Mind the Matter: Active Matter, Basal Cognition, and the Making of Bio-Inspired Artificial Intelligence

David Jacob Harrison

Selwyn College

University of Cambridge

July 2023

This thesis is submitted for the degree of Doctor of Philosophy

Table of Contents

Declarations

• This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

• It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

• It does not exceed the prescribed word limit for the History and Philosophy of Science Degree Committee

Acknowledgments

Doing philosophy for a living can be draining work and I entered the Ph.D. with something like a ‘crisis of faith’. That is, until I encountered a small book by Gilles Deleuze and Felix Guattari entitled *What is Philosophy?* [*Qu'est-ce que la philosophie?*]. Of enduring influence on me was their notion of the *philosophical persona* that accompanies a philosopher and their work as a penumbra of thought: something that mediates, shapes, and moulds the way one thinks and who one becomes. This, importantly, stresses the need to cultivate one’s persona, one’s personality—which at the time was necessary for me to hear because I believe when we are first taught philosophy, we are told to remove ourselves from the picture we seek to paint.

We are taught that philosophy is an epistemic exercise – striving towards an ideal, an end point, at which the subject themselves recedes into the background to let the veracity of what is spoken shine through—unmediated by the subjectivity or idiosyncrasies of the author. In some sense, philosophy encourages this kind of self-effacement. One thing I learned from this small work of Deleuze & Guattari is that, rather than undermining one’s philosophy, developing a strong sense of self, becoming who you are, is the real fount for a philosophy worth having. As important as pioneering a new philosophy, a revolutionary theory, an inspiring ideal, is the personal adventure of becoming what one is. Wrapped up in this (and I promise I have a point here about acknowledgments) is the *interpersonal* process of learning who one is due to the love, support, and encouragement of those with whom we are surrounded. This thesis, then, is dedicated to all those who helped me through this time and encouraged me to develop myself emotionally, intellectually, and personally; who have contributed to my ‘philosophical persona’, ways of thinking, and ways of being; and who taught me that emotionality and intellection are not different beasts played off against each other (to their detriment), but constitute a whole from which true philosophy—yes, including philosophy of science—can begin.

I would like to thank my family for their unconditional (a word one should never use lightly) support and the many times I could call them across time zones and oceans to receive support—in all kinds of ways and on more occasions than I care to recall. Especially, I would like to thank my sister, Katherine Harrison, whom when I visited during her studies at Harvard (the *other* Cambridge) drove me to pursue philosophy of science in the way I did. Her studies in Neuroscience and Artificial Intelligence undoubtedly inspired me to be where I am today.

My philosophical colleagues, friends, and mentors—too many to name here—helped lay down my path for walking. I would especially like to thank George Deane, Urte Laukaityte, Wiktor Rorot, Andrea Gambarotto, Auguste Nahas, Alejandro Fábregas Tejeda, and Matt Sims. Of the many professors and mentors I have had, I would like to thank Tim Lewens, Mark Sprevak, Michela Massimi, Alexandre Bird, Jessica Berry, Colin Klein, and Andy Barron.

This work could not have been completed without the love and support I had from my friends. To Martin Chivers, Nate Roundy, Megan Murdie, Lilian Kroth, Gediminas Lesutis, Fergus Hamilton, Andrew Haehn, Marcus Collins, Jesse Sperling, Stefan Liberadzki, Bennett Golder, Jason Ball, Joe Post, and many, many more: thank you for all your support and I look forward to making more memories in the future.

Lastly, and I dare say most importantly, this work is dedicated to perhaps one of my greatest supporters: my supervisor, Marta Halina. I could not have asked for a better and more understanding person to accompany me on this journey for the past four years. She consistently pushed me to not give up, helped me through the times when things were going rough, challenged my thinking in encouraging and helpful ways, and has been a great mentor. Perhaps more than anyone else, this work would not have been possible without her support.

Abstract

This thesis explores a recent trend in theoretical biology and cognitive science called the biogenic approach to life and mind: the view that living processes of self-organization and self-maintenance are intimately linked with agential capacities underpinning survival and reproduction. It addresses a central question within this body of literature: what is an agent? What are the conditions that must be in place for agency to emerge in nature? What systems can be called agents in a realist sense; and what systems are metaphorically ascribed agency? Because the systems the biogenic approach examines (such as protists, unicellulars, and slime moulds, to name a few) do not exhibit sharp distinctions between their materiality and their cognitive capacities, these questions boil down to how mind relates to its material substrate more generally—connecting to debates in the life and mind sciences concerning the epistemic and ontological status of ascribing agency to natural phenomena. That is, it addresses the naturalisation of agency: how do we reconcile the purposive nature of organismic processes with our naturalist commitments to scientific methodology? Historically, this amounted to showing how the supposedly agential nature of organisms can be reduced to mechanistic processes. We treat organisms as if they are agents, but ultimately this locution can be elided when ontologising. In this thesis, I pull from recent developments in Active Matter Physics, Basal Cognition, and Bio-Inspired A.I. to argue that this conception of materiality and physicality needs to be reworked to reflect the state-of-the-art. When we pivot away from a conception of materiality as entailing reductive physicalism, we find that the above adumbrated disciplines are in fact conceptualising and operationalising notions of agency (and closely related notions of emergence and downward causation) that are both realist with regards to agency and non-reductive when exploring materiality. To conclude, I argue that the biogenic approach can better overcome concerns of reductionism and as if teleology when equipped with this updated view of materiality.

Introduction
Mind the Matter: The biogenic approach to life and mind

> What is life? What is it that enables living things, apparently so moist, fragile, and evanescent, to persist while towering mountains dissolve into dust, and the very continents and oceans dance into oblivion and back?
> - Robert Rosen, *Life Itself* (1991: 6)

The goal of this work is to articulate and elucidate some of the philosophical foundations and implications of a novel approach in theoretical biology and the life and mind sciences: The biogenic approach to mind and the framework of Basal Cognition (hereafter, simply BC). Briefly, BC can be characterized as examining a set of unconventional model organisms—typically those lacking sophisticated nervous systems or the articulated bodies that exemplify mammalian organization (more on this below)—and highlighting a deep homology between paradigmatically living processes (e.g., self-organization, metabolism, existential needs) and cognitive ones (e.g., goal-directedness, memory, learning). In other words, that understanding the origination of cognition (and potentially identifying it in diverse systems beyond evolution, as seen in the field of robotics and Artificial Intelligence (AI)) requires a re-examination of its material and physiological substrate, such that the "cognitive sciences should at last join the rest of the life sciences in the way they approach their quarry" (Lyon et al. 2021: 1). To this end, the fulcrum of my discussion turns around notions of *intrinsic purposiveness*, *agency*, and *goal-directedness*: concepts that scaffold and animate much of this discourse that speaks of 'biology's next great horizon', consisting of understanding how "cells, tissues, and organisms [are] agents with agendas" (Levin & Dennett 2020).

In other words, I explore the extent to which BC, in its examination of the life-mind relation, comes down on the notion of *naturalized agency* and how the field is actively furnishing a robust account of agency that goes beyond the *as if* teleology that currently dominates the life and mind sciences—as is exemplified in the liberal sprinkling of scare quotes we find in discussions of 'agency', 'goal-directedness', and 'teleology'. In contrast to this, I view BC as providing an account of naturalized agency that accommodates the agential modalities expressed in the organic and inorganic world. As such, BC, and its philosophical articulation under the biogenic approach, seeks to *reframe cognition* such that agential features governing physiological regulation in accordance with existential needs become front and centre on this new account. This is an approach that depicts cognitive activity as immanent and inherent in diverse constellations of materiality and living activity—such that physiology and cognition are braided together in a way that cannot easily be dissociated.

As we will see below, one of the biggest stumbling blocks facing previous attempts at naturalizing agency is an outmoded conception of matter and materiality assumed at the outset of the debate—one of matter as inert, passive, unorganized, and wholly lacking the directionality of relevance to accounts of agency and teleology. This is a view of matter often associated with a 'Cartesian-Newtonian' mechanism, reductionist physicalism, and philosophical materialism.[1] However, it is a view that has grown increasingly suspect and questionable in light of discoveries in Basal Cognition (Lyon 2006, 2015, 2020; Van Duijn et al. 2006; Newman 2016, 2019, 2020; Levin 2019, 2020; Bechtel & Bich 2021; Lyon et al. 2021a, 2021b)—and, as we will see, is being actively reshaped and reformulated in the domains of Active Matter Physics and Materials Sciences (Hanczyc & Ikegami 2010; Needlman & Dogic 2017; McGivern 2020; Egbert 2022) and Soft Robotics (Man & Damasio 2019; Blackiston et al. 2021; Bongard & Levin 2021; Kaspar et al. 2021; Kriegman et al. 2021). Thus, my contribution to the literature consists of the following: I propose that reframing the matter side of the mind-matter interaction is as, if not more, essential to discussions on naturalized agency (and the account the biogenic approach aims to provide) as reframing our understanding or approach to mindedness and its distribution. Stated less antagonistically, I am suggesting that reframing cognition requires a reframing of matter and materiality. Additionally, rather than this being a (strictly) philosophical exercise, I suggest that it is in fact philosophy of the life and mind sciences that are increasingly dissociated from the (non-reductive) understanding of matter and materiality explored in the domains of BC, active matter physics, and soft robotics (to name only a few). These disciplines are increasingly reshaping the landscape in which discussions of the material and the physical take place and are of huge importance to philosophers of the life and mind sciences: such that the quick and easy dissociations of living activity from cognitive activity that occurs in debates in these areas grows increasingly untenable.

What I am currently arguing is that the materialism shadowing much of the philosophical debates on naturalized agency—a materialism that denies top-down causation, prohibits (strong) emergence, and eliminates agency—is disjointed from a novel approach to material

[1] I agree with recent accounts on so-called new materialism that criticize the sloganistic and propagandistic usage of a Cartesian-Newtonian mechanism that is inconsistent with the nuances of both Newton and Descartes formulations of mechanism.

and biological agency of current focus in empirical approaches in the above adumbrated disciplines. In effect, what I am suggesting here is a position expressed in John Dewey's *Experience and Nature* when he writes "the notion of matter actually found in the practice of science has nothing in common with the matter of materialists—and almost everyone is still a materialist as to matter, to which he merely adds a second rigid structure which he calls mind" (Dewey 1958: 74). We do not stand in a different position in contemporary philosophy of cognitive science, except this materialism has transformed. Writing more recently, Peter Godfrey-Smith (2016a: 496) recapitulates a similar sentiment: "philosophers of [cognitive science] often operate with a picture in which living activity is a kind of non-mental substrate, and then evolution lays a computer—the nervous system—on top of the merely living, after which cognition and subjective experience result". Thus, summarizing the current attitudes of the field, Michael Levin remarks, in a passage worth quoting in full:

> [Cognitive] capacities are taken to be the properties of a fixed, embodied Agent [sic]: the fact that it is a collection of cells or subcellular fragments, which proliferate and actively interact to build its body, is relegated to developmental biologists. That phase of a subject's life is usually ignored as behind-the-scenes setup, after which real study can begin. … But this paradigm is importantly incomplete: it is now essential to begin to unravel the plasticity of both, bodies and minds, within continuous life histories that highlight the fact that no agent is a monadic (indivisible) mind—all are made of parts, and those parts can rearrange. (Levin 2020: 2).

What I hope to outline with this is a problem, a goal, and, as will become clearer later, a strategy: the problem is the broad prohibition on agency and teleology that characterizes contemporary biology and (certain) approaches to mindedness in the philosophy of cognitive science. The goal is to articulate the manner in which agentic and teleological concepts can play an explanatory and guiding role in the life and mind sciences—in a way that is compatible with our naturalistic commitments. The strategy will be to take a different tack than that which is commonly pursued: working through the matter side of the debate and rejecting a conception of materiality that blocks, denies, or precludes an emergentism that is essential for a non-eliminative, yet naturalized, account of agency.

This approach is rendered more tractable when one picks up Levin's suggestion above: beginning from the constituent self-organizing processes that make up the multicellular forms of a cognitive agent. In other words, BC, and the biogenic approach more generally, explores how cellularity expresses a distinct kind of agency and minimal cognition from which higher-level cognitive sophistication emerges (Lyon 2006; van Duijn et al. 2007;

Godfrey-Smith 2016b; Levin 2019; Lyon 2020). Thus, in examining the close imbrications of plastic bodies and cognitive plasticity, I explore how we can meaningfully speak of the parts and subunits of organisms—cells, tissues, networks—and the materiality that comprises them as cognitive and agential. What I want to suggest, then, is not (or not simply) that the body and its materiality matters to cognition (a view that tacitly supposes the body in service to a more or less unified, more or less higher-order cognitive subject), but rather that the body *itself*—at varying levels, and to various degrees, of organization—exhibits proto-cognitive capacities through and through: from cellular activities entrained to regulating morphology, development, and intercellular communications; to tissue complexes and system functioning; through to the more baroque appearance of cognition manifested in cephalopod, arthropod, avian, and mammalian brains—Darwin's 'endless forms most beautiful' (Darwin, *Origin of Species*). In a phrase expressed elsewhere, this is *cognition all the way down* (Levin 2019; Levin & Dennett 2020) or *mind everywhere* (Levin 2022). As an exemplary expression of the BC framework and the biogenic approach, making sense of this idea is a driving goal of the present work.

To recapitulate briefly, I have suggested that much of the contemporary impasses relating to agency and teleology in the life sciences stem from an outmoded conception of materiality, matter, and embodiment, such that reframing the matter side of the matter-mind interaction becomes of paramount importance. This is because we often work with a view of materiality devoid of potential for emergence and top-down causation: two concepts of critical importance for naturalizing agency. In a world that is nothing but "the foresightless play of fermions and bosons producing the … *illusion* of purpose" (Rosenberg 2018: 19), it is truly mysterious how the agential, goal-directed, indeed *purposive* activities of organisms is possible in anything other than an illusionist or soon-to-be-eliminated sense (as Alexander Rosenberg suggests). This vision of the foresightless play of fermions and bosons is evidently meant to imply that the real 'causes that kick' (as Thomas Jefferson once put it) are 'at bottom' 'nothing but' (or 'just are') the activity of fundamental particles. This 'nothing buttery', as Paul Davies (2014) memorably phrases it, as well as its relegating of agency and teleology to irreality, has a long and distinguished history in the life and mind sciences (see Richards 1993, 2002; Zammito 2003; Lewens 2003; Riskin 2016; Gambarotto 2017; Zammitto 2018; Wolf 2021).

For instance, as will be explored in detail later (Chapter 2: 'Origins of Organization'), I will document how the Kantian formulation of self-organization and intrinsic purposiveness, as well as the banishing of teleology from the domain of 'proper science', casts a long shadow over contemporary debates in the life and mind sciences vis-à-vis teleology and agency. For now, it is simply worth noting that a pedigreed tradition of philosophical and scientific thinking has sought to reinforce the broad prohibition on agency in biology that is being reconsidered today. Contrastively, a large part of the philosophical heavy-lifting in my argument will be to work against this view of materiality, which is not only philosophically suspect, but increasingly at odds with what the physical and chemical sciences are revealing about the world. In doing so, we might not necessarily *close* a gap between agency and materiality (indeed, many pressing questions will still linger); but I do wish to suggest that the gap becomes significantly narrowed as a result.

I have kept the above discussion relatively schematic in order to limn the overall argumentative structure and direction of this work. With the broad-stroke position and approach on the table, I now turn to an overview of the thesis in outline:

Chapter 1, 'A Vocal Minority: Exploring the Life-Mind Continuity Thesis', begins with an introductory discussion on the nature of the biogenic approach to life and mind, and elaborates features of the basal cognition framework. I would like to draw attention to two central theoretical contributions of basal cognition that, I believe, reshape our philosophical engagement with the material and the mental: namely, the physiological and biophysical nature of information processing; and the nested nature of collective intelligences. If work on BC is on the right track, then it suggests that the emergence of multicellular forms and cognitive agents should be understood as a process of *collective intelligence*— "that [it] is more effective to understand intelligence as being collective 'all the way down'" (Faraday et al. 2023). The reason for drawing attention to these two facets of BC research is that they highlight the importance of reconsidering how the cognitive relates to its physiological substrate, in a way that helpfully opens a path for incorporating insights of active matter physics and materials sciences into a wider theoretical armature for understanding agency and mind. One upshot of this chapter is that debates in Basal Cognition—specifically, the ascription of cognitive functions or mindedness to comparatively simpler organisms—would benefit from the identification of an agential dimension that is antecedent to, and fundamental for, cognition: and it is on naturalised forms of agency that most of the thesis will focus on.

Chapter 2, 'Origins of Organization: Intrinsic Purposiveness, the Kantian Legacy, and the Post-Kantian Response', represents a historical and philosophical examination into the current state of affairs surrounding teleophobia and the omnipresent *as if* teleology that still defines modern biological praxis. As the title suggests—and consistent with the wider literature on the topic—I situate some of the current anxiety at the origin of the life sciences and the rich philosophical engagement with it, specifically relating to the works of Kant and Schelling. Specific attention is given to the notion of intrinsic purposiveness and the type of reciprocal causality and revamped materiality required for its 'naturalisation'. Ultimately what I want to canvas in this work is two views of 'nature', the Kantian and Post-Kantian respectively, and how depending on the materialist and naturalist commitments one adopts, the purportedly 'natural' credentials of agency and teleology look correspondingly different.

Chapter 3, ''Just Physics'? Rethinking Emergence and Causation in Physics and Developmental Biology' telegraphs the considerations of Chapter 2 into a modern context and introduces some novel insights from Condensed Matter Physics and Materials Sciences to suggest that the sciences of matter do not support the reductive views associated with philosophical physicalism and materialism. By revisiting these sciences and their active exploration of notions of emergence and downward causation, I suggest that we expand the 'physical' within physicalism to include a wider range of disciplines: such that the supposedly inscrutable nature of emergence becomes more demystified in the process. This will support the overall thesis that what is required for naturalised forms of agency is a rethinking of materiality and physicality, as assumptions thereof directly influence our theorising about agents and minds.

Chapter 4, 'The Extraordinary Liveliness of the World: Active Matter, Agency, and a Relational Ontology', further explores the rethinking of materiality promoted in this work. If in Chapter 3 I focus on properties of condensed and emergent systems, Chapter 4 pivots to the thermodynamic and bioenergetic context thereof which serves as its compliment. Introducing Active Matter Physics into the debate on naturalised agency aids the discussion as it provides principled boundaries between the kinds of systems that can sustain agential activities and those that cannot. This draws together the lines of argument from Chapters 1, 3, and 4 to argue that one path to dispelling anxiety surrounding agential reasoning in the life and mind sciences is to show how the concepts required for its naturalisation—emergence and some form of downward causation—are not only amenable to contemporary physical sciences approaches, but actively part of emerging research programs.

Chapter 5, 'A Road Map for Naturalising Agency', then draws the threads together and explores in nested detail a question at the heart of this thesis: what is an agent and what needs to be done conceptually and shown empirically in order for agents to be considered part of scientific and philosophical ontologising. Specifically, I will build off a roadmap put forth by Potter & Mitchell (2021) to suggest a naturalistically defensible way in which agency can be explored. This comes down on eight criteria: (1) thermodynamic autonomy; (2) Persistence; (3) endogenous activity; (4) holism; (5) lower-scale indeterminacy; (6) some aspect of multiple realizability; (7) historicity; (8) normativity. While my account here builds off Potter & Mitchell, it diverges in important places and pulls from different literature to make the point where necessary.

Lastly, in Chapter 6, I pull from my published material (Harrison et al. 2022) to explore the extension of these ideas and the biogenic approach into the domains of soft robotics and what is called bio-inspired Artificial Intelligence [AI]. My interest in making this move is twofold. First, soft robotics and matter-level innovations represent novel and exciting research avenues in which the relevance of materiality can find real experimental traction. This makes the field of soft robotics a promising 'sandbox' where the ideas of this thesis can be, and have been, put to the test. The second reason is more theoretical in nature and pertains to a distinctive modality of living systems explored and defended in this work: that organisms have existential needs and, to put it in a somewhat colloquial way, *care* about their existence. The philosopher of cognitive science John Haugeland once quipped that "The problem with A.I. is that it doesn't give a damn" (1979). We know—because we indeed ourselves *are*—that organisms paradigmatically *do* 'give a damn', so one of my interests is the extent to which being a precariously, vulnerably embodied being (here again is a materiality point) in some sense bootstraps the affective and valence driven aspects of organismic existence: and if introducing some analogue dimension in AI or robotics could represent an advancement in the direction of artificial systems that might one day exhibit the intrinsically purposive and motivational aspects of living beings.

Chapter 1
A Vocal Minority: Exploring the Life-Mind Continuity Thesis

1.1 Introduction

Research in the cognitive and neurosciences has traditionally focused on high-level features and capacities that are restricted to a slender section of the evolutionary tree: humans and closely related primates. The focus is usually on, e.g., transitive reasoning, game-playing (Chess, Atari, Go), and ratiocination as the hallmarks of intelligence. Accordingly, it is not surprising that attention is largely placed on the human brain—often depicted as the coronation of neural evolution—and its constitutive components: neurons and neural networks. This is a view encapsulated in, and articulated under, the neuron doctrine (Yuste 2015). While originally postulating that discrete cells serve as the basic features and components of the nervous system (instead of reticula as Camillo Golgi suggested), the neuron doctrine extended to encompass the domain of mindedness more generally. It is therefore suggestive of the fact that the best inroad into understanding minds—its emergence and evolution—is via the neurosciences, given our methodological commitment to naturalism.[2] In other words, if the neurosciences (and their basic commitment to the neuron doctrine) reveal that the mind is mediated, enabled, and manifested by neurons and their concatenations of spike trains, then this is the most appropriate avenue through which psychological, cognitive, and mental phenomena should be explored (Gold & Stoljar 1999: 809).

In the introduction to this work, I suggested that the unifocal approach to nervous systems had the corollary of marginalising non-neural modes of information processing—which we now know to be ubiquitous both in animals with sophisticated nervous systems and (multicellular) organisms that lack neurons entirely. In this chapter, then, I will overview some of the core features and claims associated with the biogenic approach to life and mind. I will highlight three positions in particular: first, in Section 1.1 ('A Vocal Minority—Situating Mind in Life') I overview the privileged position of the neuron doctrine and how recent work

[2] Understood as the belief that what there is in the world is that revealed by scientific practice. However, as a driving component of this thesis, what shape this naturalism can take is highly contested, and the naturalistic commitments I make in what follows would differ significantly from a more conventional naturalist picture. See section (1.4) in particular for this discussion.

in the neurosciences reveal it to be untenable. Although the biogenic approach is not restricted to (or even focused on) neuroscience, I nevertheless show how some of the most compelling arguments in favour of the biogenic approach are in fact native to the context of Human and affective neuroscience. If the contention that neurons are the primary or sole basis of cognition and mind is untenable even in the neurosciences, then this will help dislodge some of the prejudices against taking non-neural modes of cognition seriously. Thus, section 1.2 proceeds through multiple steps to arrive at the conclusion that cognition cannot be easily extricated from its living, metabolic, developmental—indeed, *material* context.

Section 1.3 ('A Society of Selves: Cognition All the Way Down') proceeds to another central claim of the biogenic approach: namely, it identifies the nested nature of cognitive structures that follows from the conclusion in sections (1.2) and I overview recent literature which has elucidated the hierarchical and heterarchical (Bechtel & Bich 2021) nature of information processing and how we might consider even tightly integrated organisms (ourselves included) as consisting of cognitive structures 'all the way down': from the level of neural networks, through to organ structures, individual cells, and indeed symbiotic and multi-organismal interactions that we know to play a constitutive functional role in cognition. Having defended these two positions in sections 1.2 and 1.3, I proceed in Section 1.4 ('Working Against a Reductive View of Matter and Mind') to philosophical and metatheoretical commitments to 'naturalism' and 'materialism' to highlight how much is outdated or outmoded in contemporary philosophical approaches to these terms: specifically where being a 'naturalist' or 'materialist' is taken to prohibit ascriptions of agency to nature; block top-down causation; and cast doubt on (strong) emergence—concepts that are essential to any attempt at naturalizing agency and 'intrinsic purposiveness'. In contrast to a naturalism that rejects these features, I propose a position that takes seriously the causal role and function of emergence in the life and mind sciences such that advancing the discourse on naturalizing teleology and agency becomes more feasible as an outcome. I conclude with remarks on the rest of this work in Section 1.5 ('The Biogenic Approach: Rethinking Matter to Rethink Mind').

1.2 A Vocal Minority: Situating Mind in Life

The past decade has witnessed a flurry of research across the life and mind sciences that explores the strong continuity between paradigmatically living processes and cognitive ones, such that "the principles of biological organization and the requirements of survival and reproduction present the most productive route to a general understanding of the principles of cognition," indeed, that "cognition, whatever else it may be in the future, is naturally a biological process and biological function" (Lyon 2006: 12). This, in effect, is the biogenic approach to life and mind. While my focus in this work will be to shift the goal posts away from *cognition* to one of a more general understanding of *agency*, the literature often covers both interchangeably and it is this literature I now review. That said, in this section I explore how the biogenic approach receives support from various domains in the life and mind sciences—from the 'top-down' (1.1.2), as it were, in affective neuroscience, and the 'bottom-up' (1.1.3) when looking at Basal Cognition [BC]. My goal here is to show how many streams flow into the biogenic approach—encouraging the wider reframing of cognition that I wish to advance in this work.

1.2.1 The Physiological Basis of Neural Cognition: Recent Insights from the Mind and Brain sciences

"In a way", neuroscientist Rafael Yuste writes, "The history of neuroscience is the history of its methods" (2015: 487). Under consideration here is the neuron doctrine: the long-running thesis that the basic processes underlying cognition are individual neurons computing in the form of action potentials—and the spike trains that emerge in networks as the concatenation of these individual 'computations'. The neuron doctrine spans much of the history of the cognitive sciences (Piccinini 2010) and posits that the 'stuff' of cognition consists of neurons, action potentials, and spike-trains—with somatic interactions marginalised or construed as either 'background conditions' or merely a 'hardware problem' with little import into how we understand, investigate, or theorise in the cognitive sciences. Godfrey-Smith describes this state of affairs well in a recent paper (2016a). He notes that the rapid advancements of Artificial Intelligence [AI] in the latter half of the 20th century allegedly demonstrated that cognition was mechanizable in non-living systems: "There seems no question of life being present in a classical AI system… and given that there seems a real possibility that such a system might realize all of mentality, there apparently cannot be too close a link between life and mind" (2016: 482). In construing neurons, the electrical activity present in spike trains, and the brain (isolated from the body) as the medium through which mindedness is realised, "computation, rather than life, became the bridging concept between mental and physical" (ibid.).

