "items in that domain" in Craver's (2002) terms. If this is what M mean by heuristic, and heuristics are, for all intents and purposes, theories, then the nature of the heuristic/theory distinction seems uncertain.

Contrast this notion of heuristic with how other mechanists use it. Bechtel & Richardson (1993) advanced the influential dual heuristics of decomposition and localization as guides to discovery for mechanistic science. They characterise their heuristics as "fallible research strategies" (Bechtel & Richardson 1993, pg. xxx) not specific proposals for what kinds of items can figure in cognitive explanations. Using Bechtel & Richardson's (1993) notion of heuristics, a clearer distinction from theories can be made. Heuristics are different to theories is that they do not make claims about the possible items of explanation, nor are they linked to notions of falsification. A theory can both propose features of nature (representations, computations) and be wrong about those proposals. Heuristics, on the other hand, do not propose the existence of things, instead providing guidance about how to proceed with a scientific investigation (likely within the parameters of possible items of explanations as provided by theories). Similarly, heuristics will not be dismissed if they prove to be fallible during the course of investigation, unlike theories. If a particular system proves to be non–decomposable, for instance, Bechtel & Richardson (1993) would consider this perfectly acceptable—heuristics are only meant to be fallible strategies, after all.

3.4.2 Smuggling cognitivism; Or, some heuristics are more equal than others

M seem to somewhat adhere to this characterisation of heuristics, but also blur the lines of this heuristic/theory distinction when it is enactivism under consideration versus cognitivism. Unlike cognitivism which gets to retain its capacity to propose entities and activities at work in cognition one made into a heuristic, enactivism is stripped of this kind of consequence. This is done via progressively weaker and less accurate definitions of enactivism and its goals. M start with a description of enactivism that many of its proponents would probably agree with:

> The enactive approach to cognition recognizes a crucial inter–dependency between an autonomous agent and the world it inhabits. Cognitive activity is wholly determined neither by the agents nor by their environment, but rather it emerges from the inter–dependency between the two.(pg. 3)

Compare this with the very similar first commitment of enactivism according to Gallagher (2017):

> Cognition is not simply a brain event. It emerges from processes distributed across brain–body–environment...From a third–person perspective the organism–environment is taken as the explanatory unit (pg. 6)

But later when building their mechanistic–heuristic account, M change their description dramatically:

> ...the enactive perspective (at least in its non–classical version) points to participatory negotiation of how the activity is perceived by various agents involved...(pg. 4)

> ...the enactive perspective will stress the importance of dynamic coordination between pilots and consider how the environment is structured in terms of various affordances.(pg. 7)

In the former set of descriptions, cognition *emerges* from interaction. The agent and its environment are *interdependent*. Cognition is not internal but depends on an agent interacting with an environment. These interdependencies are items of explanation in their own right. But without warning, this enactivism is transformed almost unrecognisably. The stronger language of emergence and interdependence (with the potentially deeper ontological or metaphysical connotations these terms carry) becomes quite weak and merely suggestive instead. Enactivism as a heuristic *points to* interaction; it *stresses its importance* for explanation. Something has happened in between these two sets of descriptions, and M do not say what.

M may intend for these weaker descriptions to represent the heuristic version of enactivism, and hence consider the weakening to be justified. But their conversion of theories into heuristics was justified originally by the claim that enactivism made no substantial proposals in the first place—conversion to heuristic was supposed to be a lossless process. So–called theories were really just heuristics all along, were already used as such, and their differing claims were in practice just expository differences. So, M upset their own case by tinkering with enactivism's claims. If the deformation of enactivism is necessary for conversion into a heuristic, then (if we follow M's reasoning behind redeploying enactivism as a heuristic) there was no rationale for converting enactivism into a heuristic in the first place.

This difference in standards boils down to a strategically inconsistent notion of heuristic. When it benefits the maintenance of a traditional cognitive science, heuristic means an account is stripped of its meatier theoretical implications and turned into a reminder for good conduct of mechanistic explanations. However, when computationalism and representationalism are turned into heuristics they retain their full import, able to function as theories, proposing and constraining the possible items of explanation. To make the contrast in treatment clearer, enactive proposals like SMCs are not given this benefit of the doubt—they are not assumed to be a necessary item of explanation in cognitive science unless proven otherwise.

One can certainly treat enactivism as merely a set of suggestions or reminders to look for inter–agential causal relations when mechanistically explaining some cognitive phenomenon. The defeat of a straw enactivism however has few implications for living, breathing enactivism as a theory of cognition: while a truncated enactivism suits M's proposal it is not an enactivism anyone actually endorses.

3.5 Integration

M, as we have seen, consider the notion of integration a powerful motivating force for their criticism of existing theories of cognition , and for developing their own replacement account. They suggest that integrating disparate theories—"wide" and cognitivist—will make for a better cognitive science. One way they motivate this is by suggesting that intractable theoretical dichotomies (discussed in Section 3) can be dissolved via integration—the new unified synthesis of both halves of the dichotomy will lead to a better cognitive science. I agree that it is a proposal worth investigating.

3.5.1 Conditions for integration

If we grant for the sake of argument that theoretical quagmires and dichotomies are a problem, this raises the question of how integration will resolve that problem. Not all dichotomies are best revolved through integration. It would not, for instance, seem like a good idea to integrate heliocentrism with geocentrism simply because they are theories that sit on either side of a theoretical dichotomy. This raises the possibility that the opponent theories of enactivism and cognitivism are not, much like helio– and geocentrism, theories that we would want to (or even could) integrate.

Contradictory proposals—the sun orbits the Earth; the Earth orbits the sun—do not seem integrable without one account or the other turning out to be false, which would prevent integration. On the other hand, we may find theories in opposition, but where genuine integration is possible. Segundo–Ortin et al (2018) for instance discuss the apparent incompatibility between enactivism and ecological psychology (Gibson 1979). Enactivism and ecological psychology agree on many fronts: both argue that representations are unnecessary for perception, that perception is an embodied affair, and that explanations of perception can be dynamical. However, some enactivists have claimed that the two accounts are at odds, targeting what they perceive as an excessively representation–friendly notion of *information* on the part of ecological psychology (Myin 2016).

Information is a sticking point because, as Segundo–Ortin et al (2018) relate, enactivists and ecological psychologists seem to be understanding the concept in two different ways. Ecological psychologists claim that perception involves the picking up of information in the environment by an organism. Enactivists worry that ecological psychologists therefore think that "this information *specifies* or *is about* the environment..." (Segundo–Ortin et al 2018, pg. 5, emphases original), which sounds dangerously close to the view that information is a kind of contentful representation of the environment for perception.

But for ecological psychologists the terminology carries a different meaning: to pick up or specify features of the environment does not mean producing a representation of the outside world for internal processing, but rather using the environment as structured by the organism's body and its movements:

> Consider...the example of optic flow...an individual's movements in her environment lawfully produce invariant patterns in her sensory array. For instance, as any animal moves toward an object, the image this object projects in her retina lawfully expands, causing the object to expand in her visual field...(Segundo–Ortin et al 2018, pg. 12)
>
> Due to the interaction between the light and the objects in the room...the optic array gets structured, and insofar as this structure corresponds to

> the structure of the surroundings, the former can be said to specify or "contain" information about the latter. (Segundo–Ortin et al 2018, pg. 8)

So understood, information specifies invariant relationships between the incoming light from the organism's surroundings, the sensory organs and the organism's movement. Hence specification is not contentful or representational (providing a replica of the environment for internal uptake, processing and so on) but simply a form of lawful covariation between organism and environment, which organisms thereby pick up directly via action. It thereby does not imply contentful representations as core to perception. This clears up this expository difference between enactivists and ecological psychologists, and Segundo–Ortin et al (2018) clear the path for genuine integration.

A few key observations can be made here. Both accounts agree on most features of cognition—that it is embodied, nonrepresentational, and dynamically explicable, for instance. The apparent incompatibility between the two (the ecological notion of information) proves to be resolvable to their mutual satisfaction. The critical back and forth between the enactive and ecological approaches opens up the possibility of genuine rapprochement, and ultimately integration.

3.5.2 Unity or conquest?

Does M's proposal resemble this kind of integrative work? There are, similarly, differences in commitments between cognitivists and enactivist. Though perhaps—following Segundo–Ortin et al's (2018) example—these differences can be shown to have a mutually satisfactory resolution. And M do present their account as a pluralistic view on cognitive science, committed to getting on with the real business of explaining cognition and setting aside intractable expository differences via integration.

The problem is that unlike Segundo–Ortin et al (2018), M do not engage with actually existing enactivism to find a path to integration. Instead, a straw enactivism that is only committed to reminding cognitive scientists to look for inter–agent interactions is incorporated into cognitivism. This hardly seems a case of genuine integration: if there are deep disagreements between accounts, then they bar the way for integration (for instance, one cannot hold that representations are both required for explanation and unnecessary for explanation simultaneously). Their strategy of weakening enactivist claims means that enactivism and cognitivism never get the opportunity to settle their score. The anti–representationalism of enactivism, for instance, is simply not addressed. Cognitivism, on the other hand, gets to retain its representational commitments, and hence does not have to face any critical perspectives. M are attempting integration, but they do not attempt to show how fundamental disagreements between cognitivism and enactivism can be resolved to allow it.

I suggest that M's account does not genuine integrate differing theories of cognition, because it fails to engage with the substantive disagreements that stand it its way. Their manoeuvre has been a multifaceted avoidance of this problem; by pressing Mechanism into service as a replacement for theories, so as to avoid theoretical disagreements; by misrepresenting enactive cognitive science to avoid theoretical disagreements; by converting enactive theory into a straw–version heuristic.

On the other hand, it may simply be M's intention to bury enactivism alive: to perceive that it is a living, breathing account of cognition but nevertheless put together a unificatory story that buries it as an independent research tradition, turning it into a mere heuristic for Mechanism instead. In this latter case, this intention shouldn't be presented as an even–handed treatment of the theories populating cognitive science. This winds up being counterproductive—if it is an end to enactive cognitive science that M want, then their case should presented as such, and not under the guide of unity. Otherwise any such argument will fail to achieve its goal: enactivists will likely not buy into this uneven story and accept integration. To put an end to opponent E–cognitive accounts, traditional cognitive science ought to show how they fail to do the job of theory; that they are somehow wrong or misguided in their explanations of phenomena.

3.6 Heading off the Non–mechanistic Option

Having established their mechanistic–heuristic account for cognitive science, M make a case to undermine a potential dynamical alternative. We have already seen how cognitivism and mechanism share a close and harmonious relationship—enactivism and mechanism, far less so . Traditionally, many enactivists appeal to dynamical models as their primary descriptive and explanatory tool (Chemero 2009). M attempt to head off this explanatory alternative by attacking the possibility of dynamical explanation—though I will show how the move also fails. Their first charge is a reiteration of the causal relevance concern. I will not labour the point, and simply respond that the interventionist account of dynamical explanation proposed by in Chapter 1 goes some way to providing a solution here. However, M have one further, more novel criticism of dynamical explanations.

3.6.1 Vacuous mechanisms

To head off dynamical explanation, M move to expand the definition of a mechanism to include all causal relations. The authors initially propose a familiar and fleshed–out definition of mechanisms:

> A mechanism is an organized spatiotemporal structure responsible for the occurrence of at least one phenomenon to be explained. The orchestrated causal interaction of the mechanism's component parts and operations explains the phenomenon at hand." (pg. 5)

Both mechanists and dynamicists would likely agree to this definition. Immediately afterwards this is truncated to the claim that mechanisms "are only causally organized spatiotemporal structures." (pg. 5) This truncation is not a harmless summary, because it quietly broadens the definition of mechanisms to make it, as I will show, excessively permissive.

If we accept M's definition of a mechanism for a moment, the consequences are worth considering. Even if dynamical models appeared to provide us with causal explanations, we would always turn out to be misled. Since any "causally organized spatiotemporal structure" is by definition a mechanism, even a single relationship of cause and effect might count as a "causally organized spatiotemporal structure" and therefore be counted as a mechanism. One might consider this reading unfair, but this does seem to be the intended meaning on the part of M, since they claim a direct equivalence between the causal explanations identified by dynamical models, and their mechanistic counterpart:

> It is extremely difficult to find dynamical explanations that do not appeal to causally organized systems; and this is what makes purported dynamical explanations equivalent to mechanistic explanations. (pg. 12)

Ultimately "causally organized spatiotemporal structure" is condensed to "causally organized system", which functions identically to "cause" in their account. While M may not intend this outcome, it seems any cause is now a mechanism. This is somewhat of a departure from the more measured 3M requirement (Kaplan & Craver 2011) which suggests that dynamical causes *can be reduced* to mechanistic causes; 3M denies that dynamical causes are actually causes. Here M are saying that dynamical causes *just are* mechanistic causes, since a sufficiently broad definition of mechanism can include any causal story dynamicists can come up with.