The strongly entrenched nature of the neuron doctrine therefore side-lined subcellular, developmental, or living processes in favour of digital modes of computation that were more amenable to the ambitions of the computer and cognitive sciences at the time: making cognition implementable in a diverse range of media—indeed, its 'generality' was the main feature speaking in favour of this version of multiple realizability (see Polger & Shapiro 2016)—from more common examples of digital computers and artificial-neural-networks, through to more fantastical scenarios of delicately arranged beer bottles falling at the right angle and right speed. The upshot of this was that the materiality of cognition—its internal dynamics and self-organizing processes—were of little relevance to understanding the nature of mindedness, such that "we could be made of Swiss cheese and it wouldn't matter" (Putnam 1972). Evidently, if materiality matters at all here it is only as a set of background conditions that, once established, can support cognitive function. It is precisely this picture that BC aims to destabilize: in emphasising non-neural modes of cognition and self-organizing dynamics underpinning a range of living and cognitive processes, researchers in BC highlight the importance of 'being a beast machine' (as Descartes memorably put it) for being a cognitive agent (Man & Damasio 2016).

More perspicuously, if the history of neuroscience can be detailed in terms of the history of its methodologies, as Yuste suggests, then the history of the philosophy of mind can be elaborated in similar terms in combination with the most available or dominant technologies of the time. And perhaps no metaphor or technological paradigm has exerted as much of a pull over philosopher's imaginations than computers and their distinctive mode of information-processing: digital computations. Indeed, the view that cognition is a substrate-independent process has a distinguished history in 20th century cognitive science and philosophy of mind. Although the neuron doctrine is, as the name suggests, a partially substrate-dependent thesis, when it is abstracted from its neuroscientific context and generalized to a wider theory of mind, it serves as a broader position regarding the nature of mind-matter interaction. Because action potentials exhibit an all-or-nothing way of firing, they are often presented as analogous to digital computations found in computer science. This is on clear display in a famous thought experiment that David Chalmers presents, called 'Fading Qualia' (1995). Although departing from a much different angle of conscious experience, there are several themes in 'Fading Qualia' that invite us to reconsider the relations between life and mind—or, as Chalmers makes clear, the lack of relationality between them.

Chalmers invites us to envision a robotic isomorph to the human brain—specifically, with neurons supplanted with silicon analogues that do the exact thing that neurons do (or so it is stipulated: "The causal patterns in the Robot's processing are the same as mine" (1995)). We are then asked to imagine a series of intermediary systems between the robot and the human where only incremental steps are made throughout such that the 'functional organization' of the brain is maintained. This could take the form of neurons being replaced progressively with silicon chips until biochemical mechanisms have been dispensed with entirely ("except at the periphery", we are told—though the comment scarcely makes any sense in the context of the thought experiment: why would a silicon functional analogue not exist here as well?). After this intricate swapping of neurons for silicon chips is performed, "and there are no biological mechanisms playing an essential role", we are asked whether there would be any experiential disturbances (at best) or cessations of conscious experience (at worst) in the process.

Because Chalmers has stipulated that the silicon chips behave in functionally isomorphic ways, it is implausible, we are told, that the intermediate stages through to the final mechanized (silicon-based) brain at some point loses consciousness (hence 'fading qualia'). If the silicon analogue recapitulates the functional organization of neural organization, then we should concede conscious presence and qualia to the Robot isomorph. While there are many things to discuss here, the point I want to draw our attention to is that functional organization is maintained such that whether the system is composed of biochemical mechanisms or silicon chips is ancillary to the discussion of one's mental life. This is meant to show, among other things, that the mind is a kind of 'software' that is comprehensible and explicable in terms independent of its biological basis.

While I will have more to say about this question of artificial or digital minds in a later chapter (and the all-important question of multiple-realizability; see Chapter 6), for now I want to suggest that Chalmers' thought experiment holds interest for the present discussion in four ways. First, it highlights the hybrid dependence of philosophical thought experiments in the cognitive sciences on both a predominant thesis in the neurosciences (maintaining its naturalist credos) and its reliance on the dominant machine metaphor of the time: that of the digital computer. Second, and as a corollary, it marginalises the non-neural and somatic dimensions involved in information processing—whether of relevance to consciousness or not. Third, the extrication of biochemical mechanisms and biomechanical processes from the question of mentation has the implication that there cannot be too strong a link between life and mind here. Fourth, it demonstrates how the question of matter is seen as wholly independent, ancillary, and orthogonal to the question of whether a system can or cannot support a certain cognitive software (due to considerations of multiple realisability).

However, it is precisely this view that is currently being prised apart from the dual angles of contemporary neuroscience and BC (as we will see in the following section). As Chalmers acknowledges (but only acknowledges), it is entirely an empirical question as to whether analogue *units*, let alone analogue *systems*, could be constituted so as to maintain the functional organization required for sustaining our mental economy—whether full-blown conscious experience or humbler cognitive processes geared towards life regulation. Rather than speculating or stipulating *a priori* that such a system could exist and then drawing conclusions about the importance of the materiality and matter to being the kind of system it is—and having the associated interiority and intrinsic purposiveness (discussed further below)—philosophers should be more attentive to other scientific disciplines that are progressively elucidating the feasibility of their meditations and the posited extricability between the material embodiment of a cognitive agent and its associated mental life. Stated differently, and to pre-empt the coming discussion, philosophers should be more attentive to the range of disciplines that are making headway on this most pedigreed of problems, and therefore should have more of an eye towards materials sciences, active matter physics, and, as we will see shortly, Basal Cognition.

Pulling from Godfrey-Smith (2016a) again, one of the main points I want to highlight here is that the silicon analogues will, in fact, *not* do the same thing. As Godfrey-Smith notes, the claim "is not that non-biological materials that *do* all the same things might not count because of their physical nature. Rather, the usual candidates offered as a non-biological basis for mentality will *not* do the same things. They will be functionally different, not merely different in 'hardware' or 'make-up'" (2016a: 501). Building on this, one of the most striking omissions in the thought experiment is abstracting out subcellular, intracellular, and non-electrical intercellular processes (such as the use of chemical messengers and other neuropeptides that transmit important information outside of the 'digital' mode of action potentials). This omission, to an extent, makes sense—Chalmers presented the thought experiment during the height of the 90s when the neuron doctrine was still the dominant position in the mind and brain sciences (the paper was published in 1995). But it also exemplifies a more general suspicion towards the relevance of life (and here we might add 'matter' more generally) to mind. Thus, although Chalmers' thought experiment concerns the nature of conscious experience, the philosophical moves on display here are emblematic of what Karen Barad calls a 'deep mistrust of matter' (2007: 101); that is, they typify a specific attitude regarding the life-mind or matter-mind continuities and interconnections that have arguably been the dominant mode of thought only until very recently. Introducing the thought experiment thus allows me to frame an entrenched attitude regarding teleology and agency that deserves to be revisited. That is, while Chalmers thought experiment addresses qualia and consciousness in particular, I argue that it conveys a broader 'mistrust of matter' that pervades debates on teleology—insofar as the latter is concerned with the place of normativity, intentionality, and goal-directedness in the literature. [3] Indeed, what I am currently arguing is that the fine-grained functional details of the cellular and developmental context – details quickly eliminated from the structuring of the thought experiment – matter a great deal more than is commonly supposed.

In contrast, then, to a view which depicts developmental, morphogenetic, or material processes as a 'hardware' problem sitting below a cognitive 'software'[4], many converging lines of evidence across the mind and brain sciences suggest that the interrelations between physiology and cognition are so intimate that demarcating a higher-level software with autonomy from its hardware becomes an unstable strategy. For instance, neuroscientists have progressively elucidated the importance of the gut-brain axis and the role of the endocrine system in mediating and enabling cognitive performance (Richter & Smith 2006; Mayer 2011; Cryan et al. 2012; Mayer et al. 2014; Carabotti et al. 2015; Heck et al. 2017; Mossad et al. 2021; Borsch 2021). Additionally, the discovery of the importance of non-neural modes of communication in the brain—with glial cells and astrocytes being crucial mediators and performers—suggest that chemical communication is as essential to cognition as the

[3] It is interesting to note that many of the problems that lead to naturalist and materialist difficulties vis-à-vis consciousness or qualia—its normative and value-laden nature, its distinctive intentionality, and so on—are replicated in a variety of debates regarding the place of normativity, goal-directedness, or otherwise teleological processes in nature (and here Deacon's criticisms are particularly appropriate; see Deacon 2014). With that being the case, it is not a stretch to see something of relevance between my question concerning teleology and the general anxiety involving phenomenology and the role of consciousness and experience in the cognitive and neurosciences. Indeed, as we will explore below, the 'naturalization' strategies that Shaun Gallagher (among others) explores for phenomenology will be utilized for my arguments on naturalizing agency. In both cases, what is at stake is the nature and ontological status of the *subjects* of experience, the role that subjectivity plays in natural processes, and how dimensions thereof are compatible with our naturalist commitments and methodological approaches to the world.

[4] It is interesting to note an implicit 'layered cake model' nascent in much of the emergentist literature here on multiple realizability—something that will be broken down further both in Chapter 3 and 5.

electrical activity on which the neuron doctrine depends (Wake et al. 2009; Tremblay et al. 2009; Fields et al. 2014; Liu et al. 2016; Benarroch 2016).[5] Moreover, several non-neural cell types have been identified that exploit similar bioelectric principles to neurons, and indeed bioelectricity is one of the key mechanisms explored in the coming sections, highlighting this ubiquitous information-processing mechanism across all domains of life (Sundelacruz et al. 2009; Sullivan & Levin 2016; Pietak & Levin 2018). A point I want to highlight is that recent work in the mind and brain sciences re-situates neurons within their *cellular* context as participating in, depending on, and being enabled by more common and 'biogeneric' (Newman 2019) features of wide distribution in the organic domain.

This, in effect, situates neurons in an evolutionary continuum along with other non-neural 'minimal model systems' in BC, such as biofilms and slime moulds, in order to better understand the homologous mechanisms that serve various life-regulating and cellular roles—of which the generation of an action potential is just one, if highly significant, process. For example, the discovery that non-neural creatures such as bacterial biofilms generate electrical currents or express intercellular communication mechanisms present in neurons has encouraged reconsidering the biophysical generality of information-processing mechanisms (examples include the potassium-gated ion channel that is crucial for maintaining membrane potentials; as well as the utilization of gap junctions to form reticular structures with shared ion distribution and electrical charge). These have moreover been mapped onto behaviours relevant to survival, flourishing, and reproduction in non-neural organisms—such as learning, goal-directedness, memory formation and retention, proactive and anticipatory behaviours, and solving navigational problems (Baluska & Levin 2015; Beekman & Latty 2015; Prindle et al. 2015; Nunes-Alves 2015; Gershman et al. 2021). Jointly, these empirical and theoretical advancements have encouraged a reframing of cognition in a manner that decentres neurons and elevates the body and the self-organizing processes of cellular activities to a heightened place of cognitive relevance.[6] Moreover, they complicate the quick dissociations one might make from the armchair when reflecting on the material and physical basis of mind.

---

[5] Chalmers, for his part, believes he has pre-empted this, remarking that he leaves aside the question of glial cells being important and noting that, if they are, then they too will be replaced in the thought experiment by the silicon analogue.
[6] This is to say nothing of the extensive and dense psychological literature which discusses perception as sensorimotor and research in embodied and embedded cognition which elaborates on the inextricable relations between bodies and minds. Given that these are extensively explored elsewhere (Noe 2005; Gallagher 2017), I do not explore them further herein.

In light of recent empirical advancements and theoretical articulations of the importance of life processes for mindedness (Friston 2013), we can see how these insights collectively emphasise "the deep physiological roots" of higher cognitive function (Seth & Tsakiris 2018: 979). Seth & Tsakiris, in a recent review, note how work in affective neuroscience—specifically, work on interoceptive inference, or the modelling of one's internal milieu—signals an importance of taking the 'beast machine' not as a medium to be derided, marginalised, or abstracted out, but as the key dimension through which the value-laden and goal-directed orientation of cognitive processes is expressed. As they write, "There are intimate connections between the functional imperatives imposed by our physiological reality, by the drive to stay alive that animates all creatures, and the predictive machinery that implements [cognitive processes]"[7]. This is a view that targets popular and sedimented views of the mind "as substrate-independent forms of information processing" (ibid.). They conclude by highlighting the implications of exploring cognition through its physiological roots (through 'being a beast machine'): "At minimum, it suggests that mind and self cannot be understood without deep appreciation of the constraints and opportunities afforded by embodiment… More radically, it underpins a strong continuity between life, mind, and consciousness… We perceive the world around us, and ourselves within it, because of, and not in spite of, the fact that we are beast machines" (ibid.).

Thus, to conclude this section, a further corollary of taking the 'beast machine' approach seriously is that it puts into focus the far-from-equilibrium and self-organizing processes exhibited at every order of scale in biological systems: from the nervous system, to tissue complexes, and all the way down to individual cells—"cognition all the way down". Thus, similar to Seth & Tsakiris, Ciuanica & Levin (2022) have recently shown how "biological organisms [their focus is on the human organism] with complex bodies need to develop and sustain robust yet flexible self-organizing and self-regulatory mechanisms implemented via multilevel hierarchical organisation: organelles constitute cells which form tissues which in turn form organs, etc." (2022: 2). That is, their work highlights how neurons both build on evolutionarily ancient information-processing mechanisms *and* work in tandem with a range of other cell types and cognitive structures in order to subserve self-organization and adaptive behaviour at local and global levels of the human organism. I thus argue that my approach focusing on more 'basal' elements—and emphasising the material, self-organizing, far-from-

[7] Their focus is on what they call instrumental interoceptive and Bayesian inference—which are key features of cognition generally but core concepts from their theoretical framework of predictive processing.

equilibrium processes thereof—plugs hand in glove with recent work in theoretical neuroscience that situates the human organism within a strong life-mind continuity. Having suggested how there is compelling reasons native to the mind and brain sciences for moving beyond nervous system activity, I turn in the next section to my main empirical testbed for exploring strong life-mind continuity: Basal Cognition.

1.2.2 Reframing Cognition: Insights from Basal Cognition

Basal Cognition [BC] is an area of theoretical biology that seeks to highlight the evolutionarily ancient nature of certain cognitive capacities on which higher-order cognition depends. This includes studies on memory (Pezzulo et al. 2010), the evolution of learning (Jablonka et al. 2017; Ginsburg & Jablonka 2019, 2021), the origins of goal-directedness and valence (Lyon & Kuchling 2021), and decision-making (Bechtel & Bich 2021). Researchers working within a BC framework identify how these capacities and associated mechanisms are present in both neural and non-neural modes of information-processing—exhibiting a patterned distribution across the biological world (the tapestry of organisms on display can be seen in Figure 1 below). Researchers elect a series of 'unconventional modelling organisms' (Bechtel 2014), covering organisms as diverse as *Escherichia coli* (Hazelbauer et al. 2007); bacteria biofilms (Yang et al. 2020); the slime mould *Physarum polycephalum* (something of a mascot for the BC framework; Dinet et al. 2021); through to plants (Cagliano et al. 2016, 2018); and even somatic and cancerous cells of multicellular animals (Nilsonne et al. 2011).

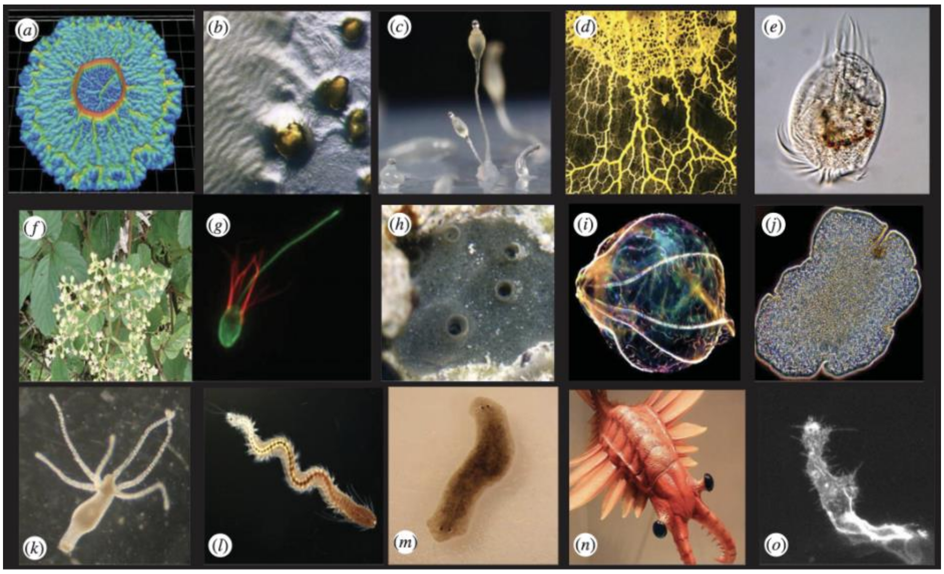

Figure 1. A selection of organisms considered within Basal Cognition for their significance in providing insight into more basic modes of information processing. In order: (a) a stress response signal and structural re-organization in *Bacillus subtilis* biofilms. (b) *Myxcoccus xanthus* and characteristic 'rippling' effects in its multicellularity in response to prey. (c) *Dictyostelium discoideum* is a social amoebae used as a model for studying the evolution of multicellularity. (d) *Physarum polycephalum*, an acellular slime mould, is something of a mascot for BC. It is a giant plasmodium that expands in vast networks and exhibits several interesting mechanistic analogues to neural systems, such as oscillations, vascularity, and electrically charged currents. (e) *Euplotes* and predatory ciliates that exhibit behavioural plasticity due to the evolution of the distinctive mode of eukaryotic electrical excitability. (f) *Cayratia japonica* tendrils, which have been suggested to show elements of 'self-recognition'. (g) *Salpingoeca rosetta*, the closest living unicellular relative to animals, has been shown to contain almost as much of the molecular infrastructure for producing neurochemical and neurosecretory vesicles as (h) the multicellular sponge *Amphimedon queenslandia*. (i) *Beroe sp*. with recently swallowed prey, which is thought to have independently evolved a nervous system. (j) The aneural placozoan *Trichoplax adharens* possesses numerous signalling peptides that are homologous to those used in animal nervous systems. (k) *Hydra vulgaris* (Cnidaria) which is amenable to whole-organism level neural image analysis. (l) *Platynereis dumerilii*, a polychaete worm used for studying the origin of nervous systems. (m) a favourite of the Levin lab, planeria (flat worms) have been explored for their remarkable regenerative abilities including the recruiting of bioelectric signals to coordinate morphological regeneration. The one pictured has been artificially pushed to regenerate two heads. (n) artist's impression of *Larapax unguispinus*, a large marine predator from the Cambrian era. Ginsburg & Jablonka (2019) have proposed the Cambrian explosion, with the vast radiation of complex morphologies, as the origin for complex forms of learning and consciousness. (o) a blood vessel sprout grows in a Zebrafish embryo. These sprouts exhibit collective decisions based on 'active perception' and sensorimotor feedback. Photo borrowed from Lyon et al. (2020).

Interestingly, many of these organisms lack the highly articulated nervous systems, functionally differentiated cells, and complex bodies that are commonly seen as the coronation of cognitive evolution and thus as necessary for sophisticated behaviour. As Michael Levin (2020: 4) writes, in what I take to be a programmatic overview of the framework,

> (Basal) Cognition does not require a nervous system or brain. The realization that many organisms, including aneural ones, exhibit proto-cognitive functions such as memory, integrated decision making, prediction, and ability to learn general rules from instances is very old. The emerging field of Basal Cognition focuses on the phylogenetic origins of learning and goal-directedness, drawing a continuum between the humble origins of

> information processing in the metabolic homeostatic mechanisms of ancient cells and more complex learning, representation, and goal-directed activity.

This approach is meant to purchase two things:

> First, [the insight that] decision-making and scaling of integrated information processing is ancient and ubiquitous across the tree of life … [And] second, as befits the underlying evolutionary origin of all life on earth, the mechanisms used for cognition are highly conserved and predate multicellularity, working in similar ways in the control of morphogenetic behaviour of single cells as in the control of animal behaviour. (ibid.)

This wide-ranging (and, we might add, unorthodox) exploration of diverse creatures is suggestive of the fact that there is a close link between the substrate of information processing and cognitive function—considering that many of these organisms lack a decoupling of metabolic self-organization and nervous system activity (see Moreno & Etxeberria 2005). Thus BC is motivated by the contention "that the principles of biological organization and the requirements of survival and reproduction present the most productive route to a general understanding of the principles of cognition," indeed, that "cognition, whatever else it may be in the future, is naturally a biological process and a biological function" (Lyon 2006: 12).

What this serves to highlight is a critical philosophical theme and lineage that runs like an Ariadne's thread through the disparate approaches to BC in the literature: the life-mind continuity thesis (Jonas 1966; Maturana & Varela 1980; Godfrey-Smith 1994; Clark 2002). Hans Jonas, in his influential book provides an early metaphysical version of the thesis: "A philosophy of life comprises the philosophy of the organism and the philosophy of mind… [Its] statement of scope expresses no less than the contention that the organic even in its lowers forms prefigures mind, and that mind even on its highest reaches remains part of the organic" (Jonas 1966: 128). While Jonas writes from the perspective of ethics and phenomenology (see Nahas Forthcoming), his prescient articulation of a life-mind continuity thesis was taken up in the framework of autopoiesis, which similarly highlighted broad commonalities between the living and cognitive domains (Maturana & Varela 1980). Godfrey-Smith describes this thesis as mind literally being life-*like* (1994: 83).

There are, however, two different strands of the life-mind continuity thesis worth dissociating: weak and strong versions. The strong version maintains that 'mind is literally life-*like*', and that the resources used to investigate the latter are necessary for investigating the former. This is a position held by, *inter alia*, Maturana & Varela (1980) and Evan Thompson (2007). In other words, that "cognition is an enrichment of organizational principles and properties definitive of life" (Sims 2021: xi). The weaker version of the life-mind continuity thesis states that cognition merely requires life. At first blush, it is unclear where the majority of researchers fall on this continuum of weaker and stronger theses, and it is probably unhelpful to pigeonhole the field into any one commitment. Moreover, and as will be a driving goal of this work going forward, many researchers (present author included) are interested in generalising these principles beyond life as we know it. Thus, if life is construed broadly such that nonbiological exemplars can come to embody it, then this would remain consistent with the life-mind continuity thesis. If life is construed more narrowly, then this might dash the ambitions of extending the insights BC provides beyond the domain of biology. If that were the case, then cognitive scientists—especially those in Machine Learning, Robotics, and AI—might rightfully find the biogenic approach of niche and vanishing interest to their area of application. Given my interest in engaging this literature and these disciplines, I propose that we understand the biogenic approach as presenting an account of living activity—to be elaborated further below—and an imbrication with cognition of sufficient breadth so as to include potential alien and artificial cases while still restricting the domain of cognition such that, e.g., a digital computer, tornado, or vacuum cleaner fall outside its remit. We will earmark this for now and return to this problem on so-called *multiple realizability* below.

For now, my goal in relating the biogenic approach to the life-mind continuity thesis is that it highlights and clarifies several of the theoretical claims and ambitions we find in BC literature: namely, the interweaving of development, self-maintenance, and behaviour such that cognition becomes more intimately tied to its material embodiment than quick and easy dissociations of the two might suggest (dissociations that will occupy us below, and which are arguably one of the hallmarks of the cognitive revolution (Haugeland 1997)). For example, Levin has highlighted a 'deep evolutionary conservation' of fundamental neuronal mechanisms—such as membrane potential and ion channel functionalities—that exhibit a patterned distribution across neural and nonneural modes of life, illustrating "a fundamental isomorphism between developmental and behavioural processes" (2019: --) and a 'deep symmetry between behaviour and morphogenesis' (Levin 2022: 2). Such an isomorphism is supposed to encourage us "to think deeply about the bodies which enable and constrain various cognitive functions, and their potential plasticity" (2020: 116). Accordingly, the biogenic approach highlights how the plastic and active nature of materiality and embodiment constrains, enables, and conditions the emergence of cognition in the biological world.

Thus far, I have refrained from venturing a definition of cognition, namely because such a task is endlessly plagued with difficulties in the philosophy of cognitive science. Nevertheless, it is becoming pertinent to furnish at least a provisional account. The account of cognition on which I will rely throughout revolves around the notion of *existential needs* (Froese 2017; Egbert 2022; Harrison et al. Forthcoming) and can be explicated, following Lyon and colleagues (2021a), in relation to a set of sensory and information processing mechanisms organisms have for familiarizing themselves, valuing, and interacting actively with the environment in order to meet the existential needs of survival, persistence, reproduction, and growth.

Before continuing to the second feature of BC of relevance to the current discussion, I would like to pause and reflect on the role and importance of existential needs in the above account of cognition. One of the central concerns for biological approaches to agency and cognition is accounting for and naturalizing *normativity*, such that to be an agent, on the biogenic account, requires a kind of intrinsic mattering to that agent. Without getting into the weeds of naturalized accounts of normativity and valence here, what I want to focus on is how the concept of precariousness plugs in here, that is, "The precariousness that is intrinsic to all organismic, and therefore also of all mental, existence is the original reason why things matter to that individual being" (Froese 2016: 34). Precariousness is a useful way to get a grip on understanding how a range of systems have existential needs, but only those systems embodying precariousness are said to be cognitive. Allow me to elaborate.

Beginning with existential needs, there is a broad and general way in which these can be spelled out such that even rocks can be said to have existential needs. As minimal as it gets, these are simply the conditions that must be met for an entity to be the kind of entity it is. For example, the desk I am currently writing on cannot be heated beyond a certain temperature for it to be a desk. For that matter, the computer on which I write cannot overheat (as its internal fan constantly reminds me) for it to be a suitable platform to communicate these thoughts. However, these are all non-dissipative structures, and their needs are met either passively or due to active maintenance by an external user. Dissipative structures, such as organisms, represent entirely different regimes of order and the manner in which they maintain their existence is fundamentally different. Properly speaking, non-dissipative systems do not maintain their structures at all; whereas organisms *must* act in a denumerable set of ways in order to persist. Matthew Egbert (2021: 3) captures the difference when he writes: "rocks (and other non-dissipative entities) are merely passively stable, whereas dissipative structures are constantly falling apart and yet persist thanks to processes of repair, replacement, or reconstruction". Taken further, we see:

> This means that existence for passively stable entities is the absence of a destructive event. By contrast, for dissipative structures, existing is a process—and a process that must continue for the system to persist. Processes have rates and as such it is possible to measure or quantify or respond to how 'viable' a dissipative structure is—i.e. how well it is doing at persisting … There is no equivalent measurement for passively stable systems as their existence is not a process in the same way that it is for dissipative structures. (ibid.)

This is key, as the difference in their manner of persisting re-enforces the fact that not only is an alternative physical language required for comprehending their dynamics (namely, the language of far-from-equilibrium physics), but that it furnishes an empirically tractable, sometimes even quantifiable, mode of investigating their activities and marking them out as distinct from passively stable entities—as is explored in the active matter literature explored below and at a later chapter (see Fransworth (2018), Egbert (2021), and Potter & Mitchell (2022)). Connecting this back to precariousness, we see that organisms exist in precarious (metastable) states that delimit the kinds of dispositions they can have and conditions they can inhabit in order to remain viable (read: stay alive). The biogenic approach, in its emphasis on existential needs and living activity, thus ascribes cognition to systems that paradigmatically embody this form of existential precariousness as it mandates an intrinsic relation between acting and self-maintenance.[8] In other words, it is the *organizational dynamics* maintained in far-from-equilibrium conditions that endow organisms with their characteristically teleological and agential behaviour. As Mossio & Bich put, in a remark that is central to the present work, "the organisation of biological systems is inherently teleological, which means that its own activity is, in a fundamental sense, first and foremost oriented towards an end" (2017: 1090).[9]

[8] I believe the focus on precariousness here allows us to address self-organization and its causal dynamics in a way that allows us to purchase the importance of materiality for manifesting that which is distinctive of cognitive embodiment while simultaneously evading concerns of bio-chauvinism that would insist on carbon-based dissipative structures. I explore this suggestion more completely below.

[9] To anticipate the discussion to come, this account of cognition, intrinsic teleology, and precariousness will become increasingly central to the view of artificial agency I explore in this work (discussed in Chapter 6). As I have explored in previous work (see Harrison et al. 2022), I propose that at least one potential path to the development of

We will return to this discussion later, as it is critical for the account of naturalized agency I explore in this work. For now, I summarize the present section by saying that BC and the biogenic approach is importantly directed towards the active materialities of embodiment that constrains, enables, and conditions cognitive function—thereby highlighting rich connections, isomorphisms, and symmetries between the processes associated with self-maintenance, repair, and development and the cognitive capacities required for their orchestration: memory, goal-directedness, and learning. I conclude this discussion on BC by introducing the nested notion of agency on which many discussions of higher-order agency depend.

1.3 A Society of Selves: Collective Intelligence and 'Cognition All the Way Down'

Equipped with the biogenic approach, its emphasis on life-mind continuity, and a definition of cognition that places an emphasis on dynamical processes of self-organization, precariousness, and existential needs, we can now turn to what I take to be a decisive and key second contribution of the BC framework: the notion of nested selves. This is important for the present work on naturalized agency because it establishes a gradualist and graded account of agential dynamics that stretches from humble unicellulars (including the somatic cells that constitute all of us) and the distinctive agency of the higher mammals—and perhaps, one day, to the development of AI and deployment of soft robots.

---

autonomous, genuinely cognitive robots is to incorporate the fragility and precariousness that typifies organic life and dissipative structures into the field of robotics and AI. In keeping with this, I fall closely in line with the program of soft robotics as laid out by Man & Damasio (2019). They write:

> We propose the design and construction of a new class of machines organized according to the principles of life-regulation, or homeostasis. These machines have physical constructions—bodies—that must be maintained within a narrow range of viability states and thus share some essential traits with living systems. The fundamental innovation of these machines is the introduction of risk-to-self. Rather than up-armouring or adding raw processing power to achieve resilience, we begin the design of these robots by, paradoxically, introducing vulnerability. (2019: 446)

This introduction of vulnerability and risk-to-self—of precariousness, we could say—might introduce such machines into the realm of feeling because it introduces imperatives for actions that might be the initial steps for worldly events to matter to the robot *as agent*. In this regard, it is interesting to note how Damasio & Man arrive at a domain of thought very close to BC despite working from the angle of Affective Human Neuroscience (e.g., Damasio & Carvalho 2013). For BC, too, "high-level cognition [is] an outgrowth of resources that originated to solve the ancient biological problem of homeostasis" (Man & Damasio 2019: 447).

In reading the BC literature, one is struck by just how many references there are to the agential capacities of individual cells—whether unicellular organisms or constituents of larger multicellular creatures—such that facultative and obligate multicellularity comes to be seen as a constellation of agents, a kind of collective intelligence, that differs from approaches that might take cognition as restricted to a higher-order agent. It is with this that we really arrive at the notion of 'cognition all the way down' (Levin & Dennett 2020). The idea here is that we should dislodge our understanding of cognition from a unique, indivisible cognitive agent (a higher-order self) and instead see higher-level cognition as founded on the orchestration of 'agents with agendas' all the way down.[10] Again, pulling from Levin, we see:

> Many examples of memory, anticipation, context-dependent decision-making, and learning are exhibited by organisms from yeast and bacteria to plants and somatic cells. … Single cells are very good at managing their morphology, behaviour, and physiology as needed for survival, altering their motility, and metabolism in response to, and proactively in, changing environmental conditions. They succeed in exploring their microenvironment toward optimal reproduction by selecting among numerous possible choices of gene expression patterns and behaviour. (Levin 2019: 4)

Crucially, we see that:

> While the mechanisms by which unicellular organisms' ability to accomplish specific adaptive ends is harnessed toward cooperative multicellularity is still poorly understood, one thing is clear: Somatic cells did not lose their behavioural plasticity and computational capabilities in becoming part of Metazoan swarms (bodies): They scaled them to enable pursuit of larger goals consisting of creation and upkeep of massively complex anatomies. (ibid)

Multicellularity on this account is depicted as a society of selves consisting of agents with agendas that scale up and down at a variety of levels of organization. Highlighting the nested nature of cognition as proposed here helps to reinforce a point laid out previously: namely: "In these integrated systems, intelligence, memory, learning, behaviour, and body structure are all intertwined and emerge from the multiscale dynamics" (Shah et al. 2020a: 1). As I explored in previous work, the nested approach to agency purchases a few concepts that are of relevance to the naturalization project pursued herein by providing a tractable manner for understanding emergence and top-down causation (Harrison et al. 2022). For instance, in Figure 2 (Levin 2019) below we see that the scale of goal-directed behaviour—understood in

[10] How far down is a matter of controversy for Michael Levin. I will venture my thoughts on this in a later chapter, but for now I propose we restrict our descent at the level of cells (though Levin believes Gene Regulatory Networks might qualify as well).

terms of possible perceptual states and ability for action—varies at the level of single cells, cellular aggregates (biofilms), and integrated (obligate) multicellularity. As we can see, the range of actionability for an individual cell is restricted by a sensory and perceptual range that limits its behavioural repertoire.

Cells that are coupled by gap junctions (a key biophysical mechanism that facilitates neurosensory and morphological coordination at varying orders of scale) are able to specialise to specific tasks, respond to local cues, and adjust to group-level dynamics and therefore respond to spatiotemporally more distal phenomena unavailable to a cell in isolation. Again, this reiterates the idea of multicellularity as a kind of 'swarm' intelligence that will be explored later. For now, I merely wish to highlight that the entraining of individual cells to a multicellular body that constrains the activity of constituent cells exemplifies a case of top-down causation in biology – a controversial, if pivotal, notion in the philosophy of science and theoretical biology (Noble 2012; Green & Batterman 2017; Green 2018, 2021)—a position that will be elaborated more fully in Chapter 3. This contribution on the part of the biogenic approach is critical to the reworking of materiality that I take as key to the naturalization of agency, which will occupy us in the next section.

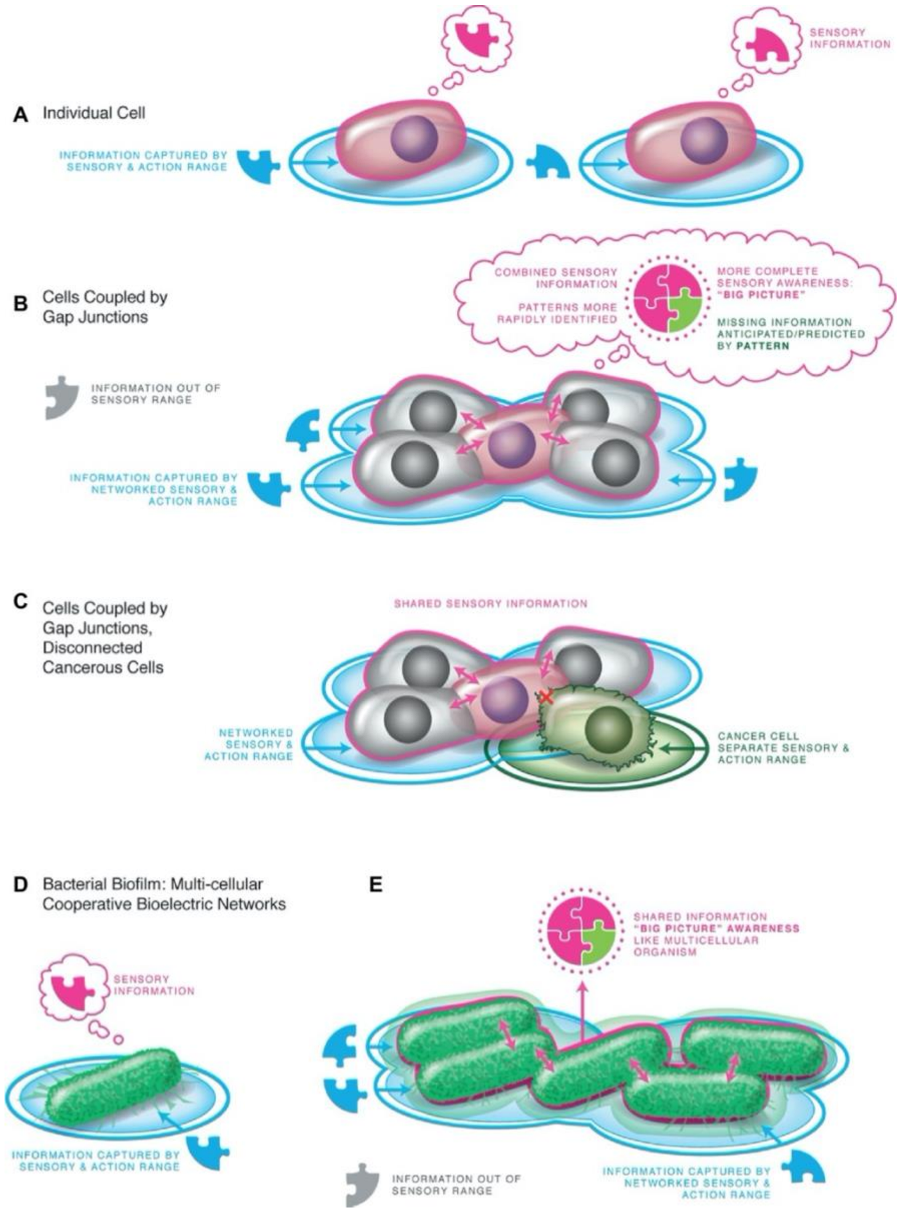

Figure 2. How cells interact collectively to pursue larger goals in space and time; an example is provided in (D) whereby cells in isolation lack the cohesiveness and scale to pursue the goals the collective can in coordination.

The foregoing discussion highlights how, for the biogenic approach, there are a cluster of concepts and features that are essential for substantiating its claims and motivating its philosophical relevance. To recapitulate, the biogenic approach furnishes an account of cognition and agency centred around the notion of precariousness and existential needs that places an emphasis on the far-from-equilibrium nature of the processes that sustain cognitive and living activity. In what follows, I will increasingly draw attention to the *material* nature of these processes, identifying the kinds of materialities that can enter into the orchestrated dynamics required for living activity. Importantly, this will not be an 'anything goes' account.

Perhaps this can be better articulated by introducing the contrast case I work against: namely, the entrenched attitude that cognitive activity can be divorced from its living, material substrate and that "we could be made of Swiss Cheese and it wouldn't matter", as Hilary Putnam memorably phrased it, so long as it is 'suitably organized' (e.g., the claims of multiple realizability; see Putnam 1971: 291). This comes down to the belief, as we saw in the Introduction, that the materiality and substrate of cognition is mostly a 'hardware' problem with the truly interesting explanandum being a cognitive 'software' that sits above—a contention that still permeates much of the philosophical and theoretical literature. However, as we will see below (section 1.4), much turns on what it means to be 'suitably organized' and it is by no means clear that any pell-mell set of materials could instantiate the complex dynamics on which cognition depends.

Secondly, the biogenic approach explores what it would mean to take top-down causation and emergence seriously. Arguably, this is one of the most important philosophical contributions of the biogenic approach as a non-reductive, non-eliminative, yet still naturalised account of agential dynamics and agential activities. Arguably, though, this is precisely what makes the biogenic approach controversial and in need of a philosophical engagement that places it in dialogue with several competing, dominant attitudes in the philosophy of biology, philosophy of mind, and philosophy of science more generally—where mention of emergence and top-down causation still trails more than a whiff of odium in its wake. As should be increasingly clear by now, my diagnosis for this entrenched attitude regarding anything teleological is that it stems from an outmoded conception of matter and materiality, one that lacks potential for higher-level causes or otherwise eliminates it from its conceptual remit. I believe this is an error in the philosophy of biology and science, and, moreover, one that is increasingly disjointed from how scientists in active matter physics, theoretical biology, materials science, and robotics work with and operationalize the concept. In the next section, then, I will sketch this diagnosis more fully as it will furnish the picture of materiality and mindedness I endeavour to work against throughout this work.

1.4 Working Against a Reductive View of Matter and Mind

> Both scientists and philosophers take ontological reduction for granted … Organisms are 'nothing but' atoms, and that is that.
> - David Hull, 1981: 281

One of the main difficulties facing an account of naturalized agency is accommodating a cluster of historically contentious concepts including emergence, top-down causation, and purposiveness in living systems. Accounting for these features—or otherwise indicating their compatibility with our naturalistic commitments to scientific praxis—is evidently a demanding task, but for my argument to be successful it is paramount to overcome (or to sketch how we might overcome) the teleophobia that has long been influential in the life and mind sciences. One way in which I propose to do this is to revisit several of our philosophical commitments on materialism and naturalism in the life and mind sciences. While Chapters 2 and 3 will provide a more robust proposal for overcoming teleophobia—and the broad prohibition on agency that results from it—this section provides some of the groundwork for understanding contemporary impasses regarding the nature of agency, and the role a concept of agency can play in scientific praxis, in the life and mind sciences. As will become clearer in Chapter 2, 'Origins of Organization': The Kantian Formulation and the Post-Kantian Response', I identify how the Kantian restrictions placed on teleology have left a problematic imprint on theoretical and philosophical investigations in biology—such that overcoming teleophobia in the life and mind sciences requires a re-engagement with Kant's analysis. Specifically, the 'materialism' I seek to problematise in this work stems in large part from the reductionistic mechanism (and its incompatibility with holism, emergentism, and agency) that is autochthonous to Kantian soil.

For now, I restrict the argument to a set of contemporary positions that reveal this kind of reductionistic materialism and a restriction to epistemic or heuristic approaches to agency in the life and mind sciences. For example, Daniel Dennett recapitulates a materialist credo when he remarks that we should understand the origins of minds and organismic diversity solely based on 'how we understand the material world'—with other approaches exemplifying problematic 'skyhooks', as he describes it in characteristically colourful language (Dennett 1993). In this regard, Dennett falls squarely within a philosophical tradition that seeks to expel agency from nature—whether from evolution, writ large, or as organismic agents that sculpt the conditions of evolvability. For him, the only admissible element here is 'brute mechanical capacities' and 'mindless motiveless mechanicity' and anything else would amount to "[smuggling] purpose back into biology" (Gould). But such an approach, as the organicist E.S. Russell once said, "leaves out of account all that is distinctive of life, the directiveness, orderliness and creativeness of organic activities, and completely disregards its psychological aspect" (Russell 1945: viii). As should be clear by now, I also think that we should understand the origination of organisms and mindedness based on 'how we understand the material world', but that in doing so what we find is not the predomination of 'mindless, motiveless mechanicity'—but a world dominated by active matter that exhibits prototypical directional and teleological dispositions in a spread that covers the organic domain and certain aspects of the inorganic world.

1.4.1 Reductionism, Physicalism, Emergentism

One of the main arguments against my position comes from reductive physicalism: the view that 'all there is' to the world is 'nothing but' the activity of fundamental particles (physicist Paul Davies calls this 'nothing buttery' (2006: 35)). It is a view that commonly denies the causal robustness of purportedly emergent features—from cells to purposive agents—and, by necessity, rejects any realist stance towards agency in nature, all things that the biogenic approach is currently challenging.[11] What is less often appreciated is just how much these two views have in common. Both view matter as a basically and inherently 'brute' or 'mindless' force that operates essentially blindly. That this is the view of matter animating much of the debates is visible when Fodor questions why there is anything *other than* physics (Fodor 1971). Or when Carl Gillet remarks on ontological reductionism that "proponents of physicalism claim that scientific evidence justifies us in taking the natural world to be a fully comprehensive compositional hierarchy bottoming out, so far as we know, with the entities of physics" (Gillet 2007: 206). Evidently, if matter and materiality are viewed through the lens of ontological reductionism then Jerry Fodor, Jaegwon Kim, and others are justified in wanting to 'save' mindedness from the abyss of matter—of 'mindless, motiveless mechanicity'. What underpins both the non-reductive physicalism of computationalism, then, and ontological reductionism exemplified in 20th century approaches to physicalism is a set of presuppositions regarding matter that depicts reality as basically composed of things, entities,

[11] This is encapsulated in Cricks' memorable 'astonishing hypothesis': "That 'You', your joys and your sorrows, your memories and your ambitions, your sense of identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules" (1994: 3). It is not immediately clear why causal powers are stable at the level of 'nerve cells and their associated molecules' and do not themselves drain out of the world to something more fundamental—something Jaegwon Kim calls the 'causal drainage' problem.

and substances. As might have become clear by now, the main metaphysical culprit here is that of a substance ontology.

Intriguingly, in a way I do think the world is 'just physics', but what that entails for Fodor (the need to save mind from mindless, motiveless mechanicity) and what that entails for my approach could not be more different. Whereas Fodorian 'just physics' is meant to indicate some absolute reductionism, from my approach, making recourse to physics is meant to capture an expansive and inflationary view of physical activities and material properties that actively accommodate the emergence of higher-level features and how they exert influence on lower-level activities.

For now, while I argue that a substance ontology, and the basically inert and crude view of matter it tends to promote, underlies both computationalist and reductionist approaches to mind, the former presents a more sophisticated and nuanced account of matter that requires a much deeper examination than can be offered here. Given my interests in connecting principles of BC and the biogenic approach to developments in AI and robotics in later chapters[12], I will reserve a more detailed assessment of computational approaches to mind until then. Nevertheless, I argue that in targeting the notion of substance and mechanism on which ontological reductionism depends we begin to rework and reformulate a notion of matter that is of greater relevance to debates in the philosophy of the life and mind sciences *and* is more representative of the state-of-the-art in physics.

Substance ontology is arguably the most common metaphysical position in the history of (Western) philosophy (Hegel once quipped that 'metaphysics is the tendency towards substance') and posits that beings are simple and hence internally undifferentiated and unchangeable (Seibt 2016) with the corresponding commitment that reality can be decomposed into particulates and discrete substance in which cause powers inhere. Moreover, causal powers are typically construed in a reductionist and linear fashion, such that not only are there discrete and localizable mechanisms, but these mechanisms interact in an essentially unidirectional and linear manner. For example, in their famous paper presenting neo-mechanism, Machamer and colleagues describe mechanistic decomposition in terms of identifying parts and their interactions with other parts as A → B → C→ with arrows representing a linear (and discrete) series of causes generating a higher-level phenomenon

[12] See chapter 6 for more detailed analysis.

(e.g., action potential generation or memory potentiation; see Machamer et al. 2000; Craver 2007).

David Lewis describes this world view well when he writes, "All there is to the world is a vast mosaic of local matters of fact, just one little thing after another" (Lewis 1986: xi). We see Jaegwon Kim, whose account of physicalism will occupy us herein, iterate a similar position when he writes that "bits of matter and their aggregates in space-time exhaust the world" (2005: 71). Each highlight that, from the perspective of substance ontology, the world is constituted by discrete and localizable parts and activities that can be progressively decomposed and in which, at bottom, there is some privileged level of causal activity. Thus, we see Craver write that "in its most austere and demanding form… [neo-mechanism] insists on a disenchanted world explicable without remainder in terms of basic causal principles" (2013: 1). Obviously, much depends on what is considered 'basic' here, but we can take the hint from its 'austere and demanding' nature that what is implied is some absolute base-level of reality consisting of fundamental particles. Accordingly, any descriptions we use to characterize higher-level features are merely the results of our epistemological approaches without elucidating any robust higher-level causality in the world *in se*.

Alexander Rosenberg presents this position when he asks "What is the world really like?", and replies in a way that cements this 'austere and demanding' worldview: "It's fermions and bosons and everything that can be made up of them and nothing that can't be made up of them" (2018: 19). Moreover, these fundamental particles are meant to *fix* all the facts, such that our inability to explain higher-level features in relation to them is a result of "conceptual incommensurabilities or epistemic limitations" (Rosenberg & Kaplan 2005) with no ontological implications (a point Soto et al. (2008) stress). Consequently, the agential features that we viewed in the previous section would be nothing but "the foresightless play of fermions and bosons producing… the *illusion* of purpose" (Rosenberg 2018: 26). It is difficult to overstate the strictness of this reductionist ontology—with even the redoubtable complexity of the brain "[being] still just fermions and bosons" (ibid.). Kim summarizes this position, writing that "All properties of a physical system supervene on, or are determined by, its microstructural properties" (1999: --). If this were the case, then there would indeed be much trouble for new-fangled approaches to biological agency: including the biogenic approach I propose herein.

Fortunately, I think there are several reasons to reject the view of physicality and materiality on which this species of reductive physicalism depends. First, I would like to emphasise that the view, e.g., Rosenberg insists on is a *metaphysical* thesis that works from a specific section of the physical sciences and treats fundamental and theoretical physics as the main and primary approach to the physical world. Such a thesis is then presented as the resultant of physics—the consequence of an 'austere and demanding' outlook or 'proper science'—but what I would like to suggest is that this is a metaphysics that has forgotten it is a metaphysics; and, indeed, a different reading of physics and materiality is possible. For instance, much of the literature I utilize is in the physical sciences, but from the areas of condensed and active matter physics that focuses on systems that are inherently emergent and relational (see Anderson (1975) in his famous essay 'More is Different') and occupy far-from-equilibrium dynamics, thus requiring a different physical language for describing their activities. These sciences do not privilege theoretical physics, and many authors (physicists!) argue that a defensible account of emergence and top-down causation is merited when examining these systems. Secondly, I want to stress just how important rehabilitating notions of emergence and top-down causation are for the present work. To achieve such a rehabilitation, we must embrace a view in which "our best physics tells us that there are no basic particulars, only fields and process", one that fleshes out "an ontology which gives priority to organization, which is inherently relational" (Campbell & Bickhard 2011: 33). With this, we begin to see some of the philosophical moves required for the naturalization of agency to succeed.

1.4.2 Nature, Naturalization, and Philosophies of Nature

> There is an important sense in which the only thing that doesn't seem to matter anymore is matter.
>
> - Karen Barad, *Meeting the Universe Halfway* (2007).

To recapitulate briefly, I have focused on the account of nature based on substance ontology—primacy of substances, mechanistic decomposition, particulates, particles, things—because I am interested in how the picture often associated with the elucidation of more 'fundamental' reality allegedly requires us to eschew teleological concepts or explanations. Associated with this is the contention that philosophy (especially metaphysics) and theories of cognition should be widely in keeping with what the sciences disclose—while this can be quite liberal, there is a broad background sense that entities revealed by the "hard sciences" are somehow more real, more causally powerful, and more metaphysically robust than those explored in the 'special' sciences. What I would like to now draw attention to is how the concept of Nature we assume constrains the possible arguments and philosophical moves available to naturalistic philosophers. That is, if one assumes (and it largely is left *assumed*) the nature of matter discussed briefly above—one devoid of directional or inherently teleological processes (even if confined to living organisms)—then what it means to 'naturalize' comes either very close to an eliminativism vis-à-vis agency (by showing how it can be decomposed into 'nothing but' its parts) or it approaches the *as if* accounts of teleology that opened up this thesis: treating organisms *as if* they are agential systems with such agential and teleological features awaiting replacement by a more 'complete' scientific description.[13] Everything, then, comes down to what one means by nature.

With this in mind, we begin to draw several of the philosophical, theoretical, and scientific threads of the argument together: I argue that the biogenic approach is not only directing attention to several areas neglected in the life and mind sciences (such as focusing on comparatively simpler organisms from which we might establish gradualist continuities between cognitive capacities found in diverse species) and thus presents a veritable research program for the would-be biologist or cognitive scientist; but—and this is where I primarily position myself in relation to the debate—as also engaging the life and mind sciences along a *metatheoretical* axis, to wit, that we must reconceptualise (some of) the terms of the debate: terms that are configurational for how we not only understand, conceptualize, explore or see the objects of our study; but also how we interpret the behaviour of those objects (organisms, cells, metabolisms). Dispelling the uncertainty behind, or otherwise demystifying, the claims of BC and the biogenic approach requires us not only sharpen the tools of empirical investigation, but also tinker with the field of ideas and concepts that tacitly and explicitly influence empirical research and theory-building.

In this regard, the present work sits closely to a range of other 'naturalization' projects in the philosophy of the life and mind science and is better called—in words of both contemporary and historical relevance; see Chapter 2—a *philosophy of nature*, which Godfrey-Smith describes as a philosophical position that seeks to better understand the overall picture scienctists (and other modes of inquiry) provide and account for broad-strokes *ways of seeing* the world. The hermeneutical and interpretive side of scientific praxis is as important as the discovery of new

[13] In this vein, it is intriguing that Neils Bohr considered the consequences of the new quantum theory to *open* up 'new prospects' regarding the 'position of *living organisms* in our picture of the world' (1931: 22), certainly not closing them off and even less entailing their elimination. For an insightful overview of Bohr's comments, see Hoyingen-Huene (1994).

facts and their configuration within theories, and a philosophy of nature (here, the biogenic approach) helps us frame work in relation to a set of persistent questions in the life and mind sciences regarding agency and the place of mind in nature.