A likely sticking point here is the discord between this and existing definitions of mechanisms. For instance, Bechtel & Abrahamsen's (2005) influential definition describes a mechanism as a structure performing a function in virtue of its component parts, component operations, and their organization" (pg. 423). Craver

(2007) defines mechanisms as "a set of entities and activities organized such that they exhibit the phenomenon to be explained." (pg. 5) Some other definitions of mechanisms are more restricted, specifying that even organised parts or entities engaged in causal activities is not precise enough and that they must have extra features like fine–tunedness of organisation and modularity (Woodward 2013). The classic Machamer et al (2000) definition included the requirement of start and finish conditions to a mechanism's operations. There is generally more meat on the bones, then, than just causally organised spatiotemporal structures.

At the core of the notion of mechanism is the important requirement that they possess discrete physical parts which interact with one another causally (or at least have the potential to). The interactions of these discrete physical parts are of explanatory interest; showing how they produce the phenomenon is the business of mechanistic explanation and what individuates it from other modes of explanation. To accept M's definition would be to reduce Mechanism to barebones interventionism (Woodward 2003). Broadening the definition of mechanisms to this extent simply detonates the mechanist project as something distinct from interventionism.

It seems unlikely that most mechanists would be pleased with this scorched–earth tactic, nor the ultimate outcome, since M's definition allows any and all causal relations to count as genuinely explanatory mechanisms without actually referring to any entities or activities—core components of any mechanistic explanation.

3.7 Conclusion

Miłkowski et al (2018) touches on several discussions around the possibility of integrating E–cognition and cognitivism. It attempts to provide a unifying account for cognitive science based upon mechanistic explanation. However, I see their integrative strategy as unsuccessful. Miłkowski et al (2018) recommend setting aside theories of cognition and replacing them with heuristics wedded to mechanistic explanation. This alternative, what I call the mechanistic–heuristic account, fails to provide a credible replacement for theories of cognition, and misrepresents enactivism in order to integrate accounts, thereby failing to actually resolve theoretical disagreements between these accounts. Additionally, their attempts to head off the possibility of dynamical explanations rely on an overly permissive definition of mechanisms that renders them vacuous.

Chapter 4

Explanatory Chimeras

Integration of dynamical and mechanistic explanation under enactive cognitive science has recently received several treatments. Two separate proposals (Abramova & Slors 2019; Walmsley 2019) have argued that mechanistic explanations can be integrated into existing dynamical explanations, and thereby serve the goals of enactive cognitive science. Abramova & Slors' (2019) account claims that dynamical models only describe at the *what–level*, not the properly explanatory *how–level*, and so enactive cognitive science requires augmenting with mechanistic models to properly explain. I argue this move defers unnecessarily to the explanatory goals of traditional cognitive science, and results in the subsumption of dynamical explanations into their mechanistic counterpart. Walmsley (2019) proposes that strategic mechanisms (Levy 2013) can be used to help improve understanding of phenomena even when that system is not especially mechanistic. I agree with Walmsley (2019) that this limited, strategic type of mechanism is useful, but I also consider this an instance of subsumption in the other direction—here mechanism is subsumed into dynamical explanation, rather than being genuinely integrated. As a result, a genuine integration is still not forthcoming.

4.1 Introduction

How should we characterise the relationship between dynamical and mechanistic explanations? Competition, compatibility, or something else? So far in this thesis I have tried to make the case that non–mechanistic explanations are able to stand alone, drawing on the same interventionist justification as mechanistic accounts (Chapters 1 & 2). Of course, one might ask the question: why go it alone? Why not somehow unify or combine these mechanistic and dynamical explanations together to yield a multifaceted, and overall more informative, explanation of a phenomenon?

At least as far as cognition is concerned, however, this view runs into some apparent problems. This occurs because dynamicism has long been wedded, within the cognitive sciences, to E–cognition approaches, particularly of the radically embodied or enactive variety (Chemero 2009, Chemero & Silberstein 2008, Di Paolo et al 2017) while mechanism has a similarly robust connection to traditional cognitivism (Miłkowski 2013, Zednik 2018). This does raise the question of whether these apparently strict allegiances can be broken out of. After all when (as I showed in Chapter 3) there are substantial and non-trivial differences in theories that consequently propose very different items of explanation, and relate very different lines of investigation.

The manoeuvre I examined in Chapter 3 was Miłkowski et al's (2018) attempt to sweep aside this these theoretical debates. They argued that mechanistic explanation can supercede theoretical differences between classic cognitivist accounts and E–cognition; sweeping them aside to get at the best overarching explanation. Here instead I will look at another attempt at integration: this time the unification of different model explanations (dynamical and mechanistic) under the same theoretical framework for cognition.

4.1.1 Compatibility & Integration

Mechanists have long held that compatibility between dynamical and mechanistic explanation is possible and often desirable (Bechtel & Abrahamsen 2010, 2013), although with the caveat that a dynamical model is only explanatory insofar as it illuminates a mechanism (Kaplan & Craver 2011). Dynamicists have been leerier of the prospect, arguing that the decomposition and localisation required by mechanistic explanation is inimical to the goals of dynamical explanations paired with E–cognition accounts (Stepp et al 2011; Silberstein & Chemero 2013). In Chapters 1 and 2 I also argued, from an interventionist perspective, that dynamical models can be causally explanatory in their own right and so do not necessarily require subsumption into mechanistic explanations.

One point of confusion in this existing discussion stems from the sort of compatibility under discussion. Mechanists generally claim that dynamical explanations aren't explanations in the first place, but merely descriptive models. Hence whenever compatibility is on the table, it is not the compatibility of two equal modes of explanation being discussed but rather the compatibility of purely descriptive dynamical models with genuine mechanistic explanations. This ends up resembling subsumption more than true compatibility—the explanation is still ultimately a mechanistic one, serving mechanistic explanatory goals, but utilising dynamical models as part of its descriptive toolbox (Kaplan 2015).

What I focus on here is instead the possibility of genuinely *integrating* the two explanatory strategies. For clarity, I therefore distinguish compatibility (which I suggest entails subsumption; with making one or the other type of explanation fit into the other) from integration. I wish to find whether both models can meet their explanatory targets in their own terms, without stepping on one another's toes: in the dynamical mechanistic explanations mentioned above, dynamical models are curtailed by claiming they are not responsible for the explanatory power of the explanation. Making such a distinction is helpful for discussing the relationship between dynamical and mechanistic explanations with a focus on a genuine and even-handed integration.

4.1.2 The virtues of integration: hybrid vigour?

Much of this discussion around mechanisms and dynamical models concerns the notion of *model explanations*, where explanatory power results from the successful exhibition of some pattern observed in nature by a model (Bokulich 2017). Mechanistic models, for instance, explain by showing how the causal events going on in a mechanism exhibit a fine–grained control over the explanandum phenomenon; dynamical models (at least on the account presented in Chapter 1) explain similarly by showing how the dynamical variables of a system causally influence its outcomes.

From a philosophy of science perspective, combining multiple models together is seen to have potential benefits for a scientific investigation. Since models necessarily leave out or abstract some details about the world, combining the insights of several different kinds of models can help control for this idealisation somewhat (Levins 1966; Potochnik 2017). Integration is also perceived within cognitive science to have the potential to bring together disparate theories of cognition, to find "a way to end the representation wars" (Constant et al 2019, pg. 2) between E–cognition accounts and mainstream cognitive science. If different camps can share and integrate their explanatory resources, it might follow that little of substance outside expository differences thereby distinguishes them. Proponents of active inference and predictive processing models have for instance argued that a shared explanatory resource can integrate E–cognition and cognitivism (Clark 2015; Constant et al 2019) with the aim of developing an overarching account that "accommodates a representationalist and a dynamicist (a.k.a. non–representational) view of cognition." (Constant et al 2019, pg. 2)

I will however focus not on these outside interventions into the dynamicism–mechanism divide, nor the cognitivist–enactive divide, hoping to integrate the two through a new overall framework for cognition. Instead I will investigate whether dynamicism and Mechanism as forms of model explanation already possess the resources to fully integrate. In this chapter I will discuss two recent attempts to integrate dynamical and mechanistic explanations within the same theoretical framework: enactive cognitive science. The first is Abramova & Slors' (2019) proposal of a mechanistic explanation for enactive sociality, and the second is Walmsley's (2019) call for the integration of a strategic mechanism into enactive accounts of minimal cognition. Both Abramova & Slors (2019) and Walmsley (2019) argue that the combination of the two models—dynamical and mechanistic—will produce better explanations than either mode alone (though they have rather different reasons for thinking so, as I will show). These proposals both aim for a kind of hybrid vigour, where a combined dynamical–mechanical account of a phenomenon will prove better than either model alone.

But how to gauge whether these accounts succeed, and do so in a genuinely integrative way? I suggest that the answer to this question lies in whether these combinations of models genuinely contribute to the *explanatory goals* of the cognitive theory in question, which are often mutually antagonistic. I will show that while these integrative cases are both partially successful, I do not think that either account wholly succeeds in providing a roadmap to the genuine integration of dynamical and mechanistic model explanations under an enactive framework. I argue that, due to the requirements inherent in the specific explanatory goals of enactive cognitive science, these integrative attempts both result in one model explanation being subsumed into the other. While they purport to integrate mechanisms into enactive cognitive science, they either result in mechanistic subsumption of dynamical models (Abramova & Slors 2019), or dynamical subsumption of mechanism (Walmsley 2019). In both cases I think that accepting one or the other sets of mutually incompatible explanatory goals—cognitivist in the first case, and enactive in the latter—seems a difficult obstacle for genuine integration.

4.2 Explanatory goals

One feature central to both enactive and traditional cognitivist approaches to cognitive science is their interest in providing explanations of cognition. This is not surprising given explanation is a core goal of science (Bokulich 2017). In order to be taken seriously, any account of cognition needs to be able to explain cognitive phenomena. One difference between the enactive and cognitivist approaches is their *explanatory goals*, which I argue stem from their characterisations of the explanandum phenomenon; “[r]esearchers in different research communities and with different research projects often possess different explanatory goals” (Walmsley 2019, pg. 2). These explanatory goals dictate the questions investigators ask about phenomena, and consequently the kinds of answers considered suitable for explanation.

For instance, cognitivist accounts characterise cognition as a chiefly internal affair, performed via the central nervous system; “the sequential processing of strings of static symbols by means of some formal algorithm”(Slors et al 2015, pg. 241). To take one relevant example, social cognition on this view should be understood as the result of a specialised mindreading module, which processes incoming sensory input (according to ingrained rules) to figure out what others are thinking (Carruthers 2016, 2017).

This characterisation at the *what-level*, the level of specifying the nature of the phenomenon, frames the explanatory goals of cognitive science. So understood, we now aim to understand how the symbol-processing and rule-following of a particular mental module produces the phenomenon, the thing we have characterised already at the what-level. Consequently the explanatory goals of traditional cognitive science track these aims: to consider how a system might, in theory, be able to realise these functions, and to find the actual realisers in the cognitive system and show how these realisers produce the phenomenon in question.

This traditional perspective often coincides with an affinity for mechanistic explanations of these cognitive phenomena. At least in part this close relationship can be understood as the cohesion of the explanatory goals of cognitivism, and the kinds of model explanations Mechanism provides, as exemplified by Marr’s (1982) levels of analysis (discussed in Chapter 3):

> “The idea of algorithms governing the transformations of representations is familiar from computer science but, as we will see, is not limited to the sorts of representations and algorithms used in digital computers.Rather, it allows for the identification of the organization of the mechanism that explains how representations are manipulated to generate the phenomenon.”(Bechtel & Shagrir 2015, pg. 313)

If an investigation is committed to the classic cognitivist picture of cognition (here represented by Marr (1982)), then regardless of whether it is utilising mechanistic or dynamical models (or a combination thereof) “...there is reason to believe that all of these research programs...aspire to discover and describe mechanisms.”

(Zednik 2018, pg. 390). Once the explanatory goals of cognitivism have been accepted, it is a natural next step to appeal to mechanisms which align with the what-level characterisation provided by cognitivism.

Mechanism is well–suited to addressing the explanatory goals of cognitivism because it engages with both the *what–level*, which describes the phenomenon, and the subsequent *how–level* which describes how the phenomenon was underlain by the operations of an underlying causal mechanism (Abramova & Slors 2019). This how–level provides the step that turns a mechanistic description/model into an explanation. Hence mechanism both locates physical realisers and shows how they realise, in synergy with the explanatory goals of cognitivism.

On the other end of the spectrum, enactivism (and related versions of radically embodied cognitive science) entails very different explanatory goals. Consider Di Paolo et al's (2017) characterisation of the explanatory strategy of enactive cognitive science, focussing on sensorimotor interaction as a core feature of cognition:

> In contemporary terms the sensorimotor loop is manifested in the way in which motor variations induce (via the environment) sensory variations, and sensory changes induce (via internal processes) the agent to change the way it moves. The regularities that emerge from recurrent sensorimotor cycles are constitutive both of action and perception. These are also the regularities that the cognitive system is sensitive to and attempts to manage.(Di Paolo et al 2017, pg. 17)

Analysis of the coupled agent–environment system, typically through dynamical systems techniques, is the prevalent research method for studying the emergent regularities in closed sensorimotor loops."(Di Paolo et al 2017, pg. 18) In enactive cognitive science the explanatory goal is to find out how features of the world are coupled dynamically such that they produced the phenomenon. This is because Di Paolo et al's (2017) characterisation of what cognition is—not the implementation of computational processes, but the dynamical couplings the agent–environment unit constitutes. From this we can sketch some explanatory goals. With cognition so characterised, the explanatory goals of enactivism are the successful description of how couplings between relevant features of the world—brain, body, and environment—produce different outcomes as those couplings unfold over time. Furthermore, this fits with an interventionist, counterfactual image of explanation—showing how things would have been different had these couplings been different (see Chapters 1 & 2). Hence dynamical models, which say little about physical realisation, and a lot about the unfolding of couplings over time, serve the explanatory goals of enactive cognitive science well.