For example, writing from the perspective of 'naturalizing phenomenology' ("How does consciousness and experience fit into our naturalist views of the world?") Shaun Gallagher has argued against a substance ontology picture of the world, wherein the organism, "if not life itself", has always presented an enigma to classic naturalism (2018: 128). For instance, if 'naturalism' is meant to imply some kind of reductionism (understood as intertheoretic reduction, à la Oppenheim & Putnam (1958)) then something being 'naturalized' means it should fit within this picture of theory-building as enabling bridge-laws and conceptual reductions. Although this is precisely the view of naturalism rejected by non-reductive physicalists, even these non-reductive approaches tend to hedge their bets by arguing for an heuristic or *as if* approach to the objects of the special sciences (including biology).[14] In other words, I take it that neither the reductive nor the non-reductive physicalist interested in the life and mind sciences takes the view of Nature that supports their metaphysics as particularly problematic. Rather, the latter merely do what they can to square away mindedness, agency, teleology, biological processes, and so on with an 'austere and demanding' approach to Nature present in reductive physicalism. This would mean that contemporary approaches offering alternatives in the form of emergentism are dogged by "an intellectual framework which perpetuates the metaphysical assumptions common to physicalism and Cartesianism" (Campbell & Bickhard 2011: 34). And, for this emergentist alternative to succeed, "those shared presuppositions [must be] rejected—in particular, the assumption that what is ontologically primary are things, entities" (ibid.).

It is the identifying, targeting, reconceptualising, and, in some cases, rejecting the presuppositions common to the previous metaphysical approaches that marks the biogenic approach, as I defend it herein, out as a metatheoretical project. Again, to get a flavour of this, consider Gallagher when he writes "A rethinking of the concept of nature itself, that is, a thinking that takes nature… not as a totality of objects, or particles, but as relational, situational, involving/including, irreducibly… would necessarily be a rethinking of the nature of science" (2018). While the larger (and more prickly) question of pheneomenology

[14] Fodor, for instance, himself wavers between metaphysical and epistemological interpretations of non-reductive physicalism (See Loewer 2009).

and experience is not the focus of the present work, which is the primary concern of philosophers like Gallagher, the commonalities in our projects should be clear: agency, just as much as phenomenology, is all but mysterious according to the standard naturalism and reductive materialism.

Moreover, it is here that the biogenic approach harmonises with new materialist philosophies and ontologies found in political science and feminist philosophy (Coole & Frost 2010). Briefly, new-materialism marks "a return to the notion of matter" (Butler 1993) to re-incorporate materiality as an important dialectical element in the social construction of (human) reality. It is an attempt at reworking several of the implicit assumptions that historically haunted debates on social-constructivism (see Bennett 2010; Hasslanger 2016), assumptions that either deprive materiality of any creative capacity or put it off as a reductive element from which social, human, and emergent activity had to be saved. Thus, despite working from very different angles and dealing with very different explananda, the biogenic approach converges with new materialism in giving priority to organization and relationality such that emergent features are ubiquitous across the organic and inorganic world:

> Inorganic matter-energy has a wider range of alternatives for the generation of structure than just simple phase transitions.… In other words, even the humblest forms of matter-energy have the potential for *self-organization* [sic] beyond the relatively simple type involved in the creation of crystals. … When put together, these forms of spontaneous structural generation suggests that inorganic matter is much more variable and creative than we ever imagined. And this insight into matter's inherent creativity needs to be fully incorporated into our [ontologies]. (de Landa 1997: 16)

What Manuel de Landa suggests here is central to my argument, and in fact is increasingly key to BC and the biogenic approach. It is an exploration into material processes that "makes manifest the extraordinary liveliness of the world" (Barad 2007: 91). Similarly, Nick Brancazio has recently argued for taking a relational ontology that emphasises interactive agential dynamics. This requires an engagement with active matter physics: "The fundamental force driving active matter systems derives from expenditure of energy at the level of the individual units involved, whether these are living entities … or non-living objects" (2022: 1). Active matter physics, then, furnishes an account of agency 'in the broadest possible sense': "an individual who is the causal source of asymmetry in differentiating itself from and acting towards its environment" (ibid.: 2). Thus, while it seems this discussion departs significantly from 'cognition', I argue that accounts thereof—especially insofar as they look at 'minimal' systems instantiating it—which rely on a notion of agency involving maintenance in far-from-equilibrium conditions merits a consideration of the metabolic, material, and physiological basis of proto-cognitive or agential behaviours (McGivern 2020).[15] With this, we see how the biogenic approach provides a more expansive account of cognition that begins with a demystification of how prototypically agential, teleological, and directional processes emerge from self-organizing, relational, and material ones.

1.5 Looking Forward: The Biogenic Approach and Revisiting Materialist Assumptions of Matter and Agency

To recapitulate, debates in Basal Cognition and the life and mind sciences more generally are typically dogged by two alternatives for understanding the place of directional, teleological, or subjective modalities of existence within a largely physicalist world disclosed in the natural sciences. The dichotomy tends to consist of: the coronation of neural evolution in the form of the human (or, at most, mammalian) brain on one side—and a kind of 'mindless, motiveless mechanicity' on the other side. In other words, when the bells and whistles of 'cognition' (restricted sense) are not present, then reductive mechanism predominates. When framed as such, a stark and contrasting picture holds philosophical debate in its grips, manifesting as a sharp contrast between the active, agential, and experiential modalities of being an embodied, living organism; and the dominion of a physicality where causal powers inhere at a level of 'fundamental physics'. Researchers in BC reasonably identify a set of interesting capacities and model organisms that they believe—and they are right—cannot be understood *reductively.* But I believe the urgency to label certain basal dispositions 'cognitive' is itself a symptom of the dichotomy that needs to be overcome. Effectively, the label 'cognition' is a placeholder in some cases for 'a process you previously thought was simple but is actually complex' (but see Figdor 2017)—in some ways serving the propagandistic function of drawing attention away from structures seemingly required for complex cognitive behaviours (such as neural architectures) to more 'basic' biomechanical and biochemical embodiments that are capable of redoubtable behavioural feats.

[15] Consider Patrick McGivern, whose account on active materials and minimal models of cognition will concern us later, when he writes on how "Recent work on active materials has highlighted a wide variety of ways in which non-living systems can exhibit many properties and behaviours normally associated with living systems" (2020: 442).

The point I would like to conclude on is as follows: I believe some of the tension that exist in the debate—between on the one hand those that believe something more elaborate is required for cognition and those who think even minimal, non-living systems can qualify as cognitive (e.g., Hanczyc & Ikegami 2010; McGivern 2020)—can be loosened if we shift the goal posts slightly away from a *thicker* sense of cognition to establish a more nuanced middle ground that focuses on an agential dimension, in the sense Brancazio suggests of " an individual who [being] the causal source of asymmetry in differentiating itself from and acting towards its environment" (2021: 2). While substantiating further why this qualifies as agential is a project stretched over the next 5 chapters (with an emphasis on Chapters 4 & 5), I introduce it now in order to promote a view wherein whether a given basal system qualifies as cognitive is reapproached from the perspective of agency, something I contend is a precondition for the elaborate cognitive capacities that form the bulk of the debate.

Chapter 2

Origins of Organization: Intrinsic Purposiveness, the Kantian Legacy, and the Post-Kantian Response

2.1 Intrinsic Purposiveness and the Spirit of Naturphilosophie[16]

> No inquiry has been surrounded, for the philosophers of every age, by so much darkness as that concerning the nature of matter. And yet insight into this question is necessary for all true philosophy, just as all false systems are shipwrecked from the very outset on this reef. Matter is the general seed of the universe, in which is concealed everything that evolves in later developments.
>
> - Friedrich J.W. Schelling, *Schelling Werke*

This chapter introduces and specifies one of the guiding concepts in this work: that of *intrinsic purposiveness* (as opposed to *extrinsic* purposiveness). This concept is key because it helps us identify systems that can sustain the agential-cum-cognitive activities that are of most importance to the biogenic approach to life and mind—specifically the organisms and systems that are of current interest to the field of Basal Cognition. Having suggested in Chapter 1 that the thermodynamics of *self-organization* is central to accounts of naturalised teleology, what I want to suggest is that the biogenic approach advocates for a notion of *intrinsic purposiveness* that ties cognitive and agential capacities of organisms to the existential dimensions of self-maintenance and self-organisation which implies a regime of order that exemplifies a unique mode of existence—the one most appropriately ascribed to the living domain and one that does not have an analogue in equilibrium processes in which mere objects reside[17] (Walsh 2018; Harold & Morowitz 2022). Thus, one key distinction intrinsic purposiveness is meant to carve is that between *organisms* (and, as we will explore in Chapters 4 & 6, intriguing non-organismic, but out-of-equilibrium, cases) and *artefacts* (e.g., the standard passive models often invoked as analogues in debates on the machine-organism

---

[16] This chapter is heavily indebted to my friends Andrea Gambarotto and Auguste Nahas, whose conversations and feedback at various stages of development moved this into a more rigorously structured chapter than I could have achieved without their insight.

[17] More specifically, we introduced the notion of *existential needs* and highlighted certain causal structures bound up in being an intentional or goal-directed system. What I will suggest in this chapter is that existential needs can be fleshed out further by saying that only those systems that exhibit *intrinsic purposiveness* have *existential needs* in a way that leads to the forms of agency that concern me in this work.

distinction; cf. Nicholson 2013).[18] Before proceeding to the empirically oriented chapters, then, it is important to emphasise what I believe this merits: a return to the Kantian formulation of purposiveness (and its naturalisation via self-organization).

This is key as the Kantian notion is arguably the cornerstone of biological and theoretical discussions on the naturalised agency in the life and mind sciences—even if only as a namesake (Maturana & Varela 1979; Varela 1977; Varela et al. 1993; Thompson 2007; Huneman 2001; Moreno & Mossio 2015; Newman & Moss 2013; Mossio & Bich 2017; Weber & Varela 2002; Wolfe 2017, 2022). Not only was Kant the originator of the intrinsic-extrinsic purposiveness distinction, but he also introduced the notion of self-organization as a central concept to render it naturalistically amenable to empirical inquiry.[19] Although I ultimately value the intrinsic/extrinsic distinction as Kant formulates it (indeed, Hegel once called it "Kant's great service to philosophy" [2010: 654/6:440-441]), my ultimate aim is to encourage a wider reframing of the debate that requires moving *beyond* Kant into (what I find to be) a more fertile philosophical landscape: one we find in Post-Kantian *Naturphilosophie* and the historically maligned Romantic Biology, the spirit of which I seek to champion in this work. Getting there will take some steps, however, so I will begin in order.

What is fascinating about the Kantian formulation—and arguably what has led to the most spilt-ink concerning its status—is the simultaneously hopeful and yet sceptical conclusion that Kant draws in the *Critique of the Power of Judgment* (from hereon, simply the *Third Critique*[20]), a tension that results in the (in)famous *antinomy of teleological judgment.* Simply put, this comes down on the fact that we must, as a matter of heuristic value, *judge* or assume that organisms exhibit a purposiveness whereby the wholes are the cause of the parts and the parts the cause of the whole, to wit, that organisms[21] are self-organizing and holistic systems motivated by an intrinsic end towards their perpetuation, nourishment, and reproduction.

[18] The Machine-Organism distinction receives further articulation in Chapter 6 in the chapter on Soft Robotics.
[19] Intriguingly, as we will explore later, these debates were centrally located at the origin of the life sciences as a scientific discipline and, although they no longer assume the centrality they once did, have always represented an alternative way of cognising and construing organisms and agents within the biological sciences (Zammitto 2002; Riskin 2017).

[20] So-called because it constituted the third and last main text of Kant's critical oeuvre.

[21] This word, while present sparingly in the discourse at the time, is not used by Kant where he instead talks of self-organizing *beings*. However, the extension of the latter term has discursively been wound up in the notion of organism and is textually supported by the examples Kant himself provides. I will stick with the term 'organism' for simplicity.

However, after seemingly defending the unique status of organisms, Kant (in almost the same breath) undermines the ontological status of these systems. He remarks that the best physics of the time, Newtonian mechanics, not only sets the standard and metric for legitimate scientific explanations, but posits the primacy of mechanism and the causal primacy of the parts: parts cause wholes and not the other way around (something called *efficient causation*, a fulcrum of the current chapter). This leads Kant to conclude that we can never know if organisms are truly animated by this intrinsic purposiveness and must instead treat them *as if* [*als ob*] they are teleologically structured. We will have time to get into this further below (see section [2.2 The Teleological Judgement]), but for now I would like to stress that this tension still persists in our contemporary ways of theorising agency and organisms (as we saw in Chapter 1). That is, there is a tension between those who believe the agential status of organisms is nothing more than a heuristic *gloss* pasted over an otherwise mechanistic world. *As if* teleology is arguably one of the defining features of 20th century life sciences (cf. Gambarotto & Nahas 2022), making a reconsideration of its origin and subsequent insinuation in biological debates even more worthwhile.

What this serves to highlight, then, is certain conceptual and argumentative isomorphisms that stretch between Kant's prescient articulation of intrinsic purposiveness and our own deep-seated conceptual anxieties surrounding agency, teleology, and goal-directedness, such that "investigating the way Kant's legacy is currently understood provides important insights on the general stances towards teleology currently at play in our philosophical landscape" (Gambarotto & Nahas 2022: 48). Indeed, continuing the theme from the previous chapter, we see that one of the main impasses—stretching like an Ariadne's thread from Kant to contemporary theoretical and philosophical practice—within the Kantian formulation is the commitment to a certain species of physicalism dependent on a ('Newtonian') mechanism and model of efficient causation. A term of art often employed in this area, efficient causation is a dominant model (or presupposition) in the literature which posits a connection that "constitute a series (of causes and effects) that is always descending", implying that every effect has a cause that is necessarily external and antecedent to it (Nassar 2022: 38). The first part of this chapter will break down the mechanistic and physicalist assumptions underpinning Kant's formulation of intrinsic purposiveness as a tool we can then transfer to examining our own presuppositions about the nature and behaviour of matter.

Much of the current literature on autopoiesis, biological autonomy, and intrinsic purposiveness explicitly begins from Kant's formulation of self-organization and intrinsic purposiveness—and there is a "peculiar resonance between Kant's treatment of teleology and [our] continued struggle to make sense of the place of [the organism] within the life sciences" (ibid.). Accordingly, Charles Wolfe remarks that there is "a curious consensus without examination of presuppositions [that] reigns today in the various sub-fields concerned with the properties (organizational, systemic, etc.) of living beings", which involves invoking the authority of Kant's formulation while equipped with the resources of modern scientific practice (normally, some deployment of dynamical and complex systems theory; Wolfe 2022). Evidently, Kant, or at least a Kantian invocation, still exerts immense gravity over debates on naturalized teleology. This, then, provides a clue as to why I find an engagement with Kant to be pertinent to my discussion on materialism and agency in current discourse on agency and mind: namely, because "The framework within which the philosophical encounter with nature remains, after two hundred years, that established by Kant's *Teleological Judgment*" (Grant 2006: 17). Given the pervasiveness of Kantian concepts, let alone invocations and references, in the literature, more detailed engagement with the Kantian formulation seems paramount for understanding the promise of many projects that *depart* therefrom.

More perspicuously, this 'peculiar resonance' and 'curious consensus' comes down on two interrelated components: First, "an invocation of Kant … usually presented as an epistemological component" and an ontological delineation concerning "the inherent features of organisms" (Wolfe 2022), such as self-maintenance, occupying far-from-equilibrium conditions, having existential needs, and so on. Wolfe continues (and I concur),

> I suggest we have a problem here, verging on a category mistake: it is not quite right to invoke the authority of the Kantian 'projective' approach in order to assert a set of ontological specificities about organisms. First, because this is precisely what the Kantian regulative ideal [eternalised with the notion of *as if teleology*] was designed to avoid… Second, because it is not clear in any case why it counts as an argument against 'mechanism' or 'reductionism' to say: here is a list of key features [the ontological features I described above, e.g.]. (Wolfe 2022)

This is not to say that certain components of the Kantian formulation do not hold relevance to contemporary debates or discussions vis-à-vis naturalizing agency and intrinsic purposiveness (very much the opposite, in fact). Building off (and perhaps into a slightly different direction from) Wolfe's argument, I rather want to suggest that we cannot ultimately leave the presuppositions latent in these approaches *unexamined*, and that, once we interrogate the Kantian formulation, what we find are several tensions native to that context which complicate these projects. This leads to a set of interrelated claims to be explored in this chapter: that our current theorising on organismic agency and naturalised teleology is hindered by remaining with the Kantian formulation; that a re-examination of certain elements present therein (which I contend still animate current discourse) is required; and, finally, that moving *beyond* Kant could open a philosophical landscape that allows us to better conceptualise, theorise, and interrogate the role and place of agency in nature.

But what does this landscape look like, and what perspectives does it afford? Encouraging a pivot away from Kant into what I find to be a more fertile landscape leads me to pull from an unlikely source in the history of philosophy and science: that of *Naturphilosophie* and Romantic Biology[22], theoretical camps that included philosophers and experimental scientists working during the wake of Kant's Critical publications (from the time of the publication of the *First Critique* in 1781 to the *Third Critique* in 1793)—with my specific interest in the philosophical work of Friedrich W.J. Schelling [1775-1854], but also, *inter alia*, Johann Gottfried Herder [1744-1803], Johann Wolfgang von Goethe [1749-1832], and Alexander von Humboldt [1769-1859]. I will explore how these philosophers and naturalists were keenly interested in the problems of teleology, life, and biology set up in the *Third Critique*[23] and that (some of) their projects can be understand as an attempt to work beyond the Kantian strictures on teleology as set out in the *Critique of the Power of Judgment*, and, especially in the case of Schelling, the view of matter and causality expounded in Kant's *Metaphysical Foundations of Natural Science* [1786/2010] and *Third Critique*. Moreover, in commonality with contemporary process philosophy (Dupre & O'Malley 2006), these authors were interested in espousing a metaphysics that could accommodate relationality, organization, and holism—thereby explicitly working against various forms of reductionism and mechanism that arguably still hold sway in metaphysics.[24] The main insight I would like to take away from

[22] So-called because it covers a suite of concepts and theoretical concerns that animated the Naturphilosophen yet formed legitimate research programs in the scientific practice of its time.

[23] Herder is a bit more of an exception in this list, having been more contemporary with Immanuel Kant.

[24] A metaphysics that prioritises relationality is key for a robust substantiation of emergence. We have seen in Chapter 1 that anti-realist attitudes towards agency and teleology tend to eschew relations as ontologically primitive, and thus restoring relationality to its own legitimate scientific status is a key move of this work.

the *Naturphilosophen* is their insistence that Nature *looks* different depending on the methodologies, concepts, and ideas we bring to our investigation (Nassar 2022). This plugs naturally into my contention, put forth in Chapter 1, that the biogenic approach offers a new way of *viewing* agency, goal-directedness, and organisms not afforded by a scientific picture that expels these features from an ontology of nature.

In light of this and given the concerted effort to rework the Kantian formulation—not to mention Kant's own dissatisfaction with the antinomy as settled in the *Third Critique* (van den Berg 2014)—it seems even more astonishing that we rest content with Kant's formulation, not least because it was one of the central sources of tension in its own time. My approach in this work is better seen as having closer affinity with the many strands of *Post*-Kantianism explored in the latter part of this chapter, reminiscent and inspired as it is by *Naturphilosophie*. It is precisely the rethinking of the metaphysical basis of mechanism, materialism, and reductionism that are autochthonous to the post-Kantian context that encourages me to revisit Kant and his critical reception. In doing so, I hope to paint an image of this relatively un-(or under-)explored philosophical landscape and sketch an alternative path in the history of science and philosophy that sought to accommodate and naturalize (in a non-eliminative way) the agentic dimension of organisms—rather than expelling them from a scientific world view that seems to follow inexorably from naturalism.

The remainder of this chapter will be structured as follows. Section 2.2 explores the 'Teleological Judgment' of the *Third Critique* with a specific emphasis on the notion of intrinsic purposiveness. Section 2.3 analyses in further depth the physicalism, reductive causal model, and understanding of materiality (famously, one as passive, brute, inert, and crude) that prevents Kant from a more thoroughgoing ontologising of the organism. This will set the stage for Section 2.4, where I introduce some background to *Naturephilosophie* and Romantic Biology (specifically as encapsulated in the work of Schelling). Section 2.5 overviews Schelling's prescient rethinking of the nature of matter as a way to naturalise teleology. Section 2.6 concludes by synthesising and translating the discussion into a contemporary context. I will highlight the conceptual affinities between the rethinking of nature and matter that is arguably the main theoretical contribution of *Naturphilosophie* and the contemporary impasses and anxieties surrounding naturalised forms of agency. Before proceeding, I give one last preamble for why this section of the work—immersed as they are with philosophical and historical conditions of our thinking on agency and teleology—are of relevant to my wider argument and the more general audience.

2.1.1 A Small Preamble: Why History Matters for Debates on Naturalised Teleology

History will get you eventually, if your eyes are open.
- Hasok Chang, *Is Water H2O*, xxi

I have already stated that much of the debate on naturalising teleology, from a biological point of view, begins from Kant's formulation of intrinsic-extrinsic purposiveness. It is less clear, however, why I need engage this historical front at all. The sceptical philosopher or scientist might reasonably ask: Why does it matter? Or, as might be some reader's wont, can I just skip this section as window-dressing and proceed to the later, more empirical chapters? I believe this would be to miss the point of what is being advocated in this work, and so, for the sake of clarity, I repeat: what interests me in this thesis is the overall shape of the conceptual landscape in which researchers, experimental scientists, philosophers, and theoreticians operate *today*. It is in this sense that I construed (in the Introduction) the biogenic approach as a *Philosophy of Nature*: it examines, explores, and enquires into the picture of the world proffered by the sciences *in tandem with* how certain philosophical presuppositions, social attitudes, and metatheoretical assumptions frame or otherwise mediate these various *Weltanschauung*. This more 'zoomed-out' perspective is where I believe philosophers, historians, and social scientists are uniquely positioned for investigating the origination, sedimentation, and solidification of a particular set of attitudes vis-à-vis nature and scientific practice.

Thus, to pull a leaf from integrated history and philosophy of science research, what I am primarily interested in here is how we have *inherited* a deep scepticism towards agency in nature—rather than this scepticism being a natural product of an unblighted scientific (physicalist!) view of the world. That is, its incompatibility with our naturalistic commitments has long been taken as *prima facie* accepted such that overcoming any impasses in debates on teleology does indeed require examining how such an attitude or worldview has become so unanimously accepted (Rosen 1991). Stated differently, questions concerning the nature of agency and life are of course scientific questions or open to scientific analysis; but tensions and antipathies within debates on the nature, place, and role of agency "are only symptoms of a far deeper situation" (Rosen 1991: 2)—and investigating the roots of what is involved in such a situation is one piece to the picture I paint in this work.[25]

Another way of putting this point could be as follows: science is as much a hermeneutical practice as it is an experimental one. Again, as Rosen writes, "a fact or datum, by itself, is essentially meaningless; it is only the *interpretation* assigned to it that has significance" (1991: 17). We receive specific assumptions and presuppositions that become 'sedimented' in praxis, theory, ways of seeing the world—and these conditions have a historical becoming that can be unrooted and dislodged, not least by interrogating how we came to view the world in the way we have. In other words, the *methods* and *assumptions* we bring to, not just experimentation, but interpretation can hinder the process: "in [many] cases, it was the absence, not of data, but of imagination that created difficulty [for advancing science]" (ibid.). As Dalia Nassar notes in her analysis of Herder, this is precisely where a key insight of the *Third Critique* shines forth: "If there is one lesson to be taken from Kant's *Critique of the Power of Judgment*, it is that nature *appears* differently, depending on the tools, assumptions, and methods deployed" (2022: 46).

All of this, then, points to the importance of revisiting a series of episodes in the history of philosophy and science around the time of the development of biological science at the end of the 18th and beginning of the 19th centuries when two viable paths were under exploration: one of a reductionist and eliminativist attitude towards agency, emergence, and top-down causation; and one that actively sought to re-examine the presuppositions leading to this difficulty, accommodate teleology and agency, and conceptually re-engineer the notion of mechanism and matter on which the former position depended. In this vein, there has been a large body of literature in the past four decades that has explored the Kantian legacy and its insinuation in various modes of philosophical thought up to today—as well as the alternative

[25] Rosen considers the cognate case of experimental physics' push to ever greater 'generality' or universality, writing that its inability to branch into, or otherwise accommodate, biological phenomena stems from certain hypotheses, assumptions, and presuppositions that are auxiliary to the scientific process itself. He writes, "it becomes a matter of finding, and removing, whatever tacit hypotheses are limiting the generality of contemporary physics in these directions. We cannot find them by looking at contemporary physics itself; it, and everything in it, are already the consequences of imposing these very hypotheses" (Rosen 1991: 38). To ameliorate the situation, Rosen suggests that we return to a "conceptual stage" to explore how physical and biological practice embodies a set of 'restrictive hypotheses' that need to be re-examined. As should be becoming clear, I propose and promote a similar re-examination of the conceptual stage here, arguing that, although we stand two centuries apart from Kant, much contemporary philosophical work on agency and teleology nevertheless still embodies these 'restrictive hypotheses'.

approach to philosophers and life scientists at the turn of the 19th century that promoted an active materiality for overcoming the antinomies latent in the Kantian formulation.

The *locus classicus* of this analysis is Timothy Lenoir's *The Strategy of Life* (1981), although problems with Lenoir's historical exegesis have now long been established (Zammito 2012). Better known might be the works of Robert Richards who, across an entire career and several books (1993, 2002, 2007), has explored the Romantic and *Naturphilosophie* context in which the biological sciences congealed and evolutionary theory flowered—and his historiographical work is integral to my own reconstruction herein. Adding to this, Fredrick Beiser (2002), Maurizio Esposito (2013), Jessica Riskin (2016), Andrea Wulff (2017), Andrea Gambarotto (2018), John Zammitto (2020), and Dalia Nassar (2022) have all dedicated book-length treatments to recovering particular facets of *Naturphilosophie* and an 'active mechanism' with notable similarities to the view I promote in this work.

Jointly, this scholarly work points towards a "paradox at the heart of modern science", one of "particular significance for scientific accounts of life and mind" (Riskin 2016: 337). Jessica Riskin, in a chapter titled "History Matters", remarks that this 'paradox' originated in the 17th century when modern science and the machine model began to take shape: a view that banished agency, sentience, and purposiveness from 'merely' mechanistic processes—whether proprietary to organisms (a world not coined until Leibniz) or to a teleological understanding of the cosmos. It also saw the provincialisation of agency to a strictly human domain (where it largely remains today[26]).

In her work, Riskin seeks to discover and rehabilitate a competing scientific narrative consisting of a "parallel development… that naturalized rather than exported purpose and agency" (ibid.). Riskin's analysis has clear implications for the strategy I pursue herein, not least because it shows how the "history of these competing forms of science is profoundly relevant to current debates in Artificial Intelligence, cognitive science, and evolutionary

[26] As Moss & Nicholson (2012: 88) note, the assumption that agency is proprietary to the human domain is still taken as unquestioned in Anglophone philosophy. Indeed, a recent collection of work, edited by De Caro and Putnam, explores naturalistic approaches to agency and normativity that never 'dares to ask': "is normativity really exclusive to the human domain?". Moreover, "normativity, for these thinkers, is identified with the realm of the human" (ibid.). What is required, from their perspective, "is that of a more radical re-thinking of just what is the relationship of normativity to nature all the way down. Must our point of departure always be at the level of normativity as a distinctively human phenomenon? Might it not be the case that it is just such a presupposition that dooms the conciliatory enterprise to failure from the start?" (ibid.).

biology" (2016: 338). She highlights a pervasive and deep-seated principle, "so elemental as to be essentially invisible" (ibid.), that agency cannot be a primitive, elemental feature of the natural world and that, as a corollary, "the idea of natural forms of agency has no place in legitimate science". But as I have remarked above, this is a *Weltanschauung*, not a necessary consequence of how we approach or study the objects of biological or humanistic inquiry. Moreover, given the deluge of research seeking to account for causally robust forms of natural agency (see Chapter 1, 4, and 5) it is becoming increasingly pertinent to find ways to *accommodate*, not eliminate, agentic, directional, and normative dimensions of cognitive embodiment. The upshot, then, is this: "by recognizing the historical roots of the almost unanimous conviction … that science must not attribute any kind of agency to natural phenomena, and recuperating the historical presence of a competing tradition, we can measure the limits of current scientific discussion, and, possibly, think beyond them" (Riskin 2016: 338-339).

This is a view that is in dialogue with that of 'complimentary science' which Hasok Chang proposes (2014). Writing in a much different historiographical context on the Chemical Revolution that Lavoisier initiates (and the details do not concern us here), Chang remarks that, even by the perspective of 18^{th}-19^{th} century science, Lavoisier's oxygenist chemistry (2014: 8-12) left much to be desired—despite its overwhelming triumph over the competing phlogiston theory. The latter theory accommodated certain aspects of chemical theory that the oxygenist model eliminated or elided. In fact, Chang argues, co-maintaining both theories would have enabled a more productive theoretical space in which to explore features of the Chemical Revolution and address certain issues for which the oxygenist model alone could not account. Instead, however, of proceeding in this manner, the phlogiston theory according to Chang "*was prematurely killed*" (2014: 12). As he writes further,

> What I seek is a complete view: I want to know what contributions to scientific knowledge the phlogiston theory did make, what contributions it could have made if it had been kept longer, and what contributions it could still make if it were revived. If all those categories of contributions have been lost or missed because of the premature abandonment of phlogiston theory, then we should recover, imagine and create them. If you would object that such an enterprise is neither history nor philosophy nor real science, so be it: I call it 'complementary science'. (Chang 2014: 12).

I would like to follow suit here. Let us revive certain elements, modes of viewing, and regulative concepts in this much maligned domain of science and see what *could* come from it.

The particular facet of *Naturphilosophie* that interests me here is the revisiting of matter theory that is central to Schelling's own work.

What I am suggesting, then, is that we look at certain elements of *Naturphilosophie* and Romantic science, or what Dalia Nassar calls Romantic Empiricism (2022), from a similar perspective: as a historical, scientific vantage point from which problems of agency, normativity, and teleology become less mysterious and transformed into veritable objects of investigation. It is undeniable that *Naturphilosophie* trails a whiff of odium in its wake when discussed—and its scientific irrelevance was all but a consensus in 20th century philosophy and biology (see Ernst Mayr's *Growth of Biological Thought* (1982: 388), where he dismisses the ideas of *Naturphilosophie* as 'fantastic if not ludicrous'). The antipathy towards this vein of scientific and philosophical thought in part stems from the complex and arcane metaphysics that has been seen as privileging speculation and "a denial of the value of experimentation" (Snelders 1970).

But this is historical caricature—not fact, and various reconstructions of the relationship between *Naturphilosophie* and science are now available that cast doubt on this common telling. Indeed, sometimes the triumph over *Naturphilosophie* is seen as the origin story of the biological sciences and evolutionary theory—but this is no longer tenable, both on exegetical and philosophical grounds. Ron Amundson, in his influential history of developmental biology, challenges these entrenched attitudes, remarking that biologists working within a Modern Synthesis vein reject Naturphilosophie (Amundson treats it more generally as *Idealism*) due to its speculative flare and generation of 'faulty scientific theories' (2002: 18). Again, this is simply not true—and whatever disagreement there is about the nuances or particularities of *Naturphilosophie* and Idealism, "it has given rise to extremely important scientific theories. These theories played a crucial role in the history of evolutionary thought, and they do not deserve the disdain to which they have so long been subject" (ibid.).

What I hope to accomplish over this chapter, then, is to spark a discussion in which a rehabilitated---and suitably updated—*Naturphilosophie*, that is, a *Philosophy of Nature*, could weigh in on the many debates and issues surrounding the problems of agency, teleology, and normativity in the life and mind sciences. To sketch, in other words, an alternative and competing approach that generates a way of seeing the world in which these concepts lose their mysterious character slightly. Pulling again from Gallagher, this amounts to a "rethinking of the concept of nature itself", that is, "a "thinking that takes nature not as a totality of objects, or particles, but as relational, situational, involving/including…agents" something that would amount to a partial "rethinking of the nature of science" (2018: 134). Indeed, as we saw [Chapter 1: 'Nature, Naturalisation, and Philosophies of Nature'], what ones means by naturalisation depends directly on what one means by nature (ibid.), making the *Naturphilosophie* episode even more illuminating. What I want to suggest then is that playing the different formulations of naturalism found in Kantianism and the Post-Kantianism against each other demonstrates this point: the nature of Kant and the nature of Schellingian *Naturphilosophie* are indeed very different, and the objects, phenomena, and sets of relations that can be taken as legitimate scientific targets are, correspondingly, very different. My allegiances lie with the latter, but it is undeniable the Kantian context sowed the seeds for the many relevant projects of *Naturphilosophe*: most notably, the centrality of *intrinsic purposiveness*. Having said this, I now turn to the conception of intrinsic purposiveness found in the *Third Critique* before identifying the problematic elements that have historically blocked taking this dimension of organismality more seriously.

2.2 The Teleological Judgment, Intrinsic Purposiveness, and the Primacy of Mechanism

Choosing where to start with as monumental a text as the *Third Critique* is a formidable task—and the work itself does indeed present itself as a conical mountain that affords many different sides for approach. Be that as it may, I restrict myself to but a slender portion of this immensely interesting work in order to highlight the problems that, in relief, would be most pressing for the Post-Kantian *Naturphilosophen*, the program of Romantic Biology, and subsequent debates on the nature of mechanism and place of subjectivity in nature. I will proceed as follows: first, [Section 2.2.1] I begin briefly with the notion of intrinsic purposiveness and what Kant calls *natural purposes* [*Naturzweck*] before proceeding to [Section 2.2.2] where I explicate the notion of mechanism on which Kant's account—and ultimately the antinomy—depends.

2.2.1 Intrinsic versus Extrinsic Purposiveness

Before continuing to the intrinsic-extrinsic distinction, it is important to familiarise ourselves with the context of Kant's Teleological Judgment. In it, Kant is concerned with the *faculty of judging*, that is, our ability to perceive and judge something *as* something without being committed to the ontological status of the thing thereby judged (a project mainly associated with the *Critique of Pure Reason*). One of the main objects of concern in the Teleological Judgment section is that between *natural products* and *natural purposes*. Kant makes clear early on [CJ §64: 369], when we take natural products as natural purposes we suppose they possess a different causal structure, namely a teleological one: "To say a thing is possible only as a purpose is to say that the causality that gave rise to it must be sought, not in the mechanism of nature, but in a cause whose ability to act is determined by concepts".

It is interesting to reflect on what Kant means here by 'determined by concepts', as the insistence on the causal role of *concepts* is what partially leads Kant to reject that natural products can be natural purposes because it would imply the causal efficacy of a divine creator—something Kant rightly rejects as impermissible for empirical investigation. Kant illustrates this point of 'determined by concepts' with that of a geometric figure in the sand. Say you are walking along a beach on the Gold Coast of Australia and come across a triangle sketched in the sand. We would never initially judge such a figure to have appeared by a cause merely in accordance with blind mechanical forces (it is not impossible, but vanishingly unlikely that such a figure could be produced by the activities of wind and waves alone) but would probably suppose a (perhaps eccentric) person had created the figure. But this would be a *natural product* for Kant, and it is ultimately the notion of natural purpose that concerns him.

The key difference between the two, for Kant, comes down to the fact that with the geometric figure we suppose its construction—its *cause*—is something *distinct from the figure itself* (Sedgwick 2012: 48). As Sedgwick helpfully elaborates: "We do not suppose in the case of artifacts … that their parts are themselves in any way responsible for the purpose or form of the whole" (ibid.). Instead, we assume that a concept *extrinsic* to the artifact determines the presence of the parts – and, in turn, the whole. Another way of putting this is that we could never divine the cause of the figure through an identification of the parts themselves. Instead, we assume that the cause of the figure is not the product of 'nature' at all but, rather, 'art' ([CJ§ 64: 370]). Thus, as Kant clarifies, the cause of the figure in the sand—or any artifact, for that matter—"is distinct from the matter of the thing… [and] distinct from a thing's parts" ([CJ § 65 [373]]).

With this, I believe, we begin to see the defining feature of a *natural purpose*. Pulling from Sedgwick again, we see that "the distinguishing feature of a [natural purpose] … is that it has 'in itself and its inner possibility' [*in sich selbst und seiner inneren Möglichkeit*] relation to purposes" (2012: 48). To state this differently, natural purposes—paradigmatically, organisms—are not determined by concepts extraneous to their parts and activities but rather are "*organized* and *self-organizing* being[s]" (CJ §65 [374]). Kant elaborates:

> In such a product of nature, each part, at the same time as it exists throughout all the others, is thought as existing with respect to the other parts and the whole, namely as an instrument (organ). That is nevertheless not enough (because it could be merely an instrument of art and represented as possible only as a purpose in general); the part is thought of as an organ producing the other parts (and consequently each part as producing the others reciprocally). Namely, the part cannot be any instrument of art, but only an instrument of nature, which provides the matter to all instruments… It is then—and for this sole reason—that such a product as organised and organising itself, can be called a natural purpose. (CJ, § 65: 374)

This passage drives home why Kant believes the notion of a natural purpose implies a *different* model of causality: natural purposes are not determined through blind mechanical forces (for then the fine-tuning and intimate relations between parts would be absolutely contingent, something Kant thinks we cannot even understand), nor can we understand them as produced through concepts of a rational being; rather, we must posit a *recursive causality*, one whereby the parts come into existence due to the whole and the whole due to the parts. In a phrase, natural purposes are *causes and effects of themselves.*

However, immediately after suggesting this recursive causality, Kant cautions that this concept of "self-organization" is 'inscrutable' and mysterious: it is a form of causality that "has nothing analogous to any causality known to us" ([CJ §65 [374]]). To anticipate the coming discussion, Kant boxes himself into this position by remaining stalwartly committed to a view of Newtonian mechanism and a reductive causal model (efficient causality, which we will return to below), such that positing a self-organizing, recursive causality would contradict 'the essential character of matter': "Lifelessness, *inertia*, constitutes the essential character of matter" (CJ: [394]). Materiality, in other words, lacks 'in itself and its inner possibility', the 'inner animating principle' required to be lifted out of mechanism and generate the diversity of forms of elaborate structures that exemplify the biological world, and which were fascinating naturalists in the 18^{th} century. The fact that Kant contends we can never know how, or from which source, this diversity sprung led him to his famous proclamation that there could never be a Newton of the blade of grass—something which Schelling categorically rejects as "accomplish[ing] nothing more than to crush the ambition of the researcher" [1798: vi].[27]

Returning to the specification of a natural purpose, Kant illustrates what he means by something being cause and effect of itself with the example of a tree. First and foremost, a tree can be considered as cause and effect of itself as a member of its species: it is itself the product of previous generation of its species and brings forth another generation via reproduction. So, they are cause and effect of themselves by virtue of membership in a species. Second, the tree reproduces itself as an individual via processes of growth and repair. Lastly, a natural purpose is cause and effect of itself insofar as its parts depend on each other for their existence—their involvement in this or that organized being and self-organizing being is not extrinsic or, as is sometimes said, 'indifferent' to the whole. As he writes, for a natural purpose and its parts "(as it concerns both their existence and their form) must depend on their relation to a whole" [CJ § 64: 369]. Kant therefore concludes that for something to be judged as a natural purpose "in itself and in accordance with its internal possibility, it is required that its parts reciprocally produce each other, as far as both their form and their combination is concerned, and thus produce a whole out of their own causality" ([CJ § 65: 373]). That is, a natural purpose is where "each part is conceived as if it exists only *through* all the others, thus as if existing *for the sake of the others* and *on account of* the whole… only then and on that

[27]Kant's exact words are: "indeed this is so certain that we can boldly say that it would be absurd for humans even to make such an attempt or to hope that there may yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws that no intention has ordered; rather, we must absolutely deny this insight to human beings" (CJ § 75: 399). It is precisely this "crushing" of the researcher's ambition that will lead many of those in Post-Kantianism—beginning with Herder, extending through Goethe, Humboldt, Schelling, and on to Arthur Schopenhauer—to identify this core issue with Kant's Transcendental Idealism and work to furnish a more empirically satisfactory and conceptually rigorous account of materiality and mechanism that did not vitiate the budding biological sciences (Zammitto 2018).

account can such a product, as an *organised* and *self-organizing* being be called a natural [purpose]" (ibid.: 374).

With this, we can draw the threads together to better identify the difference between intrinsically organized and extrinsically constructed beings, that is, the difference between organisms *qua* natural purposes and artifacts *qua* products of a concept. We might understand *extrinsic* purposiveness as those systems whose teloi are imputed, endowed, manufactured, or created in accordance with the design, blueprint, aims, or desires of intentional beings—such as, for example, tool-using humans and non-human animals. These objects, systems, or artefacts are thus *derivatively* teleological: they are artefacts or machines whose function or purpose is determined by the actions or concepts of intentional beings: "we regard the effect [the artefact] as a product of art or [we] regard it only as a material for the art of other possible natural beings", which is called "usefulness" and is therefore "merely relative" (CJ §63). Intrinsic purposiveness implies a completely different regime of order underpinned by a distinct causal model and delineated along existential dimensions. These are systems that are "cause and effect of itself" (CJ, §64); that is, beings in which "everything is an end and reciprocally also a means as well. Nothing in it is in vain, purposeless, or to be ascribed to a blind mechanism of nature" (1793/1993: §65). When I write that they are delineated along existential dimensions, then, what I have in mind is that their *being is their doing* (as Kaufmann (1993) memorably puts it) and they are *self-maintaining*—acting in a denumerable set of ways in order to be the kind of system it is. In contrast to the artefact model on which extrinsic purposiveness depends, intrinsically purposive systems do not have their goals as something extrinsic to their parts and their organization. Quite literally, their goal-directed activity is *immanent* to their organization: "the organisation of biological systems is inherently teleological, which means that its own activity is, in a fundamental sense, first and foremost oriented towards an end" (Mossio & Bich 2017: 1090).

As Kant elaborates, and using the available scientific language of his time, extrinsically organised beings have *motive* power: "In a watch one part is the instrument for the motion of another, but one wheel is not the efficient cause for the production of the other: one part is certainly present for the sake of the other but not because of it. Hence the producing cause of the watch and its form is not contained in the nature (of this matter), but outside of it, in a being that can act in accordance with an idea of a whole that is possible through its causality" [CJ § 65: 374]. Kant gets further at the difference between the two when he writes that "one wheel in the watch does not produce the other, and even less does one watch produce another, using for this purpose other matter (organizing it) … which, by contrast, we can expect from organized nature. An organized being is thus not a mere machine, for that has only a *motive* power, while the organized being possesses in itself a *formative* power … which cannot be explained through [mechanism] alone" (ibid.). Not just their regime of orders, then, but their very modes of activity and possibility are different.

However, as we have already anticipated, right after presenting an elaborate analysis of the intrinsic-extrinsic difference, Kant proceeds to immediately problematise the ontological status of intrinsic purposiveness (that is, of judging natural products as natural purposes), remarking that this unique model of self-organization is not analogous "with any causality that we know" [CJ §65: 375]. There are myriad reasons for this, but to highlight two (interrelated) points: first, we have already seen that Kant has two permissible causal models in mind (to be elaborated further in the following section): that of mechanism, or efficient causality, that is, a series of descending parts in order of causal primacy and associated causal sequence; or coming into being in accordance with concepts. Why does Kant deny a third possibility, viz. the very one he introduced: that of a self-organizing being? Because, he contends, this would "contradict" the very essence of matter, which is supposed to lack the kind of 'inner animating principle' required for the cause to be superimposed with the being's materiality. This is what leads Kant to famously argue that intrinsic purposiveness is a *regulative* concept that applies to what he calls *natural description* [*natürliche Beschreibung*], the ontological status for which we cannot vouch. Teleological thinking (judging [*Beurtheilung*]) is "rightly drawn into our research into nature, at least problematically, but only in order to bring it under principles of observation and research in *analogy* with causality according to ends, without presuming thereby to *explain* it" [CJ § 61: 361]. It is already easy to see the affinity between this Kantian restriction and the many views canvassed in Chapter 1, and one way to better overcome this tension is to explore alternative views of matter and causal models that otherwise box our theorising and argumentation into narrowed positions whereby agency and teleology are eschewed in favour of a more streamlined ontology. With this in mind, I close this section with a brief overview of the notion of mechanism, materiality, and *efficient causality* on which the Kantian restriction hinges and, I submit, which still casts something of a shadow over contemporary ways of theorising agency in the life and mind sciences (cf. Emmeche et al. 2000).

2.2.2 Mechanism, Materiality, and Efficient Causality in the Kantian Antinomy

Kant's most detailed assessment of mechanism is worked out in the *Metaphysical Foundations of Natural Science* (1786/2011; hereafter, MFNS). In a section titled 'Construction of Matter', he explicates what we will call the *externalist view of matter*: one which is based on a mutual exclusiveness of parts and absence of interiority within their divisibility (in the philosophical jargon, this is sometimes called *partes extra partes* as it implies a mutual separability and lack of (immanent or inherent) relationality between bits and aggregates of matter; cf. Merleau-Ponty 1955; Barad 2007; Gallagher 2018). When I call this an externalist view of matter, I do so with the intention of capturing both the inertial understanding Kant has in mind as well as the affinity the view has with subsequent philosophical accounts of materiality which deny the properties of spontaneity and self-movement to brute matter. Movement, in so far as it occurs at all, is caused from a force *outside* matter. Stated differently, matter on this view lacks internal determination or immanent causal powers—and any aggregative activity of matter is determined by external forces of attraction and repulsion.

Accordingly, "All matter as such is lifeless" (MFNS: 544), a view to which Kant would remain committed for the duration of his career (as we already saw in the *Third Critique*, published four years after the MFNS, he writes that "Lifelessness, *inertia*, constitutes the essential character of matter" (CJ: 394)). A corollary of this inertial and externalist view of matter is that materiality also lacks spontaneity—something Kant associates with life and mind—and Kant's insistence on the *lifelessness* aspect of materiality will be key as it cleaves organisms from an inorganic context, erecting a gap between the nature of matter *as such* and living activity and mindedness. One goal of the MFNS is to establish the only admissible form of explanation in scientific practice, which is always mechanistic explanation, implying an ever-descending series of causes and progressively smaller parts determined in an inertial manner. For mechanistic explanations, "we say that an effect depends on a cause but not vice versa. As Kant puts it, in our purely mechanical explanations, we think of causal connection in terms of a *descending* series: "the things that are the effects, and that hence presuppose others as their causes, cannot themselves in turn be the cause of those others" (CJ § 65: 374)" Sedgwick 2012: 50). This is something called *efficient causality* [*nexus effectivus*] and is the dominant or assumed mode of causality in the sciences (Emmeche et al. 2000). This idea will be quite constitutive of later arguments (specifically in Chapter 3), so it is worth pausing and reflecting on what it implies.

Kant writes that efficient causation comprises a descending series of causes, as he proceeds to elaborate: "This causal nexus is called that of efficient causes (*nexus effectivus*). In contrast, however, a causal nexus can also be conceived in accordance with a concept of reason (of ends), which, if considered as a series, would carry with it descending as well as ascending dependency, in which the thing which is on the one hand designated as effect nevertheless deserves, in ascent, the name of a cause of the same thing of which it is the effect" (ibid.). What Kant identifies here, in so many words, is what we would nowadays call *upwards* and *downwards causation*, what Kant calls descending and ascending series respectively. Upwards causation is the more familiar species of causality from the perspective of scientific practice and maps onto the view of more fundamental causal activities preceding less fundamental ones, which usually has a spatial designation whereby mechanisms, parts, or components of a system are the real 'causes that kick'. This would have the lower-level (say, atoms, or molecules) as causes and the higher-levels (cells, organisms, behaviours) as effects—a depiction which has a tendency to leave the higher level as a "fairly impotent construction" that is "seriously threatened to be but an epiphenomenon of the lower one" (ibid.: 6). Thus, positing a kind of 'countervailing' causal interaction is seen as essential for establishing a non-epiphenomenal status for such higher-level features—both for Kant's account of *natural purposes* (intrinsic purposiveness) and contemporary ambitions of naturalising teleology.

If reciprocal causality is inscrutable for Kant, then efficient causality fairs no better when cognising or understanding the organisation of living beings. Writing in the case of the structuring of avian anatomy, we see that:

> [I]f one adduces, e.g., the structure of a bird, the hollowness of its bones, the placement of its wings for movement and of its tail for steering, etc., one says that given the mere *nexus effectivus* in nature, without the help of a special kind of causality, namely that of ends a this is all in the highest degree contingent: i.e., that nature, considered as a mere mechanism, could have formed itself in a thousand different ways without hitting precisely upon the unity in accordance with such a rule, and that it is therefore only outside the concept of nature, not within it, that one could have even the least ground a priori for hoping to find such a principle.

And thus, we are left with a relatively dissatisfying state of affairs: reciprocal causality is disallowed because it contravenes the dominant mode of scientific explanation, that of descending mechanistic explanation (what we might call the decomposition method of current neo-mechanism; see Machamer et al. 2000). But so too is efficient causality inadequate for capturing the seemingly purposive organisation of living organisms. And, for reasons we cannot get into here, invoking their design in accordance with a divine plan is even further beyond the pale for Kant. Hence why we are left with the regulative path whereby we treat living systems *as if* they are teleological while remaining committed to the mechanistic modes of explanations exemplifying scientific practice—a stance which has been called Kant's 'unstable middle position' (Gambarotto & Nahas 2022).

My reason for embarking on this tour of the Kantian formulation is twofold. For one, I do believe that the intrinsic-extrinsic purposiveness distinction as formulated in the language of 'natural purposes' (though I will stick with intrinsic purposiveness exclusively) is one of Kant's 'great services to philosophy', as Hegel put it, and sparked, indeed started, an important discussion about self-organization and how to figure this distinctive mode of organization within nature. Second, because I believe the germ of an answer to how this feature could be naturalised within a scientific framework is also present with Kant—although Kantianism is restricted from this move due to the understanding of matter and materiality on which it is based. The germ of an answer is, I submit, a situating of the activity and 'inner animating principle' within materiality itself.

To see how, we will return to the idea of a formative power that Kant thought picked out living systems as distinct from the extrinsic objects like figures in the sand. As Sedgwick describes, "In the case of [intrinsic purposiveness]… the effect of something is also the cause of that thing. Unlike artifacts, natural purposes are *self-organizing*; they possess within themselves a 'formative' force by which they are capable of maintaining and generating themselves" (2012: 50). What is critical is the following: that for intrinsically purposive beings, "their plan or purpose, rather than 'distinct from the matter of the thing' is somehow *immanent in natural processes themselves*", indeed, that because of this "we have to rely on something other than the model of efficient causation" (ibid.).[28]

[28] In an intriguing footnote, Kant illustrates what he means here by speaking of members of a polity being a means and a 'purpose' for the state: "and while each member contributes to making the whole possible, the idea of that whole should in turn determine the member's position and function" (CJ §65). Although I cannot give this more attention

The upshot of positing immanence within natural processes themselves is that it opens, I believe, a possibility for identifying a new mode of materiality: one that I will explore in depth in Chapters 3 & 4 on Condensed and Active Matter Physics. This is a position that understands materiality as containing a rich 'liveliness' (Barad 2007) and activity that would posit the ability for organized activity outside of the biological domain—rupturing the dividing line between the supposedly inscrutable and mysterious self-organizing processes of living beings and 'brute' or 'crude' matter. An intriguing upshot of the rupturing of the line between organized and unorganized modes of material organization is that it also establishes a gradualism between characteristically living and agential ways of being and more basic physical processes often seen as devoid of the directional or teleological structures that typify the living world. The rupturing here is indeed one of Schelling's main contention: that life represents an amplification of more rudimentary (though not lacking in complexity) physical and chemical precursors. This presents a felicitous affinity between Schelling's project of *Naturphilosophie* and the ambitions of current materials science programs that seek inspiration from the biological domain to engineer forms of complex spontaneity and self-organisation in physical sciences, medical, and engineering contexts (cf. esp. Tripaldi 2022).

For Kant, however, this 'rupturing' amounts to hylozoism: the idea that matter is somehow governed by an 'inner animating principle' and for him amounts to a *daring adventure of reason*, as it would eschew the distinction between constitutive principles (such as mechanistic explanation) and regulative principles ('as if' teleology). It is not my interest here to explore the fate of this distinction in a materiality that can accommodate certain components of 'hylozoism', but rather to sketch a position in which a living being's 'intrinsic purposiveness' is not "distinct from [its] parts" (CJ § 65: 373), but inheres within the activity of materiality as its ground. That is, to formulate a materials that evades the externalist view of matter just adumbrated and engages a philosophy of *immanence*. I appreciate that for now this position is highly abstract, but I believe that when telegraphed into the exciting work on Active Matter Physics, Materials Sciences, Basal Cognition, and Soft Robotics, we are progressively transitioning towards a view of materiality that captures, as Barad says, the "extraordinary

than in passing, I believe it resonates well with various views of Basal Cognition that depict multicellularity (as one form of collective action) as a 'confederation' of intelligences; or the 19th century cell state theory, which early on posited that cells maintain some degree of autonomy and a challenge of multicellularity was coordinating the activity of individual 'members' for coherent group-level goals. See Reynolds (2007) for a review.

liveliness of the world". This may amount to a 'daring adventure of reason', as Kant cautioned, but what is interesting is that his predecessors were actively interested in embarking on precisely this adventure—and it is to their exploration of an alternative view of nature I now turn.