The key point is that the characterisation at the what–level is what drives the goals of a scientific investigation. Moreover, I argue against those in favour of subsuming dynamical explanations into mechanistic ones under a cognitivist framework (i.e. Zednik 2018). Zednik (2018) points to the tight relationship between mechanism and cognitivsm and judges that due to their mutually supportive goals and assumptions, this indicates an all-encompassing perspective which the dynamical hypothesis unwittingly serves. But I see this as essentially a non-sequitur. It is true that, so characterised, cognition is neatly explained via mechanisms. It is no surprise that if we characterise cognition as a set of cognitive mechanisms, the explanatory goals this generates will be served well by mechanistic explanation. But this does not seem to say very much about the possibility of alternative characterisations of cognition, and the explanatory tools they use. One can always characterise the phenomenon differently at the what-level, and so generate different explanatory goals. These cannot be so easily subsumed.

4.3 Creating chimeras

Despite these differences in explanatory goals, there have been some recent attempts to integrate mechanistic explanations into enactive cognitive science alongisde dynamical models. The first is Abramova & Slors (2019), who argue that mechanistic explanation can augment an enactive account of social cognition. The second is Walmsley's (2019) case that the integration of mechanistic models alongside dynamical ones will lead to better explanations of minimal cognition.

4.3.1 Enactive social cognition

Abramova & Slors (2019) target discussions of enactive social cognition. Social cognition has traditionally been thought of as an act of mind–reading: simulating or otherwise modelling the inner mental states of other agents, often performed by a specialised internal mindreading component (Baron–Cohen 1995; Carruthers 2016, 2017). As part of their general criticism of traditional cognitive science, enactivists have been highly critical of this characterisation of social cognition. As an alternative, they propose social cognition is done through direct perception, not figured out at a distance by internal processes (Gallagher 2008). Social cognition should be considered not as an abstract, disembodied process of inference, but as the consequence of inter–agential coupling:

> Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the

> domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced)." (De Jaegher & Di Paolo 2007, pg. 493; also quoted in Abramova & Slors 2019)

Making this proposal more concrete, Di Paolo et al (2010) appeal to perceptual crossing experiments (Auvray et al 2009; Auvray & Rohde 2012). These experiments attempt to show how quite simple, non–mindreading sensorimotor capacities of agents are able to constitute basic social interaction. The experimental design puts two participants in separate rooms, and given a computer mouse and a tactile pad. What they cannot see is a shared virtual space, where both participants control their own avatar. Each avatar is moved along a one–dimensional horizontal line, left or right, when the participants move their mouse. If the avatars run into each other, a tactile sensation is sent to both participant's pads. They are able, thereby, to feel the presence of the other participant's avatar. There are also several stationary objects which similarly induce a sensation if a participant touches the object with their avatar. In addition, each avatar has a "shadow", which stays a set distance from their avatar but tracks its movements. Shadows can be interacted with by the other participant's avatar (the other will feel a tactile response), but the shadow's owner will feel nothing. Participants are instructed to look for the other participant, and click the mouse when they felt they were "touching" the other person's avatar, and not when they think they've encountered a shadow or object.

The experiment presents three possibilities for interaction: avatar–avatar, avatar–shadow, and avatar–object. The former kind is—according to Di Paolo et al (2010)—special, a form of basic social interaction. By contrast the avatar–shadow condition presents a situation where only one of the agents can sense the other, and sense their interaction—it is hence only one–way, not a case of social interaction.

The results showed that participants were indeed able to find the each other's avatars the majority of the time, correctly indicating via mouse clicks that they were interacting with the other person and not their shadow, or a stationary object. From this, it is concluded that "[t]he situation of sensing the other' while being sensed' turns out to be more stable than sensing an insensitive shadow." (Abramova & Slors 2019, pg. 410). Moreover, the task is not solved by highly involved mindreading capacities, but only through very simple sensorimotor interactions—"from a supra–personal perspective the task is solved by the fact that avatar–avatar interactions are more stable, more self–perpetuating, than avatar–shadow interactions." (Abramova & Slors 2019 pg. 411).

Dynamical explanations of this phenomenon have been developed further by Di Paolo et al (2008) and Froese & Di Paolo (2010). In these studies, the perceptual crossing experiment was performed by neural network–controlled simulated agents.

Despite being equipped with very simple resources (lacking explicit mindreading systems), the continued interaction of the two agents was extremely stable, much more so than the one–way avatar–shadow or avatar–object interactions. The stability of the avatar–avatar interaction is explanined dynamically, in terms of features of "the dynamical landscape, attractors, perturbation and hysteriesis" (Abramova & Slors 2019 pg. 412) much like the analysis of cell fate explanations I offered in Chapter 2.

While this is counted as a successful dynamical description of a kind of minimal social cognition, Abramova & Slors (2019) nevertheless do not consider it an explanatory success. They have three main criticisms of this dynamical explanation, to which they think the solution is integrating mechanistic models.

Firstly, they worry that despite its successes, the relative lack of uptake of enactive cognitive science within the empirical mainstream is because "when it comes to explanations, enactivists content themselves with precise, prediction–supporting descriptions of the what–level" (Abramova & Slors 2019, pg. 411). They blame the lack of popularity of enactive cognitive science on the fact that scientific commonsense seems to prefer mechanistic explanations. They propose that to improve this situation, mechanistic explanations should be used to augment dynamical models. This is less a concern over the actual explanatory power of dynamical models, and more to do with popularity within a scientific community.

For the perceptual crossing example just discussed, they suggest that the extensive social interaction between the two agents can be recast as an extended mechanism, with components stretched out across both agents. While mechanistic explanations are indeed capable of explaining phenomena where the underlying mechanism is spatially distributed, I consider this a mixed blessing for two reasons.

For one, the lack of immediate mainstream appeal seems a debatable reason to throw in the towel on pursuing dynamical explanation. Dissenting voices within and between scientific communities can drive scientific progress—sometimes dissenters ultimately end up occupying the mainstream position (Kuhn 1962). In any case, popularity is not necessarily concomitant with usefulness or correctness. Secondly, without delving too far into the sociological dimension, I am not convinced the failure of mainstream cognitive science to switch over to an enactive account can be blamed solely on an absence of mechanistic explanations, nor that their addition would change this state of affairs dramatically. Enactivism is a relatively young approach to cognition research and has the unenviable task of forging a path within a discipline that already possesses a well–established and venerable research paradigm. The minority status of enactivism is not cause to abandon core principles of the account.

The second criticism resembles the causal relevance concern (Chapter 1). Abramova & Slors (2019) suggest that while dynamical models of social interaction answer questions at the what–level of the phenomenon, they fail to provide an answer at the how–level. They argue that enactive sociality is missing a justificatory step: it fails to be explanatory because it is missing a description at the how–level of cognition. But this absence can, they think, be filled by a mechanistic model. In this way enactive cognitive science is "compatible with a mechanistic how–level focus of traditional cognitive science." (Abramova & Slors 2019, pg. 407) and the two model explanations are happily integrated.

But I think Abramova & Slors (2019) make two errors here. Firstly, I suggest that they miss the close connection between the what–level characterisation and the explanatory goals of enactive cognitive science. If the phenomenon is characterised as the stable social interactivity unfolding in a system, then providing a mechanistic story seems to miss the point. Namely, it is not clear what reconceiving the agents as mechanistic components in the cognitive system adds to an explanation. Froese & Di Paolo's (2010) model already explains this phenomenon in a way that matches the what-level as construed by enactivism—the stability of the interaction is a dynamical feature of the system, and it is explained dynamically.

One can, of course, argue that this characterisation of cognition as extensively constituted is wrong (i.e. Adams & Aizawa 2001). But this is ultimately tangential: the question is not who is correct about the nature of cognition, but how one's characterisation of cognition constrains the kinds of how-level explanations that can possibly serve the explanatory goals that go along with that specific characterisation.

Second, Abramova & Slors assume that this additional how–level description of enactive sociality must be mechanistic. Consider the way in which mechanistic explanations are supposed to describe the how–level. Mechanistic explanations are given "in terms of the structural nature of (causal) interactions between particular elements that comprise the phenomenon. This is known as identifying a mechanism..." (Abramova & Slors 2019 pg. 412). The "structural nature" of causal interactions is characterised in interventionist terms—as a series of invariant generalisations that describe how the system would have behaved across all manner of counterfactual scenarios. Once we have this comprehensive counterfactual story, we have a description at the how–level (and hence an explanation).

But as I argued in Chapter 1, there are good reasons to think that dynamical models can provide causal explanations along interventionist lines, and hence get at the how–level. If the how–level is construed just in these interventionist terms—as posing and answering one of Woodward's (2003) *what–if–things–had–been–different questions*—then dynamical explanations are perfectly capable of filling in the how–level. While Abramova & Slors (2018) argue dynamical models alone can't make this justificatory step, I disagree that mechanistic details are always and everywhere required for a how–level description of cognition.

As a final point, I return to the notion of subsumption versus integration. Abramova & Slors (2019) suggest that dynamical models require the deeper how–level story of a mechanism in order to be explanatory. Looking at the end result of this manoeuvre, it seems that we no longer get genuine dynamical explanations. This is because the justificatory step, the detail that provides the explanatory power to the model, is for this account mechanistic. This closely resembles the existing story about dynamical models and mechanistic explanation advocated for by mechanists (Bechtel & Abrahamsen 2010, 2013; Kaplan & Craver 2011; Kaplan 2015). Taking this route leads not to genuine integration, but to subsumption.

Their third concern has two parts. If the dynamical explanation offered by Froese & Di Paolo (2010) for instance is an idealisation and abstraction from actual human cognition, how can it be said to explain it? Further, they argue that "there seems to be no simple way to relate the structures and processes that generate [stable interactions] in the simulated agents versus the ones operational in humans." (Abramova & Slors 2019, pg. 412). There are really two concerns here, one about abstraction/idealisation, and the other about how to show a dynamical explanation targets some feature of the world absent a mechanistic model to ground it.

The first part can be answered without too much trouble. It is typical of model explanations to idealise and abstract away features of the target phenomenon—this is not necessarily detrimental to the explanatory power of said models, is typical of them, and in fact can enhance their explanatory power (Potochnik 2015). Science is generally packed full of idealisations and abstractions that permit explanation and understanding.The fact that dynamical explanations abstract and idealise is not therefore in itself a problem.

The second part requires unpacking the hidden assumption motivating this worry. Abramova & Slors (2019) seem to hold that a mechanistic model more capable of grounding an explanation in some real–world, physical system. But why should dynamical models be less capable in this regard? Granted, the simulated perceptual crossing model does not necessarily describe human cognition. This really leaves open several possible interpretations of this state of affairs, though. On the one hand, we might think this represents a fundamental issue with the dynamical approach to explanation, since dynamical models fail to get to grips with actual physical systems.

On the other hand, this could be seen as simply reflecting the development of a line of investigation from minimal to more complex and specific models that will provide insight into real–world cognition. And indeed, some dynamical models like Thelen et al's (2001) model of infant perseverative reaching do describe in a precise and reliable fashion the behaviour of real human infants (Chapter 1). Likewise the

HKB model (also see Chapter 1), while somewhat abstracted in its original form, has been expanded to integrate neural dynamics in the motor cortex that do tell us something about human and animal cognition (Lamb & Chemero 2014). If this is what is meant by "relating" a dynamical model to a real world system, it seems adequate. If, on the other hand, some deeper ontological grounding is required, then some argument to the effect that mechanisms are a better foundation than the dynamical features of a system would be required.

4.3.2 Minimal cognition & the slime mould

The second case for mechanistic integration into enactivism is Walmsley's (2019) discussion of minimal cognition and the slime mould. Walmsley (2019) is sensitive to the fact that enactivists use explanatory strategies and models that will best serve their particular goals. The kinds of models and explanations that can be used in service of this goal is necessarily constrained—for instance, models that appeal to computation and representation are more or less out of the picture (Walmsley 2019).

Nevertheless, this leaves quite a scope for different kinds of models to be used in service of enactivist explanations, and leaves open the door for explanatory pluralism about cognition, insofar as explanation is model–based. He also dismisses Hutto & Myin's (2017) claim that enactivism is capable of providing the *best explanation* for cognitive phenomena, on the basis that explanatory pluralism is generally a better strategy than monism. I would add to this criticism that one's definition of best explanation depends largely on the account of cognition one subscribes to. Leaving aside the question of whether there can be a singular *best* explanation of a phenomenon, there seems to be no common conceptual ground that would allow a cognitivist and enactivist (for instance) to agree on the best explanation of a given cognitive phenomenon, since they fundamentally disagree about what cognition is, and therefore what explanations of it look like.