2.3 *Naturphilosophie* and Romantic Biology

To reiterate, my interest in this chapter is delineating a form of naturalised teleology that comes in the shape of Kant's intrinsic purposiveness. I believe this is an illuminating line of research not only because of its pedigreed status in philosophical debates on normativity; but also because, à la biogenic approach, it picks out distinct kinds of systems that we can call agential and which represent steps on the way to more elaborate forms of cognitive sophistication found in the biological world. In order to have the latter, I posit that an agential dimension must be identified—at least for the biogenic approach—because it substantiates what we have called *existential needs*: the imperatives and demands of a living, out-of-equilibrium system to act in order to remain the kind of system it is. We have already seen that this implies an entirely different mode of causal organization, one that belies the ready and quick tendencies to draw analogies between organisms and artefacts (e.g., digital computers, run-of-the-mill machines, etc.).

To repeat, while I find the Kantian formulation fecund in overall shape, I believe the contours of the argument highlight the strictures and restrictions that create insuperable difficulties for naturalising teleology if we remain with Kant. In fact, one of the points I will now argue is that we have been *too* Kantian in how we approach the question of naturalised teleology. And I contend that the unmodified Kantian formulation simply does not have the resources to overcome the difficulties it sets for itself. That this is the case is evidenced within Kant's own corpus (van der Berg 2014), where Kant endlessly struggled to overcome the tensions within the antinomy of teleological judgment. Although Kant sought to posit a version of materiality that would overcome the externalist view– in the form of an 'ether' underpinning material activity – in the *Opus Postumuum* (1796/2011), this would still be one understood reductively or, as is sometimes called, *partes extra partes* (Gallagher 2018): because material becoming was still construed in an inertial manner in which self-organization could not be understood as inhering within matter. This is because Kant adopted Johann Friedrich

Blumenbach's own understanding of nature. Consider Blumenbach on the topic: "No one could be more totally convinced by something than I am on the mighty abyss that nature has fixed between the living and lifeless creation, between the organised and the unorganised creatures" (1789: 70). That is, Kant consistently maintained that it was impossible to determine the "original principle of organization" (Zammitto 2018).

Of course, self-organization is understood as a genuine mode of material becoming in contemporary science; however, there are two elements of this that fit quite naturally, and comfortably, within Kant's own strictures: first, the stable commitment to a hard and fast distinction between *organized* (living) and unorganized (non-living) beings. Second, the understanding that the principles of self-organization *can* be situated within a mode of reductive analysis that not only demystifies their material becoming but renders them amenable to the species of descending mechanistic explanation that typifies both Kant's attitude and the kind of 'scientific world view' that purportedly rejects the place of normativity, subjectivity, and agency in nature—at least in any ontologically privileged or causally efficacious manner.

For the remainder of this chapter, I will work to reject this image. If not reject, at least suggest that the ambitions of many working philosophers and researchers within Basal Cognition might find a more companionable ally in one of the many figures of *Naturphilosophie*—although my focus is almost exclusively on Schelling. While Kant's 'great service to philosophy' will indeed be an Ariadne's thread for my discussion, it is pivoting towards a Post-Kantian image of both our concept of *nature* and our understanding of the practice of science that I ultimately encourage, although I accept a fully comprehensive and convincing argument for this rests outside the domain of the current work.

Nevertheless, it is more of a Schellingian spirit that will animate the last few sections here. Consider Andrea Gambarotto on this point: "While it is true that Kant was the first to provide a philosophical formulation for the notion of teleology as internal purposiveness, it was Schelling who first conceptualised the developments of the late-eighteenth century German scientific tradition as a philosophy of nature that attempted to overcome Kant's regulative teleological judgment by defining purposiveness as an inherent, constitutive feature of living organisms" (Gambarotto 2018: 63).[29] Indeed, what is interesting to consider here is that the view of matter underpinning Kant's formulation was *already on the way out* in 18th century natural sciences, or at least being contested from within those scientific circles (Riskin 2016). This goes some way in dislodging the view that the *Naturphilosophen* were removed from, or otherwise anathema to, the sciences of their time. In fact, Schelling's work evidences a close engagement with the exciting developments of chemistry and electricity (galvanism) that vivified the scientific community of the time—and which represented points of departure with the overly restrictive view of matter found in Kant. Ironically, philosophers and naturalists more abreast of the many scientific developments bubbling to the surface in the 18th and 19th centuries were precisely those dismissed from the perspective of much 20th century biologists as "fantastical if not ludicrous" (Mayr 1983).

Picking up on this point, Dalia Nassar, in her magisterial overview of Romantic Science, writes that "While Kant drew a hard-and-fast distinction between 'organized' and 'unorganized' beings, arguing that non-organized being (e.g., crystals, snowflakes, and sand) are reducible to the mechanical laws of motion, Schelling came to the conclusion that nothing in nature is reducible to these laws. All beings, he contends, are fundamentally 'organized'" (Nassar 2022: 211). It is precisely this positing of organization as fundamental that allows Schelling to overcome certain elements of the antinomy of teleological judgment. Perhaps the most interesting argumentative move within Schelling's work on *Naturphilosophie* is the attempt to rework the mind-matter distinction (which here I take as homologous to my project of agency and materiality) from the opposite side than that of his contemporaries: that is, he sought to naturalise normative features of agents through the material side rather than substantiate an understanding of mind as autonomous and discontinuous from it. An important upshot of this is that there are no *leaps* in the natural progression from inorganic to organic nature, but a successive amplification or, what we might call following Michael Levin (2018), a *scaling up* of the inorganic into the organic.

This, I believe, has several important implications for the current work and my attempt at pulling from Basal Cognition and Active Matter Physics to effectuate the rethinking of matter I think is required for the ambitions of naturalising agency. I will focus on three for the

---

[29] Interesting because it would mean that the philosopher more abreast of the many scientific developments bubbling to the surface in the 18th and 19th centuries were precisely those dismissed from the perspective of much 20th century biologists as "fantastical if not ludicrous" (Mayr 1983).

moment using contemporary literature that, I believe, breathes new life into many of the ideas found in *Naturphilosophie*. First, there is a continuity and gradualism between complex forms of geochemistry and organic development—a position that finds new meaning in contemporary origins of life research, such as when Smith & Morowitz write that "metabolism is continuous with geochemistry" (2022: 25). Moving from simple chemical precursors of living dynamics to the modes relevant for agency under the biogenic approach represents a key move in later chapters.

Second, Schelling had a keen interest in what he called the *universal organism*: that is, the crisscrossing and mutual structuring of organic development with a wider geo- and biosphere that served both as *origin* of the former and as autonomous from it. In other words, this *universal organism* posits more generally terrestrial and chemical conditions that exist beyond the living domain while importantly being encapsulated by it—a point Nick Lane (2017) has also made at length. Again, pulling from Smith & Morowitz, we see that:

> The reactions characteristic of interface chemistry in the abiotic realm become the *constitutive* chemistry of the biosphere. This is another sense in which the ordering of processes pervades the nature of life in a way that it does not pervade the nature of non-living states. It has the consequence that the organization of living matter is anchored in the conditions for matter and energy exchange with its non-living context, at a finer level of detail than is true for physical phases of matter. (2022: 11)

While plugging naturally into the former point about gradualism, it also highlights a more interconnected and global view of life that does not take the autonomy of the organism as the end or primary goal, but rather as constitutive of wider and collective processes and interactions emblematic of life on Earth. Lastly, this plugs hand in glove with yet another project at home both in Schelling's work and the argument laid out in this thesis. What I want to promote in this work, as we explored in Chapter 1, is not organismality as an entirely novel and new mode of being whereby the constituents of a more complex multicellular organism *lose* their behavioural and cognitive plasticity (Levin 2018); rather, this represents a *confederacy of agents*, the goals of which is coordinate and collective intelligence to pursue ever more complex goals in space and time.

This, then, roughly represents the goals animating my current work and, I submit, contemporary research that seeks to understand how geochemistry folds into metabolism and cell biology: how cells, organisms, and ecosystems come to structural and interactively embody the energetic and material dynamics of a broader 'systems view of life'. Thus, to repeat, transforming our view of materiality to accommodate—in fact, to emphasise mind, agency, and subjectivity, has implications well beyond the philosophical considerations of naturalised agency, which include: a gradualism between the inorganic and the organic domain; a system's view of life whereby extra-organismic systems and interactions and relations acquire privileged ontological status; and a confederated view of life which posits, in a manner highly amenable to the ambitions of the biogenic approach, the interactive nature of cellular processes that are then entrained to pursue wider-organism level goals. The seeds for much of this, I now submit, are present within the dynamic and vibrant area of *Nautrphilosophie* and Romantic Biology—and it is the spirit of this historically underexplored and philosophically maligned program that I want to champion in this work.

This makes the present argument consistent with the growing literature of 'new naturalisms' in the philosophy of mind and nature—and a new naturalism which is distinctively *Post-Kantian*. This form of Post-Kantian naturalism departs from standard ways of theorising mind, agency, and nature as it connects the status of living organisms in scientific ontology with the place of mind in nature, something that would have been familiar to the *Naturphilosophen* but which tends to be lost in contemporary debates on naturalism where teleology is more narrowly construed as biological function as selected by the history of natural selection. Re-evaluating this older conception of teleology "is key for philosophical naturalism, insofar as it locates the wellspring of agency within nature itself" (2022: 2).

I believe this says as much about the historiography of Romantic Biology as it does a rekindling of a new vantage points and broad-strokes *ways of seeing* for contemporary issues in the life and mind sciences. Intriguingly, precisely because the organisms we examined in the opening chapter exhibit this deep entanglement between teleological structures and their minded capacities, a philosophical re-examining of intrinsic purposiveness as inhering within more basic physical processes, something I hope to establish in Chapters 3-6. In what follows, I provide a brief historical overview of the problems and programs associated with *Naturphilosophie* before proceeding to briefly overview Schelling's rethinking of matter, the outline of which will be a blueprint for the remainder of this work. That is, if Schelling argued that overcoming the difficulties of situating normativity and teleology in nature required a rethinking of materiality, then it is precisely a modern version of this rethinking that I hope to effectuate in this work.

2.3.1 Exploring a New Naturalism within *Naturphilosophie*

I pulled from contemporary work to highlight three aspects of *Naturphilosophie* that concern me in this work. The first was the continuity and gradualism extending from complex forms of geochemistry and chemical systems to biological ones, something Schelling calls the *Stufenfloge*, or graduated series of steps in nature (Berger & Whistler 2022). Second, the general interest in providing a larger organismal understanding of natural processes: taking organisms and their recursive causal structures as the basis for understanding natural processes writ large. Lastly, the confederated vision of organisms whereby a degree of autonomy is maintained in constituent parts. The degree to which any of these were admissible within the general ambit of 'proper science' (for Kant, mathematically formalizable frameworks amenable to mechanistic explanation) was something highly contested for Kant—and the joint project of naturalising teleology and situating mind in nature amounted, for Kant, "a futile and contradictory misadventure of physics into metaphysics" (Zammitto 2018: 320), that is, the 'daring adventure of reason' that closed the section on Kant.

The place and admissibility of this daring adventure of reason, then, sets the stage for *Naturphilosophie*, which sought to overcome the *as if* (and regulative) restrictions Kant placed around the biological sciences. In fact, we can see the *Naturphilosophen* as trying to establish a proper scientific—and naturalistically admissible—space for the blossoming life (and mind) sciences. Concerning the status of self-organization in organisms and nature, Schelling describes it as an "old folly to believe that organization and life cannot be explained through natural principles", calling it an 'unproven claim' which asserts "the first origin of organic nature cannot be investigated physically" [1798: iv]. Schelling thus proved himself to be a close reader of the scientific developments of his time (most notably, chemistry, which held an ambivalent position in Kant's understanding of proper science at best)—and what this amounted to was a need to reformulate the very notion of 'nature' on which Kant's philosophy (and his famous antinomy) rested. Speaking of the great German naturalist and close friend of Schelling, Johann Wolfgang von Goethe, Zammito remarks that "Goethe did not—*could not*—assimilate Kant wholesale into his 'nature'. The residual difference was decisive for his own thought and for the epoch" (2018: 290).

What I want to highlight is how *Naturphilosophie* represented a radical departure from the previous ways of formulating natural processes, such that a science of *self-organizing systems* became truly conceivable for the first time. And the decisive reformulations here were the rethinking of materiality and the re-examination of the living/non-living [organized and unorganized] distinction. Thus, contrary to Kant's own proscriptions on the daring adventure, this positively contributed to a fecund area of research championed by the likes of Goethe and Schelling, who "took from this 'adventure of reason' the warrant for drawing from Kant a program of developmental transcendental morphology and even a form of evolutionism" (Sloan 2006: 343). Or, as Le Ann Hasen Le Roy picks up, there was pressing need for this at the time because it was widely agreed "Kant had made a philosophy of nature impossible" (1985).

In light of this, Schelling's *Nautrphilosophie* presented what some researchers thought was a panacea to the many tensions, antinomies, and conflicts that were native to Kantian soil and the above contextual details provide insight into why. Carl Friedrich Kielmeyer captures this well when he remarks that Schelling's *Naturphilosophie* was not only original, but responded to real problems within Kant's formulations with which theoretically-oriented researchers struggled. This is what we saw above with Schelling's insistence that Kant's prohibitions served little more "than to crush the ambition of the researcher", and, as Zammito notes, "This was the decisive point that linked his program in *Naturphilosophie* to the ambitions of emergent life science. And there is no question that the target of Schelling's criticism was Kant" (2018: 325).

Crucially, we see that Schelling "saw himself undertaking a philosophical rescue operation for a natural science in epistemological crisis, on the one hand, and on the cusp of theoretical breakthroughs, on the other" (ibid.). If Kant's own formulations were placing bounds on the very possibility of a philosophy of nature—let alone a philosophical biology and theoretical morphology more generally, as Goethe and Schelling saw—then *Naturphilosophie* sought to move well beyond them. Zammito summarises the adoption of *Naturphilosophie* as follows:

> *Emergence* and *process* became central to the idea of nature in itself. It became inherently creative. Self-production in nature moved from the simpler to the more complex; it took on historical-developmental form. Accordingly, science needed to shift its attention from

> a set of determinate *products* to the immanent *processes* that generated them—in philosophical terms, from *natura naturata* to *natura naturans*. (2018: 326)[30]

The distinction here, that between *natura naturata* and *natura naturans*, is a philosophical one first introduced by Spinoza—a key figure in these debates, although I will not be able to discuss his work in further detail.[31] Briefly, *natura naturata* is the view of nature in terms of its *products*, or nature as *determinate product*: the resultants of a process, the sum total of all things. *Natura naturans* refers to the processes themselves, the immanent productivity of nature, the process by which it progressively individuates into products, and the inherent creativity of material activities. By making his philosophy of nature the philosophy of this productivity, Schelling displaces causal powers from an externalist conception of materiality envisioned in Kant's materialism to an inherent and immanent process autochthonous to nature.[32] As Frederick Beiser writes, for the *Naturphilosophen* "nature is nothing less than living activity or productivity itself" (2002: 530).[33] The vision of naturalism promoted here is directly amenable to the argument laid out in the following chapters. If the Kantian vision of 'nature' is one in which self-organization—and notions of emergence and reciprocal causality on which it rests—is inscrutable and mysterious, then one aim of my present analysis has been to show how a different view of nature opens up a space where such concepts can be more amenable to the philosophical ambitions of researchers and philosophers working within the biogenic approach. Again, pulling from Beiser we see that:

> The central principle of Schelling's new dynamics is that nature in itself consists in infinite productivity, absolute activity. Schelling warns against considering nature in itself as a 'substance', or even as a 'totality of things', for these terms imply a static view of nature [natura naturata]. He also advises against supposing that activity somehow inheres in some substance or thing. A substance or thing is never as basic as activity because it is only the result or product of it. Strictly speaking, then, nature is not even an organism, since that would be again only the product of its activity (*natura naturata*). Instead, nature is nothing less than living activity or productivity itself (*natura naturans*). (2002: 530).

---

[30] *Natura naturata* and *natura naturans* is a philosophical distinction first introduced by Spinoza to identify nature as *product* and nature as *productivity* (or process) respectively. It is common to understand this designation as *natura naturata* referring to the scientific practice of identifying products of nature and *natura naturans* identifying the creatvitiy and productivity of material/natural processes.

[31] See Grant (2006) for an analysis of how Spinoza's philosophy set the stage for the program of *Naturphilosophie*.

[32] That is, organization and mind were not something imputed externally from subjectively confined observes 'locked in their heads', but saw the germ for organization—relationality and emergence—as more general in nature.

[33] I do not wish to resuccitate this living aspect per se, but I do agree with Karan Barad that the physical sciences reveal a much, much more exciting and lively world than the one envisioned in an arch-reductionist like Rosenberg.

When I discuss natural sources of agency, it is something in this sense that I intend. What this involves is a reconsideration of the *source* of causal powers in nature: not something external to it, determined by position in space or inertial forces external to it. This is where Schelling's construction of matter has the upper hand on Kant's: in incorporating insights from the chemical sciences, it became clear, even in light of Kant's own time of writing, that matter could not be the homogenous mass it was presented as being in the *Metaphysical Foundations of Natural Science*.

This is a point Dalia Nassar captures well when she writes, "mechanical philosophy could not make sense of the newly discovered chemical phenomena, which implied that the structure of matter was *not* qualitatively uniform" (2020: 214). That is, chemical phenomena "reveal that change in matter is not purely external, but also internal—chemical change has nothing to do with the place of matter in space or spatial impact" (ibid.). While Kant, ever abreast of scientific developments, sought to incorporate forces internal to matter, his commitments to mechanical philosophy ultimately limited his dynamical construction of matter from doing justice to chemical—and, by extension, organic—phenomenon. Even in his later work, Kant never reneged the construction of matter and its incompatibility with purposive structures: "The system of dynamic forces of the [*Opus postumum*] remain entirely in the inorganic realm, and so can have a constitutive status without violating Kant's strictures on teleology" (Beiser 2002: 189). That is, "matter is dynamic but not organic, that it consists of moving but not living forces. Hence he argues repeatedly that matter is not living because it does not have the power to move itself, and because the idea of a purpose involves some immaterial principle" (ibid.). Thus, despite Kant's attempt at rendering his account dynamic, it is still constrained by the externalist view of matter canvased in the previous section.

I reiterate this point because it contains a key point of difference from Schelling's own view on matter: an organicist account that tries to express the *natura naturans* immanent to organic processes. What this comes down on is the reformulation—or an alternative approach—to the construction of matter that structures much of Kant's later critical works.[34] And what this amounts to is no less than an inversion of a Kantian divide consisting of *our* concept of 'Nature'

[34] Given the centrality of the concept of the organism for Kant's later work and Post-Kantian philosophers, it is surprising that Kant's *Teleological Judgment* is never once mentioned in Sellars work on the scientific and manifest images—emblematic of a general neglect in Angloamerican philosophy regarding this foundational question.

as emanating from a distinctly human point of view. Schelling thought—and I concur—that this was to begin the inquiry from the wrong way around, and did not consider nature's productivity independent of the self-constituting consciousness of the human intellect. It is in this regard that I think Schelling's *Naturphilosophie* furnishes the conceptual resources required for a more 'radical' rethinking. In other words, Schelling's project was to account for the purposive structures of organic activities—and, ultimately, human subjectivity—in naturalistic yet nonmechanistic terms. Beiser, remarking on this project, helpfully summarises it as follows:

> The strategy behind Schelling's *Naturphilosophie* was to approach the classical problem of mental-physical interaction from the opposite direction of [Kantian] philosophy itself. Rather than beginning from the subject and investigating the realm of consciousness, Schelling would begin from the object and study the nature of matter itself. He recognized that the whole problem of mental-physical interaction involves the question 'what is matter?' as much as the question 'What is mind?' Schelling contended that we should not view the nature of matter as a given, as if the only mystery were the mind and its relation to matter. Rather, we have to recognize that the very nature of matter is mysterious—it is indeed "the most obscure of all things, and indeed to some obscurity itself". (Beiser 2002: 511)

This, more than anything else, encapsulates a key motivation of *Naturphilosophie*—and there is little better expression of the following argument for what I believe is needed for debates on naturalised agency and intrinsic purposiveness than this.

2.4 Conclusion

To conclude this chapter I pull from Gilbert Simondon, a French philosopher of science whose work on biological and psychological individuality was similarly influenced by *Naturphilosophie*. In words that could easily represent a philosophical rallying call for the remainder of this thesis, Simondon identifies an energetic and proto-thermodynamic[35] conception of individuating processes and the tendency towards organization that begins from simple physical processes (such as crystallisation) and proceeds through to higher-level versions of individuation encapsulated in living self-organizing processes and, eventually,

[35] The use of far-from-equilibrium dynamics as the main physicalist basis for this form of individuality was still being specified and awaited the future work of Ilya Prigogine; see Prigogine & Stengers (2020).

dynamic human societies—resonating with Schelling's idea of the *Stufenfloge* discussed briefly above. Simondon writes, in a passage worth quoting in full, "This kind of research supposes that we consider legitimate the usage in biology of a paradigm taken from the domain of the physical sciences and particularly from the processes of morphogenesis that occur within this domain" (2020: 169). He continues,

> In this sense, it is necessary to suppose that the elementary levels of the biological order contain an organization of the same order as the one that the most perfectly individuated physical systems contain, for example, those that generate crystals or the large metastable molecules of organic chemistry. Indeed, such a research hypothesis can seem quite overwhelming; custom in fact prompts us to think that living beings cannot result from physical beings since they are superior to the latter due to their organization. (ibid.).

Critically, Simondon identifies the error present in this thinking that I have tried to express over the past few chapters: namely, how "this very attitude is a consequence of an initial postulate according to which inert nature cannot contain a high level of organization. If, on the contrary, we posited right from the beginning that the physical world is already highly organized, we wouldn't be capable of committing this basic error that results from a *devaluation of an inert matter*. … [If] from the start it is estimated that matter constitutes systems provided with a very high level of organization, then we cannot so easily hierarchize life and matter" (ibid.). Whereas Schelling could not have had the resources at the time to incorporate this energetics perspective (though his use of chemistry represents an incipient form of this), Simondon explicitly makes reference to the thermodynamic situatedness of self-individuating systems.

Ultimately, one thing I have wanted to highlight over the past two chapters is precisely this 'energetics' approach: one that captures the energetic-cum-material processes that create and constrain order in the living and non-living domains. While it may seem obvious, the importance of these processes is often not appreciated. Before concluding, then, I would like to canvas briefly the relevance of this 'rethinking matter' for contemporary debates in the life and mind sciences. Again, this is not a 'materialist chauvinism' that tries to prioritise this or that species or type of matter—to do so would be to revive a kind of 'substantialism' in the debate. Rather, what I have in mind is something closer to Schelling calls a 'materiality that is not yet corporeality', that is, it concerns the processes that can generate bodies or self-organizing systems.

Kirschner & Gerhart provide an early (2000) articulation of how the devaluation of matter, energy, and materiality has cast a penumbra over theoretical and evolutionary biology—and promote what they call, in a slightly light-hearted way, a 'molecular vitalism' that foregrounds biological organization and the potential chemical principles underpinning it (in contrast to a gene-centred view that would depict genes as the primary orchestrators of organic development). They comment that in the 19th century, when the life sciences were first congealing as disciplines (as we surveyed briefly above), hereditary was seen as an important dimension of organic life *alongside* metabolism and bio-chemistry—but the latter features were largely marginalised in mainstream theoretical biology of the 20th century, much to the detriment of our understanding of organisms. Kirschner & Gerhart propose that we now (writing at the turn of the 21st century) that we turn "to an investigation of the 'vitalistic' properties of molecular, cellular, and organismal function", which they propose we can understand in terms of chemistry.

Writing more recently, biochemist Nick Lane also proposes an inspirational return to the 'formative years of biology', where the guiding question was "what processes animate cells and set them apart from inanimate matter?" and proposes that we make reference to bioenergetic research and principles to better understand "the deep chemical coherence of the living world" (2022: 6). Lastly, Laura Tripaldi, whose work will be increasingly central for the next chapter, recapitulates the tension when she writes: "The customary terminology we have used to distinguish the living world from the non-living as always been problematic. Non-living materials have been defined as 'inorganic', 'inert', 'inanimate', or 'disorganized', but each of these terms has ultimately proved inadequate as a way to specify the precise boundaries where life begins. … Chemical matter, whether organic or inorganic, is active and dynamic, capable of forming complex organisations on different scales, evolving, and spontaneously modifying its structure in response to the environment. *For this reason, rather than considering life as an extraordinary phenomenon, miraculous and extraneous to all other behaviours of matter, chemistry provides us with the tools to think of life as one of the many different types of dynamic organisations that matter can assume*" (2022: 110).

There is thus a close affinity between the Schellingian approach detailed above and the more recent 21st century emphasis on reapproaching the bio-energetic and material aspects of living processes—the assessment of which will occupy us in part 2 of this work. Beginning with the formula that Schelling sets out, we will look at simpler Active Matter systems, such as the chemotactic oil droplet (Hanczyc & Ikegami 2010), before proceeding to the self-organizing processes examined in the biological autonomy literature. This will furnish the basis for the view of cognition that accounts for the 'cognition all the way down' perspective defended throughout this thesis. The competencies and agential modality typically associated with, and restricted to, the higher-level organism will be seen as crucially dependent on the agentic—sometimes cognitive—capacity of constitutive cellular capacities. The discussion to follow, then, will begin from the ground up as it were, for, as Richard Feynman noted, there is plenty of room at the bottom.

Chapter 3

'Just Physics'? Rethinking Emergence and Causation in Physics and Developmental Biology

3.1 Expanding the Physical within Physicalism[36]

The previous chapters explored the importance of cultivating a metatheoretical philosophy of nature that targets a wider nexus of ideas, assumptions, and concepts that structure much of the debate on naturalised agency and teleology in the life and mind sciences—with an eye towards reframing and recontextualising the nature of mechanism, physicality, and materiality as an inroad to naturalised forms of agency in the biological world. This chapter, along with the following chapter, concretise the discussion further by turning towards recent trends in Condensed and Active Matter Physics [CAMP] as well as Materials Sciences, which furnish scientifically defeasible and conceptually robust accounts of top-down causation and emergence. Causation and emergence (and their combination in the form of 'strong emergence') have been seen as constitutive in the naturalising teleology debate arguably since the idea of recursive (or reciprocal) causality was first advanced in the *Teleological Judgment* [Chapter 2: Origins of Organization]. This means that the establishment of a scientifically substantial and naturalistically admissible form of strong emergence—or its elimination—are taken as tantamount to the legitimation (or otherwise) of agential ascriptions to organic processes. As Potter & Mitchell write, "to justify agent causation… systems need to exhibit a causal power that *irreducibly* inheres at the level of the whole system (i.e., it is nonreductive) while still maintaining that the causal power is instantiated in, or realised by, the system's physical constituents (i.e., it obeys physicalism)" (2021: 1). This chapter makes the first stab at understanding what is involved in the emergence and top-down causation debates—setting the stage for how they eventuate in an account of agency relevant to the biogenic approach (Chapters 4 & 5)

[36] A note on chapter structure: this chapter was initially written in combination with the following on Active Matter Physics. For argumentative (not to mention length) purposes, I have decided to divide the two into separate pieces. However, the strategy pursued herein might still reflect their prior integration. To clarify, then, I take my exploration of both as necessary for the general rethinking of matter required for naturalised agency. This chapter focuses most explicitly on structural and material features; whereas the following will situate those properties into their bioenergetic context. References to active materiality and energetics in this chapter will be more in passing and provisional, awaiting a further elaboration in Chapter 5.