To illustrate a potential coming together of dynamical and mechanistic models under enactivism, Walmsley (2019) introduces the slime mould, a frequently used model biological organism. The mould, also called *Physarum*, has a physiology combining a spongy type of cell that has muscle–like properties (Gel) with a plasmodium (Sol) that takes up and transports nutrients around its body. Because of its bodily structure, *Physarum* is capable of a range of interesting behaviours. It will roam and search its environment in order to find food and avoid hazards, grow to reach food sources, and do so with such cost–efficiency that the networks it creates as solutions match that of real–world human–generated networks (Tero et al 2010).

An attempt at modelling *Physarum*'s behaviour can be found in Jones (2015).

This model presents a quite abstracted and idealised version of the organism, where its behaviour is modelled as particle agents (representing individual Gel and Sol cells) which move towards a chemoattractant, while constrained and directed in their movements by a lattice structure, where each space in this lattice can only be occupied by one agent at a time. From the behaviour and interactions of these agents arises an overall oscillatory, searching behaviour that closely resembles that of the actual *Physarum* organism.

Schenz et al (2017) further investigate this behaviour by devising a dynamical model that describes the dynamics of the roaming behaviour that allows *Physarum* to search its environment. *Physarum* uses its own bodily growth as a primary means of sensing its environment, since it lacks any additional sensory organs: "its growth front is its sensory organ" (Schenz et al 2017, pg. 11). To continue powering this growth, *Physarum* efficiently and adaptively lays down a vein structure, which traces an efficient path through the organism's body and transmits body mass from the rest of its body to its leading edge–the tendrils of growth that spread to explore the world around it. Hence the vein construction and flow of mass through these are constantly adapting to match the expansion of its body at the leading edge.

The joint model of these behaviours devised by Schenz et al (2017) successfully predicts and describes *Physarum*'s actual behaviour in an experimental setting. But the model, according to its creators, goes beyond just a phenomenological description of *Physarum*'s structure: it predicts and describes "key characteristics of the organism, first and foremost the centre-in-centre trajectory of the main vein, and also the tree-like vein structure, the rough leading edge profile, or shuttle streaming were successfully reproduced." (Schenz et al 2017, pg. 11)

This apparently simple but highly adaptive and responsive behaviour is proposed by Walmsley (2019) to be a good example of an organism possessing a kind of minimal cognition rooted in sensorimotor interactions with its environment. Enactivists often claim that their account of cognition does better justice to this kind of "basic" cognition than do the allegedly over–intellectualised cognitivist accounts (Hutto & Myin 2017)—hence the slime mould should make excellent fodder for enactive cognitive science. And indeed the success of the dynamical modelling approach to *Physarum* suggests a congruence with the dynamicism of enactive cognitive science.

Further, Walmsley (2019) argues that the slime mould's behaviour is difficult to characterise as being mechanistic. He introduces Skillings' (2015) spectra of isolability, sequentially, and organisation as markers of mechanistic explicability in biological systems. The slime mould's behaviour scores low on all three measures. Its behaviours involve extensive environmental features and so are not easily isolable, it behaves in a non–sequential fashion, and its activities are not finely organised. Its "cognition is thoroughly dynamical, involving not only dynamical coupling within the plasmodium itself, such as between the contracting granules described by Jones, but a dynamical coupling with the environment." (Walmsley 2019, pg. 8)

While this might seem to rule out mechanistic modelling of the slime mould, Walmsley (2019) suggests instead that a strategic mechanism (Levy 2013) could be useful in fleshing out the dynamical explanation. A strategically mechanistic approach is not supposed to contribute to an explanation by virtue of its ontologically revelatory or epistemic virtues, but rather in how it assists in the cognitive process of understanding a phenomenon, how it contributes strategically to an investigator's understanding (Levy 2013). On this view, the system is not even necessarily supposed to *really be* a mechanism—we can simply think of it as such pragmatically. In this iteration, a mechanistic model of the slime mould simply helps ground our understanding *Physarum* by connecting the dynamical model with the physical features of the mould.

So Walmsley presents an interesting combination of dynamical and mechanistic models. He accepts that dynamical models can have causal–explanatory power at the how–level (Chapter 1) and therefore do not require causal–mechanistic augmentation for that purpose. Schenz et al's (2017) model, which describes and predicts how *Physarum* moves through its environment, and the physical structures that form as *Physarum* moves, on this view doesn't need extra mechanistic detail for it to explain these features of *Physarum*'s behaviour. But in addition, Walmsley (2019) proposes that we could introduce a strategic mechanistic sketch to bolster understanding of the phenomenon:

> ...systems can be represented as more mechanistic than they are for the sake of some epistemic payoff. This means that mechanistic models compatible with [enactivism's] view of cognition may still not support straightforwardly mechanistic explanations of cognition, since mechanistic models may instead demonstrate that interaction between parts is more explanatorily relevant than individual components and their special properties. (Walmsley 2019, pg. 10)

But does this proposal produce genuine integration of two model explanations? I think that, much like Abramova & Slors' (2019) account, it resembles subsumption more than integration—though in the opposite direction. In order for the models to be integrated to serve enactive cognitive science, the offending parts of Mechanism are removed. Commitments to decomposition, to finding the organised causal entities producing a phenomenon, are abandoned by an appeal to strategic mechanism. In short, all the things that are supposed to make ontic or epistemic mechanistic explanations explanatory are absent.

Once again, I think explanatory goals are responsible for these difficulties. In order to achieve integration, the models concerned need to be able to meet the explanatory goals of the account of cognition in question—enactivism. Since, as Walmsley (2019) acknowledges, mechanism tends to run contrary to the goals of enactive cognitive science, its integration alongside dynamical models is made difficult. Hence the offending parts of mechanism—its ontological and epistemic requirements—need to be stripped out, leaving only a pragmatic, strategic mechanism that doesn't demand a mechanistic how–level be part of the explanation. While Walmsley's (2019) account succeeds, it does so in part by relinquishing mechanist commitments regarding the necessity of a causal–mechanistic description for explanation. As such, most mechanists will regard this as subsumption, not integration, wherein the form of explanation is fundamentally dynamical with some added mechanistic detail.

4.4 Conclusion

In this chapter I have argued that attempts to integrate dynamical and mechanistic models within enactive cognitive science are not successful integrations. This is largely because of differing explanatory goals: mechanistic explanation is designed to serve the explanatory goals of traditional cognitivism by providing a multi–level story of how the physical substrates of cognitive phenomena can be functionally (and causal–mechanically) decomposed and localised. Meanwhile, enactivism eschews decomposition and localisation in favour of dynamically explaining cognitive phenomena by describing the unfolding of extensive, constitutive relations between features of a system over time. Abramova & Slors' (2019) integrative proposal ends up in subsumption of dynamical models into mechanistic explanation, and hence is not genuine integration. Walmsley (2019) on the other hand subsumes mechanism into dynamical explanation, making a more enactive-friendly proposal but not one, I argue, that results in genuine integration.

Chapter 5

An Explanatory Taste for Mechanisms

Mechanistic explanations, according to some proponents, are objective explanations (Craver 2007, 2014). Mechanistic standards of explanation are derived objectively from nature, and are thereby insulated from the values of investigators, since explanation is an objectively defined achievement grounded in the causal structure of the world (Craver 2014). I call this position the closure of explanatory standards. I raise two problems with it. First, mechanists rely on several ontological claims which, while plausible, fail to guarantee the objectivity of mechanistic explanatory standards to the degree of certainty required. Second, mechanists themselves introduce a value–laden explanatory standard—the 3M requirement (Kaplan & Craver 2011)—which undermines the closure of explanatory standards. I show how in practice the standards of mechanistic explanation are guided by explanatory taste, shorthand for background contextual values that influence our standards of explanation. Ontic mechanists have a particular taste for control, and gerrymander explanatory standards in order to obtain it. Instead of being derived from nature, objectivity comes instead from intersubjective criticism, which renders visible the contextual values of communities of investigators and allows them to be controlled for (Longino 1990).

5.1 Introduction

The goal of separating the explanatory wheat from the merely descriptive chaff is a common feature of accounts of scientific explanation: what must a description do to graduate to the status of explanation? More often than not this task involves the codification of *standards of explanation*—the rules by which we judge explanatoriness. Historically, different accounts have required that explanations subsume observations under laws of nature (Hempel & Oppenheim 1948), unify phenomena (Kitcher 1989), or describe the underlying causal structure of the world that lead to a phenomenon's occurrence (Salmon 1984). Mechanism, my target in this paper, fits into this last category.

Originating with Craver (2007) one of the more influential strands of Mechanism argues that mechanistic models become explanations when they correctly and completely describe the mechanism responsible for a phenomenon. According to this account –*ontic mechanism*, sometimes called *mechanistic realism* (Craver & Kaplan 2011)–descriptions become *objective explanations* when they correspond to the real mechanisms that occupy nature (Craver 2007, 2014; Craver & Kaplan 2018). Explanations already exist in the world in the form of mechanisms; it is simply the scientist's job to describe them. Descriptive models therefore become explanations by dint of how closely and completely they represent this mechanistic explanation–in–the–world.

Craver (2014) extends this ontic picture by suggesting that by virtue of its objectivity, Mechanism is *the* guide to a cross–cultural, universal and objective means of causally explaining phenomena, at least within the special sciences (Craver 2014). There is consequently no possibility for explanatory standards to differ across times and places, across sociocultural contexts and with respect to the values and goals of investigators, other than as deviations from the correct mechanistic standards (Craver 2014). I call this position the *closure of explanatory standards.*

This line of thinking appears to motivate Craver's (2014) declaration of independence:

> ...my topic is independent of psychological questions about the kinds of explanation that human cognitive agents tend to produce or tend to accept. Clearly, people often accept as explanations a great many things that they should reject as such. And people in different cultures might have different criteria for accepting or rejecting explanations. These facts (if they are facts) would be fascinating to anthropologists, psychologists, and sociologists. But they are not relevant to the philosophical problem of stating when a scientific explanation ought to be accepted as such. In the view defended here, scientific explanation is a distinctive kind of

> achievement that cultures and individuals have to learn to make. Individual explanatory judgments, or cultural trends in such, are not data to be honored by a normative theory that seeks to specify when such judgments go right and when they go wrong. (Craver 2014, pg. 29)

Mechanistic standards of scientific explanation, on this view, are insulated from sociocultural, psychological and other influences traditionally viewed as non–scientific otherwise known as *contextual values* (Longino 1990). These contextual values are seen by ontic mechanism as an obstacle to the task of building an account of scientific explanation. In Craver's (2014) view, while the role of contextual values in science may indicate interesting vectors of research regarding how scientific explanation is done by value–driven investigators, and is potentially "fascinating to anthropologists, psychologists, and sociologists" (Craver 2014, pg. 29) it tells us nothing about what explanations actually are or how they ought to be done. Worse, by allowing contextual values to influence our standards of explanation, we potentially corrupt the hard–won independence of science from outside influences that are "not relevant to the philosophical problem of stating when a scientific explanation ought to be accepted as such" (Craver 2014, pg. 29).

With this ontic mechanistic account in mind, I pick out two potential lines of criticism. First, I probe at the ontic mechanists' claim that mechanistic standards of explanation are objective: does the underlying ontic argument support the claim, and give us good reasons to believe that mechanistic explanations are objective? Second, I test the closure of explanatory standards: do ontic mechanists provide good reasons to believe that mechanism's standards of explanation are thereby insulated from contextual values?

I argue on both counts there are substantial reasons to think not, at least based on the arguments offered by the ontic mechanist camp so far. Their account depends on a sequence of metaphysical claims that provide not just plausibility, but a guarantee that nature is mechanistic, and hence nature offers us objective mechanistic standards of explanation. Despite the centrality of this claim to the project of mechanistic objectivity, there is surprisingly little argumentation for it on offer—and while plausible, the existing arguments fail to provide the required guarantee. I also argue that the model–to–mechanism–mapping (3M) requirement endorsed by ontic mechanists as an additional explanatory standard (Kaplan & Craver 2011, Kaplan 2015) undermines the closure of explanatory standards, since in advancing 3M they incorporate an explicitly normative, value–laden standard of explanation into ontic mechanism.

This breakdown in the closure of explanatory standards reveals the value–laden character of mechanistic standards of explanation, which are constituted at least in part by contextual values. I propose that modes of explanation and their explanatory standards, the means by which we formulate and select successful explanations, are entangled with our *explanatory taste*. Explanatory taste is the product of the contextual values that impinge on our thinking about what the status of explanation requires. The selection of standards of explanation is typically an act of gerrymandering—the chosen standards allow us to elevate those descriptions we favour to the status of explanations and exclude those sorts of descriptions we have no taste for. Mechanism itself is a prime example of this explanatory taste at work, valuing descriptions that offer control over phenomena via the standard of completeness. Overall, I wish to present a cautionary perspective regarding the scope and limits of mechanistic explanation and suggest that its claims to objectivity and exclusivity are less than certain.

5.2 Mechanist explanation and objectivity

Mechanistic explanation is a popular account of causal explanation, with a particular focus on the cognitive and biological sciences. The basic explanatory unit in a mechanistic explanation is a *mechanism*, and its associated *phenomenon*. A mechanism is defined as "...a structure performing a function in virtue of its component parts, component operations, and their organization." (Bechtel & Abrahamsen 2005, pg. 423). It is this mechanism which produces, underlies or is otherwise responsible for the phenomenon.

Core to mechanistic explanation is the notion that explanations ought to speak to causes; they "must include reference to causal relationships if they are to distinguish good explanations from bad" (Craver 2007, pg. 8). Similarly important (and almost universal) in the contemporary mechanist literature is its appeal to interventionism (Woodward 2003) as a means of locating these causal relations. In brief, if an intervention on the value of X causes a change in the value of Y (all other things remaining equal) then it describes a causal relationship. These variables can stand in for various features of the world: the cell became cancerous because it mutated; the ion channel opened because of a potential difference across the membrane, etc. Describing these causal connections, which form the causal structure of the world, is ultimately what describes a mechanism.

5.2.1 The ontic argument & the aetiology of objectivity

Why does a description of a mechanism become an explanation? Some explanatory standards are required to answer that question. To that end, ontic mechanists advance an account that purports to present an objectively derived set of explanatory standards. They claim that Mechanism does not explain nature merely in the sense that its models are illuminating, enhance understanding or are useful to human endeavours:

> "Objective explanations are not texts; they are full–bodied things. They are facts, not representations" (Craver 2007, p. 27)

Good mechanistic explanations are *objectively* explanatory; and mechanistic explanations are *objective explanations* (Craver 2007). These objective explanations are meant to reveal the true mechanistic workings of nature that are waiting out there to be discovered; they shed light on objective ontic structures (Craver 2014). Ontic mechanists make a sequence of supporting claims to advance the notion that mechanisms provide objective explanations. These claims are ontological in nature; they assert that nature is mechanistic, comprised of mechanisms, and our explanations should track this mechanistic reality.

(1) Mechanisms are real

Mechanisms are real entities that really do produce or underlie the phenomena we observe in nature. They are not merely useful fictions constructed as a part of scientific investigation, but "things in the world" (Glennan & Illari 2018a, pg. 2). The view that mechanisms are real, and that this is significant for their explanatory potential, is broadly shared by prominent ontic mechanists, for instance Glennan & Illari (2018b):

> "...let us consider whether and in what sense these varieties of mechanisms are real. Our language has been realist. Mechanisms are things in the world..." (pg. 99).

Glennan's (2017) account of mechanism presents a similar view, where "...mechanisms and their constituents are things in the world that exist independently of the models we make of them" (pg. 10). Talking about mechanisms in cognitive science, Craver & Kaplan (2011) claim that ontic mechanists "expect to ground the taxonomy of cognitive and neural phenomena in objective facts about the mechanistic structure of the brain. By learning which mechanisms are distinct from which others, one can carve the mind–brain at its joints." (pg. 276).

This claim is important because it functions as a lynchpin of objectivity—it guarantees that mechanisms aren't merely forms of description, fictions or idealisations, but real things, and when we find them, we carve nature at its joints. This also points to something of a schism within mechanism between the ontic mechanists, and the epistemic mechanists (Illari 2013). These subdivisions support similar methods and explanatory standards, but provide different justifications for them.

Epistemic mechanists (Bechtel & Abrahamsen 2005; Bechtel 2015) are more concerned with how mechanistic explanations enhance understanding of phenomena; representing the ontic side, Craver (2007) is concerned with explaining by revealing the objectively existing causal structure of the world. The ontic mechanists are strict realists about mechanisms, while epistemic mechanists do not consider the ontological status of mechanisms to be important for articulating an account of mechanistic explanation.

As a consequence epistemic mechanists are criticised precisely because this view allegedly abandons anything except epistemic commitments, including objectivity (Craver 2014; Kaiser 2018). Because epistemic mechanism doesn‘t claim to identify real mechanisms, "it is too weak to serve as a guide to the norms that distinguish good explanations from bad and complete explanations from incomplete." (Craver 2007, pg. 28)

(2) Mechanisms are explanatorily obligatory

Following from (1), ontic mechanists also accept that mechanisms are not just one kind of explanatorily relevant thing among many that inhabit the world. Mechanisms are (at least within the ‘’special sciences") precisely what is relevant to explaining phenomena in those domains. There may be other things that figure in a mechanistic ontology, but the explanatorily relevant things are mechanisms (Craver 2014). Put another way, mechanisms are both necessary and sufficient for causal explanation, at least as far as cognitive and biological sciences are concerned.

This position allows non–mechanistic models to be utilised towards this end of describing mechanisms, since they are still ultimately means of getting at mechanisms (Kaplan 2015); on the other hand they fail at being explanatory if they are not mapped or grafted to a mechanism in some way. Consequently I take ontic mechanists to hold the position that mechanisms are explanatorily obligatory: descriptions of mechanisms, as well as supporting non–mechanistic models (like mathematical models) that help elucidate features of mechanisms, are necessary and sufficient for our objective explanatory purposes.

(3) Objective explanatory standards can be derived from nature.

If mechanisms really do produce or underlie phenomena, and descriptions thereof are explanatorily adequate, this then affords mechanists a straightforward way of obtaining standards of explanation. Mechanistic standards of explanation must be geared towards revealing these explanations in the world:

> The complete constitutive explanation for a given explanandum, in this ontic sense, includes everything relevant (that is, everything that makes

> a difference) to the precise phenomenon in question.(Craver & Kaplan 2018, pg. 14)

> Complete neuroscientific explanations are distinguished from incomplete explanations...by the fact that complete explanations capture all of the relevant causal relations among the components in a mechanism." (Craver 2007 pg. 61–62)

A key mechanistic standard of explanation is *completeness.* Completeness, according to ontic mechanists, is not a subjective or pragmatic criterion but rather a reference to a state of affairs in nature—what counts as complete is whatever correctly represents the objective mechanism comprehensively. Whatever allows us to get at the objective explanation waiting in the world, describe all the relevant causal and constitutive relations, and exclude irrelevant relations, are consequently the objective standards of explanation:

> Good mechanistic explanatory texts...are good in part because they correctly represent objective explanations. Complete explanatory texts are complete because they represent all and only the relevant portions of the causal structure of the world.(Craver 2007, pg. 27)

This condition of relevance also demonstrates the connection between completeness and the aforementioned notion of carving nature at its joints that ontic mechanists subscribe to. When a feature of the world is ruled to be relevant to an explanation, this means it really is a part of the mechanism producing the phenomenon of interest.

Hence completeness is the definitive mechanistic explanatory standard. Interventionism makes this standard achievable, since it permits investigators to find what features are relevant—and hence required for a complete description of a mechanism—and which features of the world are irrelevant. Completely describing (accurately) the causal structure of the world—a mechanism—underlying the explanandum phenomenon is what graduates a description to an explanation.

> (4) Mechanistic explanations are objective explanations.

If nature really is comprised of mechanisms, mechanisms are what we need to describe in order to explain, and our ways of explaining them are derived from the objective causal structure of nature, then mechanistic explanations ought to explain objectively.

I group these claims (1)—(4) together as the *ontic argument.*

5.2.2 Closure of explanatory standards, and a declaration

of independence One consequence of this account is the *closure of explanatory standards.* If explanatory standards are objectively derived from nature, then there is no room for anything else to determine them. Other potentially explanatorily relevant features of a model are not necessary for explanation. In addition, contextual values, goals, biases, background beliefs and so on simply have nowhere to get any purchase. These standards are already determined by objective, universally available facts about nature, whether we like it or not (Craver 2014).

Having established closure, Craver (2014) makes his declaration of independence, according to which the question of what explanations are (and what explanatory standards support them) is insulated from context. Contextual values are, in Craver's view, not relevant to how modes of explanations are formulated. Mechanisms are really out there producing phenomena and provide us with objective standards of explanation for those phenomena. People may mistakenly believe that context—the goals and values they want an account of explanation to support, for instance—can be decisive to the shape explanations take, but if Craver is to be believed, they are objectively wrong.

To make the position even starker, in Craver's view no features of an investigator or community of investigators can have any bearing standards of explanation, since mechanistic explanations and explanatory standards are stable across contexts. Anyone interested in explaining nature (at least where nature is comprised of mechanisms) must, to do the job properly, locate certain timeless, universal truths about mechanistic explanation in order to explain. Whether I am investigating the world in prehistory before the written word, or tapping on a keyboard in the 21st century the same standards apply; the only proper standards of explanation are mechanistic ones. Including anything else inserted into our explanatory standards serves only to muddy the waters.

5.3 Critique of mechanistic objectivity

This ontic mechanist argument raises some questions. First of all, Craver's (2014) claim about the closure of explanatory standards is phrased as a sort of guarantee. What underpins this guarantee of closure is the ontic argument. So, to be certain about the closure of explanatory standards, we need to be certain of the undergirding ontic argument. If the ontic argument is taken to really reflect *what is*, distinct from human pragmatic concerns, there must be some good reasons to believe the constituent claims.

It is important here to reiterate the direction of the justification for closure. The ontic mechanist claim is not that the explanatory successes of mechanism should make us confident of this mechanist ontology. The reasoning is the reverse: because we are already so certain about the ontic argument, we can derive from it a set of explanatory standards that guarantee objectivity; "the norms of scientific explanation fall out of a prior commitment on the part of scientific investigators to describe the relevant ontic structures in the world." (Craver 2014, pg. 41) The ontic argument is the bedrock upon which his account of objective explanation is built, at least according to Craver's characterisation.

The immediate problem with this position is that it requires an assumption—the ontic argument—to work as a guarantee. Assumptions are not guarantees, and so this "prior commitment" to a mechanistic ontology fails to provide the necessary guarantee of objectivity. But perhaps the ontic mechanistic account is able to offer further reasons to believe that a mechanistic ontology falls out of a mechanistic science.

Imagine the scenario that (1) were not known with certainty. If that were the case, then it seems that the mechanist conclusion—that mechanistic explanations are objective explanations—would be a little less certain. Unless we have good reasons to believe that nature is indeed made up of mechanisms, then there remains the plausible case where nature is not made up of mechanisms. In other words, if (1) were brought into doubt, the entire ontic argument undergirding the objectivity of mechanistic explanation would also be in trouble. If nature isn‘t really full of mechanisms producing phenomena, then there is no guarantee that striving for complete descriptions of mechanisms would lead us to an objective explanation. To be clear, these claims may still be of some epistemic value if either were shown to be uncertain on the ontic side of things. But critically they would not support (4).

I labour the point in order to make the stakes quite clear. The mechanist project puts a large stock in its objectivity. This is because if it fails to be objective, mechanism loses the edge it claims to have over other modes of explanation; it ceases to be the exclusive account of causal explanation. Most importantly, (though I leave this point for a later section) such a failure would undermine the closure of explanatory standards.

With these stakes in mind, I will review the evidence for these supporting claims, focusing mainly on (1). Let us also bear in mind the fairly high standards mechanists have set for themselves here. It is not just supposed to be likely, possible, or even just plausible that (for instance) mechanisms are real and explanatorily adequate. Claims (1) through (4) are meant to be such a certainty that mechanism can lock out all other considerations about explanatory standards.

5.3.1 The mechanist case

Craver (2007) holds to a straightforward realist argument in service of the ontological argument, specifically (1)—if we can observe entities engaged in causal relationships via ideal interventions, then we can take those things to be real. The obverse of this claim is that if we cannot empirically observe something through an intervention, it does not exist:

> There is no evidence that souls or entelechies exist...We cannot intervene with predictable outcomes to change souls and entelechies...we are justifiably suspicious of claims that such things exist. But none of these reasonable criteria fails for higher–level items in neuroscience. Molecules, neurons, brain regions, and brain systems all clearly satisfy these standards.(Craver 2007, pg. 15)

As Craver indicates here, interventionism is meant to allow the carving of nature at its joints, and provide confidence about the existence of mechanisms, since it gives us fairly clear criteria for when we have or have not observed entities engaged in a causal relationship. Being able to describe nature in terms of fine–grained relationships between causally–linked variables ought to reveal to us the true underlying structure of nature.

The ontic mechanist picture leans heavily on ideal interventions as a method that allows us to "carve nature at its joints" and locate the real entities that underlie phenomena. This also suggests that interventions and the causal relations they reveal are an objective feature of the world—that interventions give us neutral, observer–independent findings. But while interventionism is capable of doing plenty of heavy lifting for mechanism, it is not clear that it can perform this role as well.

5.3.2 Underdetermination

One concern with this realist argument is the plausibility of interventions alone delivering a mechanistic ontology. Craver's picture here seems to imagine interventions as a net that can be cast out into the world, which will bring in neutral, objective (and mechanistic) facts about nature, and which subsequently reinforce the idea that the ontic structure of the world is mechanistic. But the interpretation of that data gained via interventions as confirming or disconfirming a particular ontology, be it of neurons and brain regions or souls and entelechies, is decidedly not neutral in the way Craver requires.