This is a chapter about reductionism and physicalism. It is about how overcoming the former and reworking the latter creates a conceptual space in which naturalised forms of agency can be further explored. More perspicuously, it is about cultivating a better image of the myriad disciplines and subdisciplines that fall under the umbrella of 'Physical Sciences'. Often times, physics is presented as somewhat of a bogeyman for the life and mind sciences—and the tacit background reference here is a kind of 'fundamental physics' at which the basic causal powers and properties inhere. The 'just physics' concern (explored in Chapter 1), then, turns on an understanding of the physical sciences in which the dominance of fundamental physics is taken as uncontroversial. What is rarely appreciated here is that the sheer range of domains, physical scales, objects, and phenomena that physics deals with—covering the fermions and bosons of quantum field theory, to the behaviour of metals and engineering of new materials in condensed matter physics, through to the far-from-equilibrium and intrinsically active systems explored in active matter physics.[37] One of the key arguments of the next two chapters is to show how expanding the 'physical' within physicalism to include condensed matter physics and materials sciences (this chapter) and active matter physics (Chapter 5) correspondingly problematises rejoinders in philosophy of mind and biology that depict the causal regimes emblematic of biological organization as 'just physics'. Stated differently, it might be 'just physics', but it being so does not vitiate the ascription of agency that I seek to make here. Ascribing agency to basal and minimal systems—the very systems under investigation in the biogenic approach—mandates a consideration of how agential features relate to their physical substrate – in a manner that does not reductively imply an elimination of the former to the latter.

The argument put forth in this chapter builds off recent work in philosophy of science that rehabilitates some explanatorily acceptable notion of top-down causation and emergence that evades the reductionist worries normally plaguing the debate (e.g., Jerry Fodor's 'just physics' concern; Chapter 1). Specifically, I build on philosophers of science Sara Green & Robert Batterman (working in philosophy of physics and materials sciences) who have challenged the assumptions that physics promotes the reductionist ideal that it is commonly ascribed: "An important question that is rarely addressed is whether the ideal of progressive reduction of higher-level explanations is supported in physics, i.e., in the discipline that was taken as a model for the reductionist ideal" (2017: 19). More pointedly, they argue that "rather than

[37] One estimate has it that one-third of working physicists are employed in fields pertaining to condensed matter physics. See: "Condensed Matter Physics Jobs: Careers in Condensed Matter Physics". *Physics Today Jobs*. Archived from the original on 2009-03-27.

enforcing reductionism, physical sciences approaches can help reveal the limitations for it in [for example] developmental biology" (ibid.).

The crux of their argument comes down on two key notions: the *tyranny of scales*:--the idea that materials and physics operate different and are subject to different constraints depending on the spatial and temporal scales (e.g., nano-, meso-, and macroscales, with many systems spanning all three) at which they occur—and the notion of *constraints*—physical boundary conditions that constrain the possible configurations of lower-level physical states in a way that is dependent on scaffolding from higher-level material features (the extracellular matrix and its implication in carcinogenesis will be a key example going forward). Both concepts address addresses the scale-dependency of physical behaviours (ibid.). In looking at the tyranny of scales, we begin to see that the very reductionist ideal that worries (philosophers of) biology in incorporating physical sciences techniques, approaches, and concepts within their remit does not actually hold water in its supposedly native context:

> [In biology] physics has often been pictured as a discipline solely targeting the most fundamental 'lower levels', with a preference for simple deterministic models. Interestingly, appeals to physical science approaches in the examined cases of multi-scale modelling in developmental biology [see below] do not support this view. Rather, they show that macroscale features (i.e., those at cell and tissue level) are indispensable and irreducible to lower-level explanations. Moreover, we propose that the requirement of macroscale parameters as boundary conditions for models at lower scales provides a concrete instantiation of top-down effects. We highlight how recent insights to biomechanical aspects of morphogenesis challenge deeply entrenched presuppositions about the explanatory priority of lower scales. (Green & Batterman 2017: 22).

It is precisely in their targeting of these presuppositions—constitutive themselves of philosophical physicalism—that makes Green & Batterman's analysis particularly relevant to the present discussion. It also aids the argument laid out in the previous chapters where I suggest that philosophical physicalism and materialism are based on an *outmoded* understanding of the physical sciences, such that the reductionism and anti-emergentism it is typically seen as promoting do not follow from what the physical sciences show. And here, the subdisciplines of Condensed and Active Matter Physics become increasingly important as they inherently deal with emergent structures and often addresses out-of-equilibrium systems capable of driving and maintaining their own dynamics. In the case of Active Matter Physics (which is a subdiscipline of condensed matter physics) physicists often deal with the very same entities as biologists: organisms and the dynamics that sustain them and the materials out of which they are constituted. Thus, what I want to convey in this chapter is that revisiting the physical sciences in a way that does not privilege (but no less neglects) theoretical physics furnishes a novel ontology that tries to accommodate and capture, rather than eliminate, the nature of recursive causality and some (yet-to-be-specified) notion of strong emergence that are integral to naturalized approaches to agency.

A final note on how this chapter will relate to the coming discussion of naturalised agency relevant to the biogenic approach and research in Basal Cognition. Many philosophers view the notion of top-down causation as logically incoherent, ontologically impossible, or at least extremely problematic—usually coming with great metaphysical purchase. This is often due to the causal fundamentalism associated with 'vertical reductionism'. The idea of vertical reductionism is that for every macroscale description of a phenomenon, there is a microscale description that suitably fixes it (Potter & Mitchell 2021: 2). We saw this briefly in the first chapter, but to elaborate, this is the idea that any description of phenomena at the macroscale can be placed by a description at a lower-level: "The logical endpoint of this vertical reductionism is that everything is describable in terms of the smallest possible physical elements in the universe… and the interactions between them" (ibid.)—connecting to the contention that physicalism essentially refers to 'fundamental physics'. Importantly, Potter & Mitchell continue, "if all causation is fixed by microphysical happenings, then agents must be epiphenomenal. There is simply no room in the universe for macroscale phenomena to have the kind of 'irreducible' causal power needed for agent causation" (ibid.). I submit, in contrast to this entrenched reductionist assumption, that the modern sciences of condensed matter physics undermine the wildly shared belief that a complete physics might ultimately efface the type of emergence required for agency to be a more than descriptive gloss painted on an otherwise 'mechanical' world. We will earmark this discussion for now and return to it over the following two chapters.

The structure of this chapter will be as follows. Section (3.2) briefly surveys the notion of emergence—in both weak and strong iterations—and the account of top-down causation that will be defended in this and remaining chapters. Then, in section (3.3) I read the philosophical literature on emergence (3.2) through the contemporary empirical work in condensed matter physics to argue that a more expansive account of the 'physical' within philosophical physicalism is needed to accord with what the diverse branches of the physical sciences shows about the world. Section (3.4) will utilise the central notion of constraints to make a case for a scientifically informed account of top-down causation. I conclude in section (3.5) with a few observations about the ramifications of condensed matter physics for biology and the cognitive sciences, setting the stage for the discussion of active matter physics to come in the following chapter.

3.2 Weak and Strong Emergence—Upwards and Downwards Causation

The idea of emergence is, very simply, that "the emergent is unlike its components insofar as these are incommensurable, and it cannot be reduced to their sum or their difference" (Lewes 1879: 412).[38] More colloquially stated, emergence is the idea that the whole is greater than the sum of its parts. Emergence abounds in nature and although it is more controversially discussed in the domain of intentional, sentential, and mental phenomena, it would be a mistake to consider this the only domain dealing with emergent features. The difference amongst emergent phenomena often comes down on the weaker and stronger iterations of the thesis—with only strong emergence implying a form of downward causation often seen as metaphysically incompatible with physicalism and naturalism. This section will proceed from weak to strong emergence before turning towards the causal models on which they depend—efficient or upward causation for the former and reciprocal or downward causation for the latter.

Weak and strong emergence can be fleshed out along epistemological (dealing with modelling and predictability) and ontological (causal) dimensions (Chalmers 1995, 2006; Humphreys 2019; Turkheimer et al. 2019; Luppi et al. 2021). Beginning with the epistemic notion, Chalmers (2006) describes weak emergence as follows: "We can say that a higher-level phenomena is weakly emergent with respect to low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are unexpected given the principles governing the low-level domain. … It often happens that a high-level phenomenon is unexpected given principles of a low-level domain but is nevertheless deducible in principle from truths concerning that domain". Thus, epistemologically, something is said to be weakly emergent if the resulting structure or effect is *predictable* based solely on knowledge of lower-level realisers. While a resultant behaviour of a system might be different from the interactions between various layers and constituents,

[38] G.H. Lewes is largely credited with coining the nominal term 'emergent' and thus catalysing a moment in British philosophy properly called emergntist. Interestingly, it is largely the British emergentists that Jaegwon Kim had in mind when identifying the philosophical troubles with this school of thought. While Kim's analysis were indeed penetratingly accurate at the time, a central contention of this thesis is that the conception of emergence he argues against is not one an emergentist needs to endorse. The implications of condensed matter physics for emergentism in philosophy will be explored in the conclusion of this chapter (section 4.5).

it is nonetheless explained by, or deducible from, the interactions of 'elemental' units. Strong emergence (sticking with its epistemic specification) is when the behaviour of a system appears in a highly non-linear and unpredictable way (Turkheimer et al. 2019: 3). This means that the higher-level behaviour *cannot* be predicted based on (ideally complete) knowledge of the constitutive processes.

Something much stronger is captured when fleshing out emergence along an ontological axis—getting at issues of holism and reductionism discussed in the previous chapters. To say something is weakly emergent in this ontic way is to say that the emergent property or feature exhibits qualitative differences from lower-level realisers[39] (e.g., 'hardness' is not a property found at lower spatial scales) but is still wholly *realised* or *determined by* its constituent processes. The same could not be said for strongly emergent phenomena, features, or behaviours. Exemplary of this latter class would be organisms, agency, and consciousness. Here, we allegedly deal in a different causal model that implies a *recursive* or *top-down causation*: the emergent emerges from its constituent processes and in turn feeds back, influences, and *constrains* the possible configurations of those processes at a further timestep. George Ellis describes this form of strong emergence as follows:

> Very complex systems such as those occurring in biology [require] top-down causation, needed in order to build up the necessary biological information. This information cannot be derived in a bottom-up way, because it implicitly embodies information about environmental niches. It would be different in a different environment. Hence, higher level conditions influence what happens at the lower levels, even if lower levels do the work. This is what I characterise as top-down causation. (Ellis 2012: 127)

What Ellis tries to highlight with higher-level conditions influencing what happens at lower-levels has recently been expressed using the language of *constraining conditions* or, more generally, *constraints* (Bechtel & Bich 2021; Winning & Bechtel 2020; Winning 2021; Green & Batterman 2017; Batterman 2022; Green 2022)—and it is the language of constraints that will do the most philosophical heavy lifting later in this chapter. For now, constraints can be characterised as conditions that limit and enable certain behaviours, influencing the activities of entities and processes at multiple levels to generate novel behaviours. Constraints do not just *limit* the physical possibilities of a system; they *make possible* certain trajectories,

[39] This language of 'realisers' and 'realisation' will become increasingly important in the penultimate chapter of this thesis on Multiple Realizability, Soft Robotics, and Artificial Intelligence (based on work put forth in Harrison et al. 2021 and Harrison (forthcoming)).

configurations, and behaviours at the same time (Deacon 2011; Bechtel & Bich 2021). Undoubtedly, it is this form of strong emergence that has seen the most spilt ink and acrimonious disagreements among researchers—and it is some suitably specified version of it that is required for naturalising a kind of agent causation. To see why, let us revisit the model of causation often associated with mechanistic activities and philosophical physicalism—the *efficient causation* explored in Chapter 2 [‘Origins of Organization’].

Recall from the discussion of Kant that the causal model most commonly invoked as it relates to mechanism and mechanistic interactions is that of causes having effects in a largely linear or deterministic manner. This is the principle of efficient causality and is often the dominant and assumed mode of causal interactions found in philosophy and certain domains of science (see Emmeche et al. 2000: 6). Although it need not strongly imply upward causation, efficient causality is often invoked to suggest that putative causal powers at a higher-level can be explained via reference solely to lower-level realisers. We could even say, following Carl Craver, that efficient causality implies an ‘austere and demanding’ mechanical philosophy that depicts a “disenchanted world explicable without remainder in terms of basic causal principles” (2013: 1). Pulling as we did from Emmeche and colleagues influential discussion on causation and emergence (Chapter 2), we see that the type of upward causation assumed in this formulation of efficient causality is one in which the higher level entity or process emerges from a lower causal process—with causes leading to effects as they flow from lower to higher levels (2000: 5). This would have the lower-level (say, atoms, or molecules) as causes and the higher-levels (cells, organisms, behaviours) as effects—a depiction which has a tendency to leave the higher level as a “fairly impotent construction” that is “seriously threatened to be but an epiphenomenon of the lower one” (ibid.: 6). Thus, positing a kind of ‘countervailing’ causal interaction is seen as essential for establishing a non-epiphenomenal status for such higher-level features.

Downward causation, then, implies a reciprocal causality whereby not only the existence of higher-level features but their activities and interactions can feed back, influence, or constrain lower-level happenings (as we saw briefly above). The idea here is that once constituted, a higher-level feature is equipped with a causal prominence such that it is able to co-determine the directionality of constituent processes. To concretise this, consider the recursive process whereby semi-permeable membranes constrain the flow and direction of energy that is then harnessed to drive metabolism and thereby continue to the perpetuation of the cellular membrane; or the manner in which contextual (environmental, nutritional, behavioural) factors can mediate gene expression (Lamb & Jablonka 2011). Again, as we saw in Chapter 2, this form of downward causation can involve something stronger and more implausible (a kind of efficient causation working in the opposite direction), but as it will be used here it will take the form predominantly of a more minimal thesis involving the notion of boundary conditions and constraints. While the focus of the next section, I spare a few words now to explicitly link both concepts together.

Of the three kinds of downward causation explored in Emmeche et al. (2000), the one that I will endorse and explore in the context of condensed matter physics and materials sciences is *medium downward causation*. This can be characterised as follows:

> The emergence of a higher-level entity, process, or feature occurs when one of the many possible states amongst lower-level interactions is actualised. It implies downward causation when the information present at the higher-level at a previous timestep acts as a factor of its actualisation.

This idea can be better specified with the dual ideas of 'boundary condition' and 'constraint'.[40] We have already spared a few words on constraints, but boundary conditions are a concept employed in physics to identify the delimitation of several possible states of a system's development. In classical mechanics, a system's initial conditions are defined as a set of parameters that describe the starting point of a system (an arrow flying through the air, say) and which form the basis of an unlimited number of predictions on its trajectory through time. These systems are also, in some important sense, *atemporal* in that their behaviour can be run forward and backwards (Emmeche et al. 2000: 11). When dealing with complex systems, however, further postulations are required for understanding a system's development. Complex systems can be very sensitive to initial conditions such that changes in initial properties can change subsequent dynamics: "These are named boundary conditions because they delimit the set of initial conditions within which the properties in question will be found" (ibid.).

Note how, stated as such, neither boundary conditions nor constraints imply downward causation. To make that extra step, boundary conditions need to be linked with the aforementioned concept of constraints, specifically *constraining conditions*: "boundary

[40] Note how the notion of constraint need not imply *downward* causation but a kind of 'intralevel' or 'lateral' causation. The hill is a constraint on the ball's behaviour as it rolls downhill, but this doesn't imply any lower-level causality. However, the interactions of the ball-incline system are of a fundamentally different kind than the kind of constraint-based activity we find in many biological systems, as we will see in section 4.4.

conditions are conceived as the conditions which select and delimit various types of the system's several possible developments. The realisation of the system implies that one of these typical developments is selected, and the set of initial conditions yielding the type of possibility chosen are thus a certain type of boundary condition which has been called *constraining conditions*… [they] are the conditions by which entities on a high level constrain the activity of the lower [level]" (ibid.). Phrased as such, downward causation can be further defined as the process by which higher level entities, processes, or behaviours operate as *constraining conditions* for the emergent activity of lower-levels. The following sections, then, will work to better concretise the (so far quite abstract) discussion. I now turn to the modern sciences of Condensed Matter Physics and Materials Sciences to highlight some of the material and structural features that can be said to exhibit this mode of causality.

3.3 More is Different? Condensed Matter Physics and the Sciences of Emergence

Although condensed and active matter physics hold clear relevance for philosophers of mind, science, and biology—not to mention working biologists interested in basal cognition and developmental biology (as we will see over this and the following chapter)—there has been notably little work trying to generalise or extend recent advancements in the physical sciences to these areas (for exceptions see Hanczyc & Ikegami 2010; McGivern 2020). Philosophers of science Sara Green and Robert Batterman (2017, 2018, 2020; cf. Green 2018, 2021; Batterman 2021) represent notable exceptions. Over a series of papers, Green & Batterman explore the sciences of condensed matter physics and materials sciences both in their relevance for biological practice and modelling and in exploring the philosophical implications of this work, specifically by revisiting the physicalist-reductionist ideal that is often seen as native to the physical sciences. Given their concerted engagement with this body of literature, it is their treatment of condensed and active matter physics that will occupy me for the remainder of this section.

3.3.1 Examining Reductionist Assumptions in Physics and Physicalism

Condensed matter physics [CMP] is a branch of physics dealing with matter in its condensed forms, namely solids, liquids, and the phase transitions that occur between them. It explores the behaviour of large ensembles of interacting particles, such as atoms, molecules, electrons, and spins to better understand the nature of collective behaviour and phase transitions.

Common phenomena explored in CMP include superconductivity, semiconductors, protein folding, viscoelasticity in materials sciences (itself within the remit of CMP), and the active harnessing of energy to drive complex behaviour. Due to its handling of inherently collective dynamics, CMP is evidently critical for our understanding of emergence and emergent phenomena (Wusser 2022).[41] In this respect, it is interesting to note that many of the key puzzles for theoretical biology are shared in CMP: how simplicity can arise from complexity; how form can be generated without outside engineering; how physical processes spanning multiple spatial-scales and operating over different time-lengths can nonetheless exhibit coordinated behaviour; how (and this is the emergentist question) and by what means the interactions of parts can produce novelty that is qualitatively different from those processes? To borrow from Philip W. Anderson, what makes *more* different?[42]

The overlap in core theoretical concerns, the possibility of cross-fertilization in modelling techniques (such as renormalisation and coarse-graining), and concepts (boundary conditions and constraining conditions) suggests that the worries with reductionism are not restricted to the 'special' sciences like biology and cognitive science (Wusser 2022). It is in this respect that Green & Batterman argue that "Physical science approaches in biology with respect to reductionism should be revisited" (2017: 21). Remarking on the multiscale nature of physical and biological processes, they identify the *tyranny of scales* as a key physical concept that both

---

[41] Note how, as with the previous section, I will first advance through more mundane instances of emergence that do not imply any hefty philosophical baggage before proceeding to the cases that will be of interest to me in this, and remaining, chapters. I do not wish to overstate what CMP discloses about emergence and emergent phenomena—and whether it promotes the kind of strong emergence required for agency is far from an agreed upon aspect of the field (though in Active Matter Physics it would be possible to find a fair bit more agreement).

[42] Renowned physicist Philip W. Anderson is a monumental figure of 20th century physics and a key scientist in advancing the sciences of condensed matter. He passionately testified before Congress to oppose funding an expensive project that was intended exclusively for particle physics research, arguing for a more nuanced, multi-level, indeed emergentist understanding and picture of the physical world and the sciences that explore it. His biographer, Andrew Zangwill, captures this influence: "Anderson vented his frustrations [against particle physics receiving the lion's share of funding] in a 1972 article [the one I reference in the body of this work] where he pointed out that symmetry breaking generated novel properties in large many-particle systems. Moreover, he insisted, these novel properties are impossible to predict knowing only the properties of the individual constituent particles and their mutual interactions. He used this idea to attack the claim of particle physicists who asserted that the essential job of science was to discover the laws governing subatomic particles because all other 'laws' of Nature were ultimately derivable from them" (2022: 4-5).

Crucially, with respect to emergence, Anderson argued "the hierarchical structure of science (e.g., from physics to chemistry to biology to psychology) is not merely a convenient way to divide research practice. Rather, it reflects the existence of fundamental laws at each level that do not depend in any significant way on the details of the laws at lower levels. The higher level laws must be *consistent* with the subatomic laws, but the likelihood that one can derive the former from the latter is essentially zero. Anderson [thus] inspired a small intellectual renaissance … because (unbeknownst to him) his ideas revived a concept called *emergence* which had been proposed a century earlier" (2022: 5). The paper I reference in this footnote is Anderson's famous "More is Different", which, as Zangwill noted, addresses the emergence of qualitatively different features from constituents that independently lack them. Anderson's way of describing this kind of emergence is precisely: more is different.

physicists and biologists must understand in order to address the emergence of complexity in their respective fields. Briefly, the tyranny of scales refers to the challenge of understanding and describing complex systems at different length and time scales, while also imposing regimes of order (constraints) that restrict the scales with which it interacts. It also reflects the inherent difficulty in bridging the gap between microscopic phenomena, which occur at an atomic or molecular level, and mesoscale/macroscale phenomena—such as the collective behaviour of particles in solid of liquid form. In addition to understanding how different scales relate to each other, it also addresses how different materialities behave depending on the physical scale they occupy. Stuart Newman (--), for instance, remarks that the biological matter of a developing embryo often displays (at times) solid and (at times) liquid properties that give it its unique elasticity and plasticity.[43] Thus, understanding the material affordances and potentialities available to matter at different spatial scales is critical for understanding a range of processes of physical, engineering, and biological importance.

The reason this is significant is because it requires us to be sensitive to the scale-dependency of physical behaviours, as Oden describes: "Virtually all simulation methods known at the beginning of the twenty-first century were valid only for limited ranges of spatial and temporal scales" (2006: 29). These models often reveal, and are valid only within, a restricted domain of application and cannot fully capture physical phenomena spanning across large range of scales—such as the 12 orders of magnitude in time for protein folding; or the 1-orders of magnitude in spatial scale involved in designing advanced materials (ibid.). The main reason for this is that the principal physics governing events often changes with scale—necessitating a coordination of modelling techniques that capture the ramifications of these shifts and transformations (Oden 2006: 30). While the inability to capture these scale-dependent features with reductionist strategies might at first seem an epistemic limit or technological artefact awaiting more advanced computational techniques, Oden cautions that accommodating the diverse range of physical behaviours requires a different way of conceiving scientific methodologies, indeed, about cultivating a different scientific attitude (ibid.).

Sticking with the anti-reductionist argument, it is often necessary in materials sciences and CMP to elide lower-scale details. Moreover, this is necessary not simply from a modelling perspective, but in order to capture the property or behaviour of interest. To take an easy (or,

[43] Explored further in the following Chapter [insert section number].

easier) example, consider the bending of steel. At atomic scales, steel has a regular lattice structure, while at higher-scales steel exhibits more elasticity. This is captured with the well-described Navier-Cauchy elasticity equations (Green & Batterman 2018: 22). The Navier-Cauchy equations are a set of partial differential equations that describe the behaviour of linearly elastic materials subjected to external forces or deformations. They depict the steel as a continuum (that is, as homogenous): a modelling decision that elides atomic structures.[44] The Navier-Cauchy elasticity equations, in tandem with continuum models, then, are both needed when accounting for the elasticity and bending properties of steel (ibid.): "Very importantly, at intermediate (meso) scales, steel presents a host of other structures such as lamellar inclusions of pearlite, cracks, grains, boundaries, etc. To fully understand the behaviour of bending steel requires that one can combine models at different scales that inform each other" (ibid.).

Modelling in biology requires similar considerations: while any one cell may be subject to forces at the nanoscale, cells in an adhesive, multicellular context begin to exhibit novel properties. While it is often important to consider the structural diversity and typological heterogeneity of cells when modelling, at times we can treat cellular masses and tissues according to continuum models—a process called 'coarse-graining'[45]), which elides stochastic and nanoscale forces. Because many biological processes are collective in nature it is important to occasionally abstract details of individual cells, much in the same way that the ideal gas law ignores the trajectories of individual molecules to arrive at the behaviour of gases. Green & Batterman describe this as follows: "When modelling cell motion at tissue scales or that of the whole embryo, developmental biologists often rely on coupled partial differential equations that ignore the stochastic properties of interactions between individual molecules and cells. Similar to mean-field approaches in physics, they study the collective dynamics of the population of cells rather than the individual components". This makes sense when one considers that cells involved in tissue complexes behave differently than cells in

[44] Continuum models are essential to the physical sciences. These are mathematical representations that treat certain properties or materials of a system as smoothly varying, continuous functions. These models often simplify complex systems with many discrete components—like atoms, molecules, or cells—by assuming that the system can be described as a continuous medium. These models are particularly useful, as we explore, in cases where the behaviour of the system at a larger scale is more relevant than the individual components or interactions—such as in thermodynamics, fluid dynamics, and elasticity [**citation**].

[45] Another technical term, coarse-graining is the modelling method of simplifying details by averaging or approximating the properties of the system over a larger scale. This process reduces the complexity of a model by ignoring certain features, such as the positions or velocities of particles. This creates computationally more tractable models, especially in cases where the underlying fine-grained details are too complex to capture with our models.

isolation. That is, while indeed a modelling decision to help render models more tractable, the model also says something about the behaviour of cells as collectives more generally.

The examples of steel and cell collectives demonstrates the importance of factoring in material properties for better understanding gross behaviour, and these properties are often inherently emergent. Not only can they not be reduced to atomic or molecular interactions, but in attempt to do so would leave out something important in the picture of what we're trying to explain (the bending properties of steel and the viscoelastic nature of cell collectives, respectively). This is a point researchers in Materials Sciences make frequently, and a domain within Materials Sciences is that of Soft Matter. Soft matter covers a wide range of materials of both industrial and biological importance [citations], including polymers, colloids, gels, liquid crystals, viscoelastic matter, cell membranes, and so on. [46] Materials Sciences approaches to developmental biology identify the material properties of the embryo to better understand how development is managed.