Craver & Kaplan (2011) themselves raise this substantial problem with an intervention–generated ontology within the domain of cognitive science. They note that pre–existing theory and assumptions are decisive to how investigators interpret incoming data. For instance, the interpretation of interventions in the form of double dissociation experiments has permitted cognitive neuroscientists to localise particular brain functions within distinct brain regions. However, these findings depend on several assumptions, e.g. that the brain is functionally modular, and of the validity of the standard taxonomy of mental functions. Craver & Kaplan (2011) concede this potentially leads to a kind of circularity where "our ability to carve the brain into distinct mechanisms requires some idea of what those mechanisms do, and this requires some commitment about which capacities require explanations." (pg. 76)

This acknowledges the basic problem with Craver's realist argument, which resembles Quine's (1951, 1975) problem of underdetermination: data alone is not enough to objectively determine which of many possible interpretations of that data is correct. Any investigation of a phenomenon will bring with it pre–existing commitments that colour the interpretation of the results of any empirical test.

For instance, the ontic mechanist account assumes a necessary connection between interventions and mechanisms. When ontic mechanisms talk about locating causal relations, to them it necessarily follows that this relationship reveals (a portion of) a mechanism. But while interventionism is amenable to describing causal mechanisms, there is no necessary connection between the two. Woodward (2013a) specifies that non–mechanistic explanations can be performed using interventions as well. Much like Craver & Kaplan's (2011) double dissociation example, pre–existing assumptions are what binds together interventions and mechanisms. Although ontic mechanists assume that any successful intervention indicates a successfully located mechanism, "any use of interventions in science are epistemic and come only from a human recognizing the intervention as providing knowledge about a mechanism." (Machamer 2004, pg. 28).

Hence by themselves, interventions are typically insufficient to confirm or disconfirm the existence of particular entities in an objective manner. Since we are always entering into investigations of nature with hypotheses and theories that propose a certain state of affairs for testing (for instance, that nature is full of causal mechanisms, or that the brain is modular), the interpretation of interventions is always done through such a lens.

Consequently, a dependency on interventions to carve nature at its joints simply results in circularity. None of this is to say that interventionism cannot be used to investigate what kinds of entities might exist in the world, only that interventions alone cannot provide us with objective causal facts which we can then add right into the corpus of a mechanistic ontology.

5.3.3 Non–mechanistic reification

Another related problem for this realist argument is that interventionism can find causes without mechanisms, so long as one's interpretation of intervention data is non–mechanistic. Dynamical models, for instance, describe more abstract mathematical variables detached from mechanisms that can be intervened upon to establish causal relations in precisely the same way as mechanistic models (Meyer 2018). If the sole criterion of the ontic mechanist account to establish the existence of something is that a part of the world can be intervened upon, then the dynamics of a system ought to be just as much a part of our ontology. This would mean that a non–mechanistic part of the causal structure of the world can do explanatory work, and would suggest that part holds equal ontic status alongside more traditional mechanistic entities. Such an outcome does not square with a mechanistic ontology, which holds that mechanisms are the only explanatorily relevant things out there in nature.

Take for instance the oft–cited case of the Haken–Kelso–Bunz (HKB) model of bimanual coordination (Haken et al 1985). This dynamical model describes and predicts with great accuracy how bimanual coordination (in this instance, wagging the index fingers on both hands simultaneously) falls into in–phase and anti–phase patterns depending on the frequency of the oscillations. The model is described by the following differential equation:

$$\frac{d\phi}{dt} = -a\sin\phi - 2b\sin 2\,\phi$$

Where is relative phase, and the ratio b/a represents the inverse of the frequency of bimanual oscillations. Intervening on b/a causes changes in ϕ and hence the HKB model describes a causal dependency between these variables (Chapter 1). Consequently, this causal relationship forms part of a causal, dynamical explanation of bimanual coordination. Here we seem to meet Craver's realist requirements: since intervention data is all that is required to confirm that an entity belongs in our ontology, then these dynamical variables should be granted the same ontic status as, say, the opening or closing of ion channels during the neuronal action potential.

However, this would surely be counted as an unwanted reification of merely mathematical variables from the ontic mechanist perspective. An ontic mechanist might wish to reply by either suggesting that there is something more metaphysically fundamental or real about mechanistic components as opposed to a dynamical property of a system; or by adding new criteria to ensure that only canonical mechanistic entities can be granted this ontological status via interventions.

The former response has quite a challenge ahead of it. Mechanists hewing to this line of argument have to show how objects like neurons and brain regions are "more real" than dynamical features of a system without appealing to the outcomes of interventions. The latter response has proven the more popular one, defended in the form of the 3M requirement, which I will address later in Section 4.

5.3.4 Consequences for mechanistic objectivity

With the certainty of mechanistic objectivity in question, Craver's (2014) declaration of independence seems to be on shakier foundations. Now that the claims undergirding the objectivity of mechanism are shown to be more a set of assumptions than ontological certainties upon which an objective mode of explanation can hang, the closure of explanatory standards can be broken open.

5.4 The 3M requirement

I have just raised a few concerns with the guarantee (or lack thereof) of ontic mechanism's undergirding assumptions. In this section, I will raise another distinct problem with the account. So far, I have dealt with that ontic account's reliance on uncertain ontological claims for objectivity. Here I investigate how the ontic mechanist account undermines its own claim to objectivity by incorporating a normative explanatory standard—the 3M requirement.

5.4.1 The requirement

The model–to–mechanism–mapping (3M) requirement grew out of debates surrounding the explanatory power of non–mechanistic, dynamical models in the cognitive sciences. Dynamicists claimed that the descriptive and predictive successes of dynamical models elevated them to the status of genuine explanations (Stepp et al 2011, Chemero & Silberstein 2008). Dynamicists appealed to various arguments to this effect, including the equivalence of prediction with explanation, and adherence to a covering–law model of scientific explanation (Walmsley 2008). More recently I have used the very same interventionist method core to mechanistic explanation as an argument for causal dynamical explanations (Chapter 1).

In response to the encroachment of the increasingly popular notion of dynamical explanations into cognitive science, mechanists (Kaplan 2011, Kaplan & Craver 2011) devised 3M. It requires that, in order to be an explanation, any model must show how its features (for instance, the variables in a dynamical model) "map onto" the components of an underlying mechanism. The assumption expressed here is that any causal relationship in a dynamical model actually reduces down to a causal relation present in the underlying mechanism:

> (3M) A model of a target phenomenon explains that phenomenon to the extent that (a) the variables in the model correspond to identifiable components, activities, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these (perhaps mathematical) variables in the model correspond to causal relations among the components of the target mechanism." (Kaplan 2011, pg. 347)

By introducing 3M, Kaplan & Craver (2011) create an additional explanatory standard that works to exclude potential dynamical (or otherwise non–mechanistic) explanations that meet the existing interventionist requirements to be a causal explanation. No non–mechanistic model, at least within certain domains of science, can provide us with genuine causal explanations along interventionist lines if 3M is to be believed. This presents a serious challenge to the possibility of causal yet non–mechanistic explanations. If any apparent dynamical explanation necessarily reduces down to a mechanistic explanation, then dynamical explanations are not an explanatory alternative and are subsumed under mechanism.

However, 3M is interesting in that Kaplan & Craver (2011) explicitly acknowledge its normative character. They claim that 3M "is justified in part because it makes sense of scientific–commonsense judgments about the norms of explanation" (Kaplan & Craver 2011, pg. 602):

> The 3M constraint is the mechanist's gauntlet: a default assumption that the phenomena of cognitive and systems neuroscience have mechanistic explanations, like so many other phenomena in the special sciences, and that cognitive and systems neuroscientists ought to (and often do) demand that explanations reveal the mechanisms underlying the phenomena they seek to explain. (Kaplan & Craver 2011, pg. 603–604)

Their case against the dynamicists, they claim, is justified because dynamicists who defy or reject 3M fail to meet the ''scientific commonsense" judgement that good explanations "correctly identify features of the causal structures that produce, underlie, or maintain the explanandum phenomena." (Kaplan & Craver 2011, pg. 607). It is this appeal to the normative, the "commonsense", that comes back to bite the ontic mechanist account by undermining the closure of explanatory standards.

5.4.2 The problem of 3M

Recall the reasoning behind the closure of explanatory standards: because of their objective character, mechanistic standards of explanation are immune to all sociocultural and otherwise normative judgements. These kinds of judgements only reflect the contextual values of investigators and cannot be admitted to the genuinely objective set of explanatory standards.

Yet, 3M is explicitly just this kind of thing. It is a mechanistic explanatory standard, advanced as part of the ontic mechanist project, yet it is explicitly normative in character. Its justification is decidedly contextual—as Kaplan & Craver (2011) point out when they suggest that 3M is justified as an additional explanatory standard because it represents the judgements or expectations of a community of scientists and philosophers of science; it helps exclude potential explanations that do not suit their values or goals.

It seems difficult, then, to reconcile Craver's (2014) declaration of independence with the actual explanatory standards ontic mechanists abide by and endorse. On the one hand, he claims that mechanistic standards of explanation are objectively derived from nature, immune to normative criteria. On the other, Craver and Kaplan are happy to include a normative criterion with the same ability to determine explanatory power (or lack thereof) as the existing explanatory standards. Even if there were strong reasons to believe that mechanistic standards of explanation were objective, the inclusion of 3M serves to undermine confidence in that claim.

5.5 Explanatory taste

The ontic mechanist claim that objectivity is a bulwark against the intrusion of contextual values into their standards of explanation is not well–founded. Their use of uncertain ontic claims to guarantee the certainty of the closure of explanatory standards, coupled with the normative character of 3M, undermines the ontic mechanist guarantee that mechanistic explanations are objective explanations, outside of time and place.

How to understand these mechanistic explanatory standards, then, if not as spawned straight from objectivity? To that end, I introduce the notion of *explanatory taste*. The term refers to the taste—emphasising its contextual nature—of investigators both scientific and philosophical for certain sorts of explanation. To this end, philosophers of science habitually gerrymander explanatory standards in order to include into canon those things they see as proper explanations, and exclude those they do not.

To understand what is considered explanatory versus non–explanatory—what investigators have a taste for and what they do not—it becomes necessary to look at the uses investigators have in mind for explanations. The demarcation line for explanations is typically drawn via explanatory standards along pragmatic lines, namely the perceived end goal of a given mode of inquiry into nature. To return to a historical case, Hempel's (1965) account of covering law explanations was engineered to provide the desired *understanding* of phenomena, where understanding was equated with having shown how events in nature were to be expected, where that expectation is provided by a deductive or inductive argument. Logical positivists or empiricists like Hempel had a particular explanatory taste that reflected the broader goals of their project: understanding nature in terms of a series of verifiable statements.

5.5.1 The contextual values of mechanism

Ontic mechanists, similarly, are concerned with explanations that serve particular purposes, though theirs differ from those motivating the covering law model. They are interested not with arguments that provide understanding, but with explanations that yield practical control in the biomedical sciences and better our ability to control biology and cognition. 3M reveals explicitly the "commonsense judgements" that motivate the gerrymandering of explanatory standards. Dynamical explanations are judged the wrong sort of thing to be genuinely explanatory. They are not to mechanist tastes, and 3M is therefore summoned to bar the gates against a dynamical encroachment on explanation.

In this section I argue, quite to the contrary of the ontic mechanist refrain, that mechanism is heavily influenced by the contextual values of investigators. In fact, the centrality of various contextual values to mechanistic explanation is a pervasive theme. Consider a few exemplary quotes from Craver's (2007) canonical account of ontic mechanism:

> One way to justify the norms that I discuss is by assessing the extent to which those norms produce explanations that are *potentially useful for intervention and control.* While this is not the only touchstone that one might use, it is nonetheless one, and it is objective." (Craver 2007 pg. x, emphasis added)

> ...I develop a view of explanation that does justice to the exemplars of explanation in neuroscience and to the standards by which these explanations are evaluated." (Craver 2007, pg. 1–2)

> If neuroscience succeeds in this...goal, it will open medical possibilities that now seem like science fiction, and it will provide human beings

> (for good or ill) with *new and powerful forms of control over the human condition.*" (Craver 2007, pg. 2, emphasis added)
>
> ...unlike some areas of fundamental physics, the search for neuroscientific explanations is driven by goals of treating illnesses, improving brain function, preventing cognitive decline, and developing new ways to manipulate and record from the brain in the laboratory. *Explanation is a tool for determining how to intervene into the brain and manipulate it for our various purposes.*" (Craver 2007, pg. 38, emphasis added)
>
> To repeat a central theme: causal relevance, explanation, and control are intimately connected with one another. This is particularly true in biomedical sciences, such as neuroscience, that are driven not merely by intellectual curiosity about the structure of the world, but more fundamentally by the desire (and the funding) to cure diseases, to better the human condition, and to make marketable products." (Craver 2007, pg. 93)

There are clear themes running through Craver's rationale for an account of mechanistic explanation: concerns about human control over biomedical phenomena; pragmatic concerns about efficacy in medicine; the onerous matter of obtaining funding; justification of certain hypotheses for pragmatic ends; a confidence in neuroscience and what neuroscientists consider to be canonical explanations.

Most telling is the way Craver speaks about explanation and the pragmatic goal of *control*. Mechanistic explanation is, as Craver says above, not driven merely by "intellectual curiosity" but by a need to control phenomena "for our various purposes"; "[s]uch explanations are useful precisely because they identify loci in a mechanism that can be commandeered for the purposes of control." (Craver 2007 pg. 200). This notion of control, of explanation as an inspection that yields various levers for the fine–grained manipulation of nature, is a mechanist contextual value which finds itself reflected in mechanistic standards of explanation. Control is valued—it is to the explanatory tastes of ontic mechanists—precisely because it is perceived to serve these pragmatic ends.

I argue that the explanatory standard of completeness results from this taste for control. Completeness is born of the contextual values of mechanist–oriented investigators: specifically, the goal of control over a phenomenon. This connection between the two notions has not gone unacknowledged:

> The pragmatic import of developing norms of mechanistic completeness links to the fact that mechanistic explanations often provide the rationale for developing technologies for gaining control over phenomena, such as

> experimental techniques and medical treatments. (Baetu 2015, pg. 779–780)

Ontic mechanists want explanations that afford control over certain phenomena in certain ways—this is their explanatory taste. Mechanists themselves, as I have illustrated, frequently stress the tight relationship between their pragmatic goal of control, and the explanatory standard of completeness. To that end, they have developed standards of explanation, specifically completeness, designed to equate control over nature with explanation.

Ontic mechanists also conceptualise control a certain way, as mapping to a certain picture of the world occupied by mechanistic entities and their activities. Control is by definition achieved through interventions (or at least the possibly of interventions) on these mechanisms. While dynamical models arguably offer similar levels of control over their target phenomena, the ontological commitments that are foundational to the ontic mechanist project simply do not admit of this competition. Mechanists do not deny that dynamical models are capable of impressive description and prediction of phenomena, but they certainly do deny their ontological import and hence their capacity to exhibit genuine control. To admit that dynamical models afford control over phenomena is tantamount to denying a mechanistic ontology, which does not allow non–mechanistic features of the world to be considered explanatory. If this were allowed, it would be the same as admitting that mechanistic explanations do not uniquely capture an objective reality. Mechanistic explanations would accordingly no longer be stable across contexts by dint of their uniquely objective correspondence to nature and hence surrender their objectivity.

5.5.2 What becomes of objectivity?

I have given reasons to reject the notion that mechanistic explanations are objective explanations, at least in the manner prescribed by ontic mechanists. In doing so I have deferred to a standard view of objectivity, where the objectivity of science means value–freedom, impartiality and the absence of perspective. Objectivity in this sense "is bound up with questions about the truth and referential character of scientific theories, that is, with scientific realism" (Longino 1990, pg. 62). Objectivity is achieved when science accurately reveals the real world, when "it is a correct view of the objects to be found in the world and of their relations with each other." (Longino 1990, pg. 62); its criteria for theory selection and success are "nonarbitrary and nonsubjective" (Longino 1990, pg. 62). This is the notion of objectivity preferred by the ontic mechanists.

There is however another notion of objectivity that is detached from these commitments. On this alternative view of objectivity, it is accepted that science cannot be purged of the prior commitments or perspectives of investigators—science is not value–free, impartial and absent perspective, and so "it would be a mistake to identify the objectivity of scientific methods with their empirical features alone" (Longino 1990 pg. 75). In the case of ontic mechanism, it is a mistake to think interventions provide objectivity. Between data and hypothesis—intervention and mechanism—there remains an interpretive gap where values determine the sorts of hypotheses that are selected, and the standards by which they are selected.

Instead objectivity can be thought of as an achievement founded on intersubjective criticism. When the commitments of a community of investigators are subjected to public critique from different critical perspectives, where those perspectives represent communities holding to different contextual values, it makes those commitments visible and hence evaluable:

> When...background assumptions are shared by all members of a community, they acquire an invisibility that renders them unavailable for criticism. They do not become visible until individuals who do not share the community's assumptions can provide alternative explanations of the phenomena without those assumptions, as, for example, Einstein could provide an alternative explanation of the Michelson–Morley interferometer experiment. Until such alternatives are available, community assumptions are transparent to their adherents. In addition, the substantive principles determining standards of rationality within a research program or tradition are for the most part immune to criticism by means of those standards.(Longino 1990, pg. 80)

A critical alternative brings to light the contextual values of the community. It does not purge science of those values but makes them explicit and open to discussion and revision. Objectivity is a condition that arises when these biases have been evaluated, criticised, and controlled for. Hence objectivity is achieved not in spite of the inherently value–laden character of scientific inquiry—those features that Craver (2014) suggests might be "fascinating to anthropologists, psychologists, and sociologists", but irrelevant to objective scientific explanation—but precisely because of it.

Mechanist criticism of dynamical explanation (Kaplan 2015) has illuminated the otherwise invisible commitments of dynamicists—for instance their explanatory taste for prediction—and spurred reappraisal (see Chapters 1 and 2), further proof that "[c]riticism is thereby transformative" (Longino 1990, pg. 73). Ontic mechanism would therefore do well to open up to this sort of critique. Abandoning Craver's declaration of independence, as well as the closure of explanatory standards, is necessary to permit such criticism of the values of mechanism to be aired out. The alternative account of dynamical explanation I have provided in Chapters 1 and 2 is also serves as a criticism of many conceptual commitments of mechanists, showing how the standards used to separate genuinely explanatory mechanistic models from merely descriptive dynamical models fail to do so. While the ontic mechanist could respond that no matter the criticism, ontic mechanistic explanatory standards are objectively correct, this would be a misstep.

There is another reason beyond, to again borrow Craver's phrase, intellectual curiosity, as to why should we be particularly concerned about this process of criticism. Longino's (1990) discussion is motivated by the very real problem of biases in science that entrench or reinforce social inequalities, for instance conceptual problems in research on the biology or cognitive basis of sex and gender differences, as well as criticism thereof by feminist philosophers of science. Making the assumptions of researchers explicit through external, public criticism, generated by a community who do not share those assumptions, preserves the successes of science while also subjecting its assumptions to illuminating criticism. These criticisms can and have driven science to produce less biased and more conceptually sound research (Longino 1990). For instance, feminist critique of sex determination literature shows how background assumptions about "active maleness" versus "passive femininity" curtail a more comprehensive understanding of the phenomenon (Beldecos et al 1988). My goal is not to accuse contemporary Mechanism of holding to pernicious biases, but rather to caution against eliminating the possibility of fair and open criticism which acts as prophylaxis against them.

5.6 Conclusion

In this paper I have criticised the claim that mechanistic explanations are objective explanations, and that mechanistic standards of explanation are closed to criticism because they are objectively derived from nature. To this end I targeted the underdeveloped justification for this objectivity, in particular a reliance on revelations of a mechanistic ontology via interventionism, which fails to guarantee objectivity as required. I also showed how the closure of explanatory standards is undermined when ontic mechanists introduce the normative explanatory standard of 3M. Against the prevailing ontic mechanist position, I have argued that mechanistic standards of explanation are influenced heavily by explanatory taste; the contextual values and goals of investigators. A prime example of explanatory taste at work is the mechanist taste for control, in pursuit of which they gerrymander mechanistic standards of explanation. I also suggest that, per Longino (1990), criticism between communities holding to different contextual values is a prerequisite for objectivity, and hence mechanism would be better served by opening up to conceptual criticism.

Conclusion

This thesis has argued for a dynamical alternative for causal explanation, and against attempts to subsume dynamical explanations (and its E-cognition proponents) into mechanism (and cognitivism). It also argued against the integrative rationale of Mechanism, which I identified as the supposed objectivity of mechanistic explanations.

In Chapters 1 & 2, I demonstrated how Woodward's (2003) interventionist account can be put to work producing non–mechanistic explanations. This builds upon and improves the earlier sketches of dynamical explanation, providing an account of causal dynamical explanations in the special sciences. I also provided some reasons to reject the 3M requirement in some cases.

Chapters 3 & 4 discussed related issues of integration. In Chapter 3 I argued that Miłkowski et al's (2018) integrative account of cognitive science, based on mechanistic explanation, was not a suitable replacement for theories of cognition. Nor was it able to subsume dynamical explanation. Chapter 4 covered two accounts that attempted to integrate mechanistic explanations into enactive cognitive science. These accounts, however, both resulted in a form of subsumption rather than genuine integration. I conclude that, so far at least, no genuine integration of these research traditions—or modes of explanation—has been successfully outlined.

Returning from this foray into the philosophy of cognitive science, Chapter 5 dug deeper into the ontological foundations of mechanistic explanation. In it, I argued that ontic mechanisms claim to produce objective explanations based on objective explanatory standards is not well-founded. I brought discussions of mechanistic objectivity into contact with contextual empiricist account of scientific knowledge (Longino 1990), and argued that the goals, values and background assumptions of mechanist investigators give rise to a particular explanatory taste on their part. I suggested that genuine objectivity is not achieved through completeness and accuracy though, but through inter–subjective criticism.

I hope that this thesis has achieved its goals. It has not been my intention to undermine mechanism as a mode of explanation, but rather to recontextualise it—to build up an alternative account of dynamical explanation, to show how this alternative feeds into the special sciences in a way that cannot simply be folded into mechanism, and, lastly, to critique some of the mechanistic rationale driving claims of integration, exclusivity and objectivity.

Where to next? The terrain from here is quite open, but a few key discussions stick out. Previously wherever we have had the need for causal explanation, we had also had the need for mechanistic explanation. But if the two do not totally overlap—some causal explanations can be non–mechanistic—where, then, is mechanistic explanation appropriate? What are its natural boundaries, if indeed it has them? I think this can be fruitfully studied in terms of the explanatory goals and needs of particular theories or frameworks. I have sketched a few examples of causal explanation outside the borders of mechanism: a more comprehensive story is required.

It would also useful to investigate whether dynamical explanations should pursue (to use the mechanist nomenclature) ontic or epistemic explanations. Are some cognitive and biological phenomena, in an ontic sense, dynamical rather than mechanistic? Discussions around substance versus process ontologies (Nicholson & Dupre 2018) seem a useful resource here, and also feeds into often ontologically–loaded talk about the nature of cognition and explanation of it.

Bibliography

Abrahamsen, A. & Bechtel, W. (2012). From reactive to endogenously active dynamical conceptions of the brain. In K. S. Plaisance & T. A. Reydon (Eds.), *Philosophy of behavioral biology* (pp. 329–366). Dordrecht: Springer Netherlands.

Abramova, E. & Slors, M. (2019). Mechanistic explanation for enactive sociality. *Phenomenology and the Cognitive Sciences*, *18*(2), 401–424.

Adams, F. & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*(1), 43–64.

Aizawa, K. (2015). What is this cognition that is supposed to be embodied? *Philosophical Psychology*, *28*(6), 755–775.

Andersen, H. (2016). Complements, not competitors: Causal and mathematical explanations. *The British Journal for the Philosophy of Science*, axw023.

Auvray, M., Lenay, C. & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*(1), 32–47.

Auvray, M. & Rohde, M. (2012). Perceptual crossing: The simplest online paradigm. *Frontiers in Human Neuroscience*, *6*, 181.

Baetu, T. M. (2015). The completeness of mechanistic explanations. *Philosophy of Science*, *82*(5), 775–786.

Barack, D. L. (2019). Mental machines. *Biology & Philosophy*, *34*(6), 63.

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46.

Baumgartner, M. (2009). Interventionist causal exclusion and non-reductive physicalism. *International Studies in the Philosophy of Science*, *23*(2), 161–178.

Baumgartner, M. (2010). Interventionism and epiphenomenalsim. *Canadian Journal of Philosophy*, *40*(3), 359–383.

Baumgartner, M. & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731–756.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science*, *78*(4), 533–557.

Bechtel, W. (2016). Investigating neural representations: The tale of place cells. *Synthese*, *193*(5), 1287–1321.

Bechtel, W. (2017). Explicating top-down causation using networks and dynamics. *Philosophy of Science*, *84*(2), 253–274.

Bechtel, W. (2018). The importance of constraints and control in biological mechanisms: Insights from cancer research. *Philosophy of Science*, *85*(4), 573–593.

Bechtel, W. (2019). Analysing network models to make discoveries about biological mechanisms. *The British Journal for the Philosophy of Science*, *70*(2), 459–484.

Bechtel, W. & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421–441.

Bechtel, W. & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A*, *41*(3), 321–333.

Bechtel, W. & Abrahamsen, A. A. (2013). Thinking dynamically about biological mechanisms: Networks of coupled oscillators. *Foundations of Science*, *18*(4), 707–723.

Bechtel, W. & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge, Mass: MIT Press.

Bechtel, W. & Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information processing mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322.

Beldecos, A., Bailey, S., Gilbert, S., Hicks, K., Kenschaft, L. & Niemczyk, N. (1988). The importance of feminist critique for contemporary cell biology. *Hypatia*, *3*(1), 61–76.

Bokulich, A. (2011). How scientific models can explain. *Synthese*, *180*(1), 33–45.

Bokulich, A. (2017). Models and explanation. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 103–118). Cham: Springer International Publishing.

Boone, W. & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, *193*(5), 1509–1534.

Bressler, S. L. & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290.

Brigandt, I., Green, S. & O'Malley, M. (2016). Systems biology and mechanistic explanation. *The Routledge Handbook of Mechanisms and Mechanical Philosophy*.

Buhrmann, T., Di Paolo, E. A. & Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. *Frontiers in Psychology*, *4*.

Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162.

Carruthers, P. (2017). Mindreading in adults: Evaluating two-systems views. *Synthese*, *194*(3), 673–688.

Chemero, A. (2000). Anti-representationalism and the dynamical stance. *Philosophy of Science*, *67*(4), 625–647.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, Massachussets: MIT Press.

Chemero, A. & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science*. *Philosophy of Science*, *75*(1), 1–27.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese*, *191*(2), 127–153.

Chirimuuta, M. (2017). Explanation in computational neuroscience: Causal and non-causal. *The British Journal for the Philosophy of Science*.

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*, 3–27.

Constant, A., Clark, A. & Friston, K. (2019). *Representation wars: Enacting an armistice through active inference*.

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford : New York : Oxford University Press: Clarendon Press.

Craver, C. F. (2008). Structures of scientific theories. In P. Machamer (Ed.), *The blackwell guide to the philosophy of science*. New Jersey, NJ: Blackwell Publishing.

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: The case of biology and history* (pp. 27–52). Springer Verlag.

Craver, C. F. (2016). The explanatory power of network models. *Philosophy of Science*, *83*(5), 698–709.

Craver, C. F. & Kaplan, D. M. (2011). Towards a mechanistic philosophy of neuroscience. In S. French & J. Saatsi (Eds.), *Continuum companion to the philosophy of science* (p. 25).

Craver, C. F. & Kaplan, D. M. (2018). Are more details better? on the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*.

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science*, *75*(5), 958–969.

Davila-Velderrain, J., Martinez-Garcia, J. C. & Alvarez-Buylla, E. R. (2015). Modeling the epigenetic attractors landscape: Toward a post-genomic mechanistic understanding of development. *Frontiers in Genetics*, *6*.

De Jaegher, H. & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

De Jaegher, H., Di Paolo, E. & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447.

De Jaegher, H. & Froese, T. (2009). On the role of social interaction in individual agency. *Adaptive Behavior*, *17*(5), 444–460.

Di Paolo, E., Bhurman, T. & Barandiaran, X. (2017). *Sensorimotor life: An enactive proposal*. Oxford University Press.

Di Paolo, E., Rohde, M. & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, *26*(2), 278–294.

Diedrich, F. J., Highlands, T. M., Spahr, K. A., Thelen, E. & Smith, L. B. (2001). The role of target distinctiveness in infant perseverative reaching. *Journal of Experimental Child Psychology*, *78*(3), 263–290.

Dings, R. (2019). The dynamic and recursive interplay of embodiment and narrative identity. *Philosophical Psychology*, *32*(2), 186–210.

Dupré, J. (2013). Living causes. *Aristotelian Society Supplementary Volume*, *87*(1), 19–37.

Enver, T., Pera, M., Peterson, C. & Andrews, P. W. (2009). Stem cell states, fates, and the rules of attraction. *Cell Stem Cell*, *4*(5), 387–397.

Ferrell, J. (2012). Bistability, bifurcations, and waddington's epigenetic landscape. *Current Biology*, *22*(11), R458–R466.

Froese, T. & Di Paolo, E. A. (2010). Modelling social interaction as perceptual crossing: An investigation into the dynamics of the interaction process. *Connection Science*, *22*(1), 43–68.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

Gallagher, S. (2018). Building a stronger concept of embodiment. In A. Newen, L. De Bruin & S. Gallagher (Eds.), *The oxford handbook of 4e cognition* (pp. 353–367). Oxford: Oxford University Press.

Gervais, R. & Weber, E. (2011). The covering law model applied to dynamical cognitive science: A comment on Joel Walmsley. *Minds and Machines*, *21*(1), 33–39.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Houghton Mifflin.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*(1), 49–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, *69*, S342–S353.

Glennan, S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research*, *81*(2), 362–381.

Glennan, S. (2015). *Mechanisms and mechanical philosophy* (P. Humphreys, Ed.). Oxford University Press.

Glennan, S. (2017). *The new mechanical philosophy.* Oxford University Press.

Goldman, A. I. (2012). A moderate approach to embodied cognitive science. *Review of Philosophy and Psychology*, *3*(1), 71–88.

Graf, T. & Enver, T. (2009). Forcing cells to change lineages. *Nature*, *462*(7273), 587–594.

Griffiths, P. E., Pocheville, A., Calcott, B., Stotz, K., Kim, H. & Knight, R. (2015). Measuring causal specificity. *Philosophy of Science*, *82*(4), 529–555.

Haken, H. (2008). *Brain dynamics.* Springer series in synergetics. New York ; London: Springer.

Haken, H., Kelso, J. A. S. & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.

Haken, H. (1983). *Advanced synergetics: Instability hierarchies of self-organizing systems and devices.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Hempel, C. G. (1965). Aspects of scientific explanation. *Journal of Symbolic Logic*, *37*(4), 747–749.

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

Hildebrand, T. (2013). Can primitive laws explain? *Philosophers' Imprint*, *13*, 1–15.

Huang, S. (2009). Reprogramming cell fates: Reconciling rarity with robustness. *BioEssays*, *31*(5), 546–560.

Huang, S. (2012). The molecular and mathematical basis of waddington's epigenetic landscape: A framework for post-darwinian biology? *BioEssays*, *34*(2), 149–157.

Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical Review Letters*, *94*(12), 128701.

Huang, S., Guo, Y.-P., May, G. & Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, *305*(2), 695–713.

Huang, S. & Ingber, D. E. (2007). A non-genetic basis for cancer progression and metastasis: Self-organizing attractors in cell regulatory networks. *Breast Disease*, *26*(1), 27–54.

Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese*, *177*(2), 213–245.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265–288.

Hutto, D. D. (2005). Knowing what? radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences*, *4*(4), 389–405.

Hutto, D. D. & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content.* Cambridge, Mass: MIT Press.

Hutto, D. D. & Myin, E. (2017). *Evolving enactivism: Basic minds meet content.* MIT Press.

Illari, P. (2013). Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, *78*, 237–255.

Jaeger, J. & Monk, N. (2014). Bioattractors: Dynamical systems theory and the evolution of regulatory processes: Bioattractors. *The Journal of Physiology*, *592*(11), 2267–2281.

Jones, J. (2015). *From pattern formation to material computation.* Emergence, Complexity and Computation. Cham: Springer International Publishing.

Kaplan, D. M. & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanations? *Topics in Cognitive Science*, *3*(2), 438–444.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339–373.

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy*, *27*(4), 545–570.

Kaplan, D. M. (2015). Moving parts: The natural alliance between dynamical and mechanistic modeling approaches. *Biology & Philosophy*, *30*(6), 757–786.

Kaplan, D. M. & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, *78*(4), 601–627.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior.* Cambridge, Mass: MIT Press.

Kirchhoff, M. D. & Meyer, R. (2019). Breaking explanatory boundaries: Flexible borders and plastic minds. *Phenomenology and the Cognitive Sciences*, *18*(1), 185–204.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (Vol. 8, pp. 410–505). Minneapolis: University of Minnesota Press.

Kiverstein, J. (2012). The meaning of embodiment. *Topics in Cognitive Science*, *4*(4), 740–758.

Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science*, *48*(3), 453–468.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* University of Chicago Press.

Lamb, M. & Chemero, A. (2014). Structure and application of dynamical models in cognitive science, 6.

Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science*, *64*(3), 485–511.

Lange, M. (2009). *Laws and lawmakers: Science, metaphysics, and the laws of nature.* Oxford ; New York: Oxford University Press.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*(4), 421–431.

Levy, A. (2013). Three kinds of new mechanism. *Biology & Philosophy*, *28*(1), 99–114.

Longino, H. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry.* Princeton University Press.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science*, *18*(1), 27–39.

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

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* Cambridge, Mass: MIT Press.

Matthiessen, D. (2017). Mechanistic explanation in systems biology: Cellular networks. *British Journal for the Philosophy of Science*, *68*(1), 1–25.

Menzies, P. (2012). The causal structure of mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences.* Causality in the Biomedical and Social Sciences, *43*(4), 796–805.

Meyer, R. (2018). The non-mechanistic option: Defending dynamical explanations. *The British Journal for the Philosophy of Science.*

Miłkowski, M. (2013). *Explaining the computational mind.* Cambridge, Massachusetts: The MIT Press.

Miłkowski, M., Clowes, R., Rucinska, Z., Przegalinska, A., Zawidzki, T., Krueger, J., . . . Hohol, M. (2018). From wide cognition to mechanisms: A silent revolution. *Frontiers in Psychology*, *9*, 2393.

Moris, N., Pina, C. & Arias, A. M. (2016, November 1). Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*, *17*(11), 693–703.

Nicholson, D. J. & Dupré, J. A. (2018). *Everything flows: Towards a processual philosophy of biology.* Oxford University Press.

O'Regan, J. K. & No´e, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939–973.

Piccinini, G. & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, *183*(3), 283–311.

Potochnik, A. (2015). The diverse aims of science. *Studies in History and Philosophy of Science Part A*, *53*, 71–80.

Quine, W. V. (1951). Main trends in recent philosophy: Two dogmas of empiricism. *The Philosophical Review*, *60*(1), 20.

Quine, W. V. (1975). On empirically equivalent systems of the world. *Erkenntnis*, *9*, 313–328.

Ross, L. N. (2015). Dynamical models and explanation in neuroscience. *Philosophy of Science*, *82*(1), 32–54.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world.* Princeton University Press.

Salmon, W. C. (1989). *Four decades of scientific explanation.* Pittsburgh: University of Pittsburgh Press.

Schenz, D., Shima, Y., Kuroda, S., Nakagaki, T. & Ueda, K.-I. (2017). A mathematical model for adaptive vein formation during exploratory migration of physarum polycephalum: Routing while scouting. *Journal of Physics D: Applied Physics*, *50*(43), 434001.

Segundo-Ortin, M., Heras-Escribano, M. & Raja, V. (2019). Ecological psychology is radical enough: A reply to radical enactivists. *Philosophical Psychology*, *32*(7), 1001–1023.

Silberstein, M. & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science*, *80*(5), 958–970.

Skillings, D. J. (2015). Mechanistic explanation of biological processes. *Philosophy of Science*, *82*(5), 1139–1151.

Slors, M., Bruin, L. d. & Strijbos, D. (2015). *Philosophy of mind, brain and behaviour.* Amsterdam: Boom.

Smith, L. B. & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348.

Sporns, O. (2010). *Networks of the brain*. The MIT Press.

Stepp, N., Chemero, A. & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, *3*(2), 425–437.

Tero, A., Takagi, S., Saigusa, T., Ito, K., Bebber, D. P., Fricker, M. D., ... Nakagaki, T. (2010). Rules for biologically inspired adaptive network design. *Science*, *327*(5964), 439–442.

Thelen, E., Schöner, G., Scheier, C. & Smith, L. B. (n.d.). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*(1), 1–34.

Van Gelder, T. (1995). What might cognition be if not computation? *Journal of Philosophy*, *92*(7), 345–81.

van Eck, D. (2018). Rethinking the explanatory power of dynamical models in cognitive science. *Philosophical Psychology*, *31*(8), 1131–1161.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*(5), 615–28.

Varela, F., Thompson, E. & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

Waddington, C. (1957). *The strategy of the genes*. London: George Allen & Unwin.

Walmsley, J. (2008). Explanation in dynamical cognitive science. *Minds and Machines*, *18*(3), 331–348.

Walmsley, L. D. (2019). Lessons from a virtual slime: Marginal mechanisms, minimal cognition and radical enactivism. *Adaptive Behavior*.

Ward, D., Silverman, D. & Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi*, *36*(3), 365–375.

Waters, C. K. (2007). Causes that make a difference. *Journal of Philosophy*, *104*(11), 551–579.

Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese*, *183*(3), 313–338.

Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science*, *64*(4), 41.

Woodward, J. (2002a). There is no such thing as a ceteris paribus law. *Erkenntnis*, *57*(3), 303–328.

Woodward, J. (2002b). What is a mechanism? a counterfactual account. *Philosophy of Science*, *69*, S366–S377.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318.

Woodward, J. (2011). Mechanisms revisited. *Synthese*, *183*(3), 409–427.

Woodward, J. (2013). Mechanistic explanation: Its scope and limits. *Aristotelian Society Supplementary Volume*, *87*(1), 39–65.

Woodward, J. (2014). Explanation in neurobiology: An interventionist perspective.

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, *91*(2), 303–347.

Woodward, J. (2017). Explanations in neuroscience: An interventionist perspective. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science*. Oxford University Press.

Woodward, J. (2018). Explanatory autonomy: The role of proportionality, stability, and conditional irrelevance. *Synthese*.

Yablo, S. (1992). Mental causation. *Philosophical Review*, *101*(2), 245–280.

Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science*, *78*(2), 238–263.

Zednik, C. (2014). Are systems neuroscience explanations mechanistic?

Zednik, C. (2018). Mechanisms in cognitive science. In S. Glennan & P. Illari (Eds.), *The routledge handbook of mechanisms and mechanical philosophy* (pp. 389–400). Routledge.