Embryogenesis, then, serves as a useful contact point between the physical sciences and the biological—and how furthering our understanding of these processes requires incorporating non-reductionistic elements into our theoretical armatures. Stuart Newman describes the motivation for this as follows:

> The development of multicellular organisms involves the reshaping of cell masses and the establishment within them of precise arrangements, over time, of cells of various types. Tissues of animal embryos are soft and pliable (in contrast… to those of plants, which are solid). Animal (as well as plant) tissues are also 'excitable media', in that their mechanical configurations and chemical states (including states of gene expression) of their constituent cells are not merely passive, but draw on energy fluxes and stored energy to self-organize. (Newman & Linde-Medina 2011: 274).

Newman then draws the significance for incorporating insights from physics:

> It therefore seems reasonable to look to the physics of excitable viscoelastic matter, which pre-existed multicellularity, and indeed all living systems, as a determinant additional to genes of the origin of animal developmental systems. (Newman & Linde-Medina 2011: 275).

[46] This point will be further reinforced in the discussion on Soft Robotics in Chapter 7.

The significance of incorporating mesoscale physics into our understanding of embryogenesis is that it furnishes several key concepts such as, to name a few, nonlinear oscillators (Wang & Wolynes 2011), multi-stable dynamical systems (Prengel et al. 1994; Husiman & Weissing 2001), excitable media (Vahabli &Vicsek 2023), reaction-diffusion (Senoussi et al. 2021), viscoelastic instabilities (Hemingway et al. 2016; Emmanuel et al. 2021), and other dynamical and active processes that imprint themselves on biological media and constrain their possible configurations: "It is in fact impossible for the morphology of living systems not to bear imprint of these effects" (Newman & Linde-Medina 2011: 276). Moreover, it is precisely these material affordances and potentialities that organisms must exploit to achieve increasingly more elaborate goals and maintain complex morphologies—as we will see at the end of this section

Robert Batterman (2011), picking up on this note, suggests that the tyranny of scales had not been (and still rarely is) appreciated in philosophy. He suggests that it has implications for reductionism both in physics and biology. Most centrally, this is by reworking our very understanding of the reductionism in physics: "In many instances, philosophers hold onto some sort of ultimate reductionist picture: Whatever the fundamental theory is at the smallest, basic scale, it will be sufficient in principle to tell us about the behaviour of the system at all scales" (2011: 2). This form of 'eliminativism' is meant to imply a sort of subjectivism at worst and technological limitation at best regarding the emergent phenomena at the meso and macroscale: their behaviours and supposed causal powers being decisions we make to simplify explanations but really are 'nothing but' their constituent parts.[47] But this reductionist picture is mistaken. Not only is it mistaken about scientific practice but it is a mistaken picture of how the world works (cf. Batterman 2011: 3; Anderson 1975). Perhaps most importantly, it is not (or not entirely) a scientific thesis at all—but rather stems from prior metaphysical (and often historically sedimented) assumptions about matter, physics, and reduction.[48] In submitting that this is an outmoded view of the world, I accordingly suggest that (at least some of) the difficult surrounding emergence, realism, and naturalism in philosophy stem

[47] We saw a flavour of this 'nothing buttery' in Chapter 1, "Metaphysical Preliminaries".

[48] Barbara Drossel, working in Condensed Matter Physics, notes that reductionism qualifies as a metaphysical belief due to its lack of testability (2019). We might even say that full reductionism as a scientific thesis is "not even wrong". It requires further argumentation and support that goes beyond what can be tested. Note how it being more metaphysical in nature does not make a mark against reductionism per se—after all, I am seeking to furnish an alternative metaphysical world view that is in an important sense 'metaphysical' or metatheoretical. The problem is when reductionism is taken as an unquestioned or unobjectionable thesis that simply follows from the science. In translating reductionism into this metaphysical arena, then, I am not trying to reveal its lack of fecundity or utility, but rather open it to discussion, deliberation, and debate alongside other viable understandings and interpretations of physical phenomena.

from a *Weltanschauung* in need of reworking. One way to effectuate this shift in attitudes regarding reductionism and physicalism is to fully draw out the implications of the notion of constraints and boundary conditions, as these are importantly related to a scientifically robust account of emergence and, as we will see later, downward causation. While the previous two sections briefly referenced both concepts, I now provide a more concrete elaboration.

3.3.2 Making Sense of Constraints and Boundary Conditions in the Context of Emergence

We can characterise constraints as follows: "Constraints supplement fundamental force laws, limiting the degrees of freedom available to elementary particles, and determining that macroscopic objects behave in specific ways" (Bechtel & Winning 2018: 290). Constraints, then, enable and co-determine certain behaviours—for instance, an inclined plane functions as a constraint for a ball rolling downhill. Constraints are importantly related to restricting the degrees of freedom and possible operations of an entity in a state space (ibid.: 293). While not relevant for the present chapter, the example of state space provides an appropriate illustration. In dynamical systems theory, a state space representation is meant to capture the possible positions of each variable in the system. Any possible state of the system "corresponds to a to a point and a change in the system" and is represented as a trajectory through this state space (ibid.). That is, a state of the system is a point, and a change in the system is a trajectory through the space.

The key role of constraints is *to restrict* the possible trajectories such that certain spaces in the state-space are not traversed. Along with this negative characterisation, however, it is more important to emphasise positive role of constraints: "[they] bias a constrained object towards reaching points and trajectories in the state space that would otherwise have been practically impossible or vanishingly unlikely" (ibid.). This iterates two key points about constraints. First, it signals their causally robust nature: without constraints, certain events in a state-space would be vanishingly unlikely to occur—if not impossible entirely. Take the case of enzymes. In biology, enzymes are catalysts that dramatically facilitate and accelerate chemical reactions within cellular metabolism. More specifically, they convert substrates into products that the cell can use to perform various cellular functions. Enzymes accomplish this by helping to overcome otherwise prohibitive energy barriers—meaning without enzymes functioning as constraints in this manner substrates could not be converted into products on a relevant biological timescale.

Second, constraints are inherently *productive*—they do not just deprive a system of degrees of freedom: they enable it to perform the very functions and activities relevant to all of biology. An *unconstrained* system would have no behaviour, certainly not any of relevance for harnessing energy and driving behaviour in ways relevant to evolution and biology (to say nothing of mind and cognition). Winning & Bechtel, again, capture both the causal and productive nature of constraints as follows: "Enzymes provide an exemplar of how constraints generally account for the causal activities of mechanisms. The way mechanisms and their parts are constrained explains why both mechanisms and their components are intrinsically productive; by means of possessing such emergent powers, mechanisms and components causally produce the effects they do" (ibid.: 294).[49] An example that will become relevant going forward is that of the extracellular matrix [ECM]: the stiffness of the ECM influences the bending and elastic behaviours of cells that make up and connected through the medium. Interestingly, intervening on the quality and degree of ECM stiffness has been implicated as a factor in carcinogenesis (Green & Batterman 2017: 21).

This is interesting as ECM stiffness is an inherently emergent property and scale-dependent (e.g., it only occurs at the mesoscale with the consorting of many cells). It is not (i) reducible to the properties of individual cells and their interactions, but rather emerges at a particular spatial scale; and (ii) it exhibits novel behaviours and material properties at that scale, behaviours and properties that structure, scaffold, and constrain the interaction of cells. Without wanting to jump ahead, it is worth indicating briefly how this already provides provisional support for the notion of downward causation I defend in section (4.4) below.

The soft matter properties of biological structures highlights the importance of understanding how constraints emerge at different scales to impose order on, mediate, or otherwise co-determine lower scale happenings—further reinforcing the anti-reductionist point of this chapter. Constraints are often represented as boundary conditions in multiscale modelling. As we saw, boundary conditions can be a physical and mathematical concept. Mathematically, boundary conditions demarcate a domain of possible partial differential equations. Boundary conditions thus constrain and set the rules that define the behaviour of a physical system at its boundaries or edges. Green & Batterman (2017: 23) provide the example of a violin string to bring home this point. Solving the possible behaviours of a violin

[49] Winning & Bechtel go as far to say that constraints *ground* causality in biology (2018: 294).

string—in the form of partial differential equations—requires introducing mathematical boundary conditions that fix the endpoints of the string, in this case the bridge and the nut.

When solving the set of differential equations in this domain, one would have to assume that the bridge and the nut are completely stationary. This provides the solution to the possible modes of string activity; but, of course, an actual violin string is only sonorous when interacting with the pliability of the nut and bridge. Green & Batterman show that to capture this quality it is necessary to descend (or better, shift scales) to "focus on the molecular and sub-molecular interactions between the string and the bridge. Here the equations are of a completely different mathematical type. This is the realm of molecular dynamics governed by *ordinary* differential equations. The lesson here is that sometimes (quite often in fact) one needs to impose boundary conditions in order to efface physical details that will not allow one to model the behaviour of interest at a given scale" (2017: 23). It is important to be clear on the difference between boundaries as physical constraints and boundaries as modelling decisions made for the sake of solving partial differential equations. While the latter represent modelling decisions required for capturing a phenomenon of interest, the former represent causal constraints that influence the behaviour of the system either as a whole or along a selected domain where the behaviour functions. To see why, we will consider the case of the extracellular matrix [ECM] and how it is modelled in developmental biology.

First, sticking with the emergent and soft matter properties of the structure, the ECM is an essential component of tissues and implicated in range of biological processes of both theoretical and medical importance, such as embryogenesis and carcinogenesis.[50] It is considered an example of Soft Matter due to its dynamical, elastic, and responsive properties which are uniquely generated due to a meshwork of fibrous proteins, glycoproteins, polysaccharides, and proteoglycans—providing both structural support as well as biochemical signally (Nallanthinghal et al. 2019). Fibrous proteins, for example, include collagen, elastin, and fibronectin, which jointly provide tensile strength and elasticity to the ECM. Collagen is the common protein and endows the ECM with much of its mechanical responsiveness; while elastin ensures that the ECM can return to an original state after the deformation. Glycoproteins (such as laminin and fibronectin) essentially allow the ECM to acquire a three-dimensional structure by enabling cell-cell adhesion. Lastly, proteoglycans are macromolecules that provides the ECM its viscoelastic properties—binding growth

[50] We will consider the case of carcinogenesis in particular in the following section.

factors and regulating cell growth. Jointly, these mechanisms – to name a few – are what characterise the ECM as a Soft Matter complex with distinctive mechanical, material, and biochemical properties central to biology. Importantly, these mechanisms and materials all interact to create the emergent structure that is the ECM—and bi-directional interactions between the ECM and constituent cells are crucial for embryogenesis, wound healing, and connectivity in tissues (Checa et al. 2015). Interestingly, it is these the interactions between them that enables variation of material properties depending on the timescale of measurement—with elastic solid-like and viscous liquid-like possibilities being common.

Green & Batterman identify the ECM, along with epithelial sheets and embryonic tissues, as exemplary of the need of modelling higher-level features for understanding biological form and function. In developmental biology, as for physics, boundary conditions are introduced to represent biomechanical constraints on morphogenetic movements of epithelial sheets: "The establishment of tissue boundaries and geometrical structures during morphogenesis is mediated and stabilized by interconnected adhesions between cells and the [ECM]. Adhesion fix cells and cell populations in structures with varying degrees of freedom for bending and motility" (ibid.).[51] This description of the mediating and structuring role of the ECM allows us to further grasp the importance of how boundary conditions at higher levels of scale (here, an intercellular or multicellular context, though ECMs are not exclusive to multicellularity) facilitate cell activities that in turn contribute to the maintenance of the ECM.[52] As Checa et al. (2015) write, "It is well known that the density and distribution of cells, the mechanical properties of the ECM, and the geometry of the matrix all have an influence on traction force dynamics. From a cellular perspective, it is not the net cell traction force alone that appears essential, it is the balance between the cell tractions forces on the one side and the resistance of the extracellular matrix on the other side that causes cellular deformations and—as a consequence—alters cellular activity".

A simple way to get a grip on this to note that researchers have identified ECM *stiffness*—which like other weakly emergent properties, cannot be reduced to lower scales—contributes to cell, tissue, or organism development. Stiffness in materials sciences and condensed matter

[51] It is interesting to consider how the proteins involved in adhesion play both mechanical and signalling roles through force-transmission and mechano-transduction, that is, a conversion of mechanical stimulus into chemical cues that influence biomechanical pathways

[52] The ability of constraints to enable an internal space for chemistry—that is, generate a boundary within which metabolism may occur—that then produces the parts out of which the system can be consistently maintained is a defining feature of biological autonomy and autopoiesis.

physics refers to a material's ability (here, the ECM) to resist deformation under an applied force. It is quantified by the material's so-called elastic modulus, which is the ratio of stress to the corresponding strain (the relative deformation). Materials with high stiffness (e.g., steel) will deform less when subjected to the same force compared to materials with low stiffness (e.g., epithelial sheets and the ECM; see Green & Batterman 2017: 25). Importantly, "Since the propagation of elastic forces within the tissue goes far beyond the scale of single cells and can reach distances of centimetres, the local stress field sensed by a single cell always reflects a more global mechanistic characteristic of the ECM, which includes but is not limited to shape, its structural alignment and its overall stiffness as a composition of active cells" (Checa et. al. 2015: 2). The significance for our understanding of boundary conditions is as follows. First, it shows that they can operate at many orders of scale—with the emergent, structural features of the ECM imposing order on constituent cells and mechanisms; it also shows that lower levels can be idealised as a continuum and effacing lower-level detail for the sake of more efficiently capturing the higher-level phenomenon. Thus, a genuine place for a non-reductive engagement for higher-level features opens up here, marginalising the emphasis often placed on progressive reductionism.

As Green & Batterman show, then, bulk mechanical properties cannot be treated as epiphenomenal background conditions or behind-the-scenes set up because the materiality at these higher orders of scale "can directly influence cell differentiation and gene expression through force transmission" (2017: 29). That is, many cells are responsive to the cues that mechanical signals present, cues that often play mechanical *and* biochemical signalling roles involved in cell differentiation (this might further cast doubt on the tendency to dissociate information from its material media, as we explore in the case of multiple realizability in the penultimate chapter (Chapter 6 'Soft Robots and Bio-Inspired Artificial Intelligence')—through a key mediating mechanism called 'stretch-sensitive channels' (Miller & Davidson 2013). Engler and colleagues, picking up on the mechanical-cum-biochemical nature of the ECM, conducted an experiment wherein they cultivated mesenchymal stem cells on controllable, elastic substrates. Intriguingly, through no direct cellular or genetic intervention, Engler and colleagues manipulated the stiffness of the substrate to guide cells down osteogenic or neurogenic developmental pathways—further cautioning against an emphasis or primacy for a 'fundamental level', above which whatever is emergent is merely inert, causally superfluous, or epiphenomenal.

An upshot of this section is to show that reductionist strategies or reductionist ideals vis-à-vis the life sciences offers only an incomplete picture of many developmental and biological processes. Moreover, that expanding the theoretical armature to include and accommodate the physical sciences does not necessarily license reductionism in these domains—often times in physics, and in Condensed and Soft Matter Physics especially, it is crucial to reference emergent features to better describe or capture the phenomenon of interest. Intriguingly, as I will go on to argue presently, this work not only shows that higher scales exhibit emergence, but that their existence serves as a form of top-down causation—by far one of the most metaphysically beleaguered concepts in debates on emergence, teleology, and agency today. Using constraints as a model for understanding top-down causation allows us to not only see the emergent nature of things like the ECM or the bioelectric context of embryogenesis (Levin 2019), but it also foregrounds the discussion to come in the following chapter: namely, how developmental processes feed into, and are influenced by, behaviour at an organismic level. For now, then, I turn to more explicitly specifying the notion of top-down causation I have in mind for the remainder of this thesis.

3.4 Constraints as Exemplary of Top-Down Causation

Having broadly identified the role emergent features play in explicating many biological processes (with an explicit focus on the ECM and other physical media that must be understood in an emergentist manner), I now turn to the example of carcinogenesis wherein disturbances or defects of the ECM (and interventions thereon) are heavily implicated in whether cells develop into cancer. I present this example to show further specify a notion of top-down causation that will be operative going forward. Moreover, this example is useful as it represents a contact point between work in philosophy of physics and biology, on the one side, and the biogenic approach to life and mind on the other—often because carcinogenesis and tumorigenesis are seen as expressing a distinct kind of cellular agency.

Sara Green has written on the importance of invoking top-down causation for understanding many processes of relevance to biology—both in practice, as in the case of cancer, and in theory. Picking up on the previous discussion, she writes:

> From a biological perspective, physics is often seen as a discipline aiming for reductionist (or fundamental) explanations. … [However], I show how solid-state tissue properties in tumour progression support the importance of macroscale features of living systems.

> Moreover … that biomechanical constraints can shed light on the controversial notion of ‘downward causation’ that is often used to distinguish between reductionist and anti-reductionist approaches in philosophy of biology. (2022: 196)

This approach reconceptualises top-down causation in terms of physical scales and emergent patterns that contextualise and constrain the activities of constituents or parts at a lower physical scale—as we saw in the previous section. In doing so, it also helps to demystify the notion of top-down causation of relevance for legitimising the importance of agency and behaviour on an organismic level in a non-eliminative way (Mossio et al. 2013; Stotz & Allen 2012). Before continuing down that avenue, however, it will help to briefly summarise current literature on cancer research.

Carcinogenesis is often explored from the perspective of mutational theories involving genetic instabilities linked to somatic cells—called the somatic mutation theory (Vaux 2011). While the state-of-the-art acknowledges cancer as a multicausal pathology, the majority of research still revolves around identifying genetic markers: highlighting an entrenched, reductionist methodology in much of cancer research. However, there is growing dissatisfaction with the current state of cancer research: “Despite decades of heavy investment in research on genetic markers and molecular pathways involved in cancer development, results in terms of improved understanding and clinical control have so far been disappointing” (Green 2022: 197). The difficulty stems, in part, from the troubles involved in demarcating healthy from cancerous mutational signatures, and the difference between malignant and benign tumours—such that Soto & Sonnenschein (2011) have questioned whether a distinct category of cancer cells exists at all.

In contrast to mutational signatures, Green presents a *tissue organizational field theory*, which situates cancer development within a tissue-level context. Specifically, Green suggests incorporating elements of the so-called tissue microenvironment (of which the ECM is a part) helps us understand how “the environment of the tumour cells influences or constrains the possibilities of cell growth and proliferation” (ibid.). Evidence for this comes from experiments that intervene on the higher-level tissue environment to normalise cancer cells and resituate them within a healthy tissue setting—casting doubt on statements that depict the microenvironment as a set of ‘passive conditions’ serving ‘housekeeping functions’ for a ‘disease of the genome’ (Plutynski 2018). Pulling again from Soto & Sonnerschein, a series of experiments have reiterated the importance of contextual, environmental features for understanding cancer development. They have shown that transplanting tumour cells from a mouse into normal tissues in a host did not witness untrammelled proliferation. As Green notes, commenting on this phenomenon, "If genetic instructions determined cell growth, one would expect the hosts to develop the same rate of tumours. Instead, they observed that tumour incidence varied among the hosts and that in some cases no tumours formed at all" (2022: 199). Jointly, this evidence points towards the need for a better understanding of how higher-level physical cues influence cellular proliferation: a 'bottom-up' approach centred around genetic mutations is insufficient.[53]

Importantly, from the present perspective, it also illustrates a defensible version of top-down causation that is understood in terms of *mutual constraining relations* in the form of *context dependent constraints*: constraints that "select and delimit various types of the system's possible developments" (Emmeche et al. 2000: 25). In contrast, then, to a species of emergence that posits *autonomy* of higher levels from lower-level realisers—the fancy of philosophers of mind and cognitive scientists alike—this version of top-down causation and emergence emphasises the looped, networked, bidirectional, and heterarchical manner in which behaviours and physical processes *at different levels of scale* influence and respond to each other. The higher level is not 'insulated' from the cellular activities that come to constitute the tumour microenvironment; but neither are they determined by them. Conversely, the microenvironment and other ECM and tissue-level events—to say nothing of an organismic domain involving behaviour—contributes to the delimitation of the system's possible configurations, developments, and possibilities. When I speak of top-down causation, then, it is this 'constraint' focused account that I mostly have in mind. Green captures this well when she writes,

> By constraints, I understand features that delimit the degree of freedom of the dynamics within a given system. Constraints not only limit possibilities through restraints on the possible system states but also enable certain states that would be impossible to reach for the unconstrained system. (2022: 203).

In cancer research and modelling, an important macroscale parameter is matrix stiffness (defined by an elastic modulus; see the previous section). Stiffness, it will be recalled, expresses the deformity of a material in response to stress as homogenously applied to a surface. It also determines a material's elastic properties by treating the structure as a continuum of matter.

[53] It also highlights the importance of ontological assumptions that we bring with us to both theory and practice: making issues of top-down causation far from a niche concern in a remote region of philosophical academia.

In the case of cancer, Green writes, "matrix stiffness is inherently a tissue-scale parameter because it depicts the physical forces acting on the *integrated effects* of cell populations that are constrained by certain geometrical structures" (ibid.: 203). But there are other macroscale constraints as well that are of broad relevance to biological processes: such as cell migration, ECM fibre thickness, and cell-cell adhesion. What identifying these macroscale parameters is intended to do, then, is show how changes at a lower-level can depend on, or are influenced by, changes in high-level variables such as tissue stiffness (ibid.).[54] This, then, is one avenue for 'naturalising' downward causation. I conclude this section, then, with a glance towards more theoretical work in this area seeking to legitimate top-down causation.

3.4.1 No Privileged Level of Causation

The above discussion has canvassed some recent literature arguing for demystifying the notion of top-down causation and emergence. To do so, I have sought to show how the physical sciences are actively furnishing alternative understandings of multiscale causality and interactivity that belies the reductionist assumptions for which physics is often credited. This is intriguing, as it dramatically transforms our more general understanding of 'physicality' and 'materiality' to comfortably accommodate certain versions of top-down causation (Emmeche et al. 2000). Fodor's 'just physics' concern, introduced in Chapter 1, takes on a whole new meaning here: what 'physics' is for Fodor and what it comprises from the present argument stand very far apart: physics is the sciences of theoretical particle physics or a search for some fundamental theory, to be sure; but it is also the multiscale physics of condensed matter physics and, as we will soon see, active matter physics and materials sciences. More pointedly, it also bleeds into the domain of the life sciences—with physical sciences techniques and materials sciences concepts being integral to a fuller understanding of developmental and biological processes. However, rather than this representing a reductionist affront for biology, it actually signifies the way biologists and philosophers have been working with an outmoded conception of the physical: one intimately related to reductionism and an associated understanding of mechanism.

The above considerations dovetail to emphasise the looped and interconnected nature of biological structures—such that the desire for demarcating more fundamental levels, or otherwise blocking ascriptions of strong emergence and top-down causation, stems more

[54] Note the distinctive interventionist flavour here: interventions at one level of scale—here, the tumour microenvironment or the ECM—can lead to detectable changes at lower-levels of scale. See Wimsatt 2022 for a review of interventionism in the context of downward causation.

from philosophical assumptions about the nature of these systems than anything else. In addition to iterating the importance of revisiting a suite of philosophical assumptions that tacitly guide our theorising in the life and mind sciences (which I have elected to explore in the context of a Philosophy of Nature; see section 4.4 below), the position explored herein picks up what Denis Noble has called a principle of 'biological relativity' (2012): to wit, that there is no privileged level of causality in biology. The 'standard' way of illustrating biological happenings is presented on the left in Figure 1 below, whereas the looped and interactive nature of many biological processes is thematically illustrated on the left.

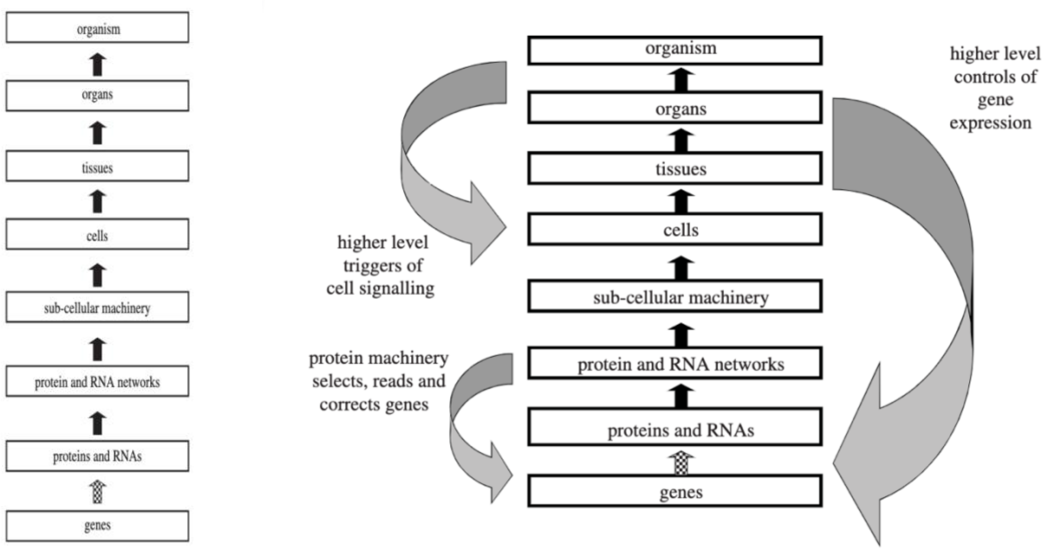

Figure 3. Denis Noble identifies two different models for understanding causality in the biological sciences: unidirectional (fundamental) interactions and bidirectional (upwards and downwards) interactions. The former is associated with causal reductionism or fundamentalism, whereas the latter identifies how the presence of higher-level features constrain or otherwise mediate lower-level interactions.

In his seminal paper, Noble cautions against literal ascriptions of levels thinking that posit discrete jumps and leaps in nature or commits to an autonomous focus on certain levels. While I have, at various points, used the term 'levels' in this work, the term 'scale' is more appropriate for my purposes. This is because it captures the more graded, continuous, and interconnected nature of systems occupying a specific scale or stretching across multiple scales: "The real reason for putting genes… at the bottom of the hierarchy is that they exist at the smallest (i.e. molecular) scale in biological systems. The formation of networks, cells, tissues, and organs can be seen as the creation of processes at larger and larger scales" (2012: 56). Scale, then, appropriately conveys the spatial nature of biological systems and, in keeping with the tyranny of scales problem above, captures the mutually constraining (understood causally in the form of 4.1.2) nature of physical behaviours and processes at different spatial and temporal scales—differences in which mediate and constrain the events at lower scales. As Noble concludes, understanding the features and properties that are relvant for integrating functions should be a matter of experimental discovery and theoretical engagement—not something to be decided *a priori* based on philosophical and methodological assumptions.

3.5 Conclusion

This chapter has covered and addressed a cluster of issues in the physical and biological sciences pertaining to emergence, reductionism, and downward causation: all of which are strongly connected concepts, such that revisiting reductionist assumptions correspondingly impacts our understanding of other contentious concepts. Importantly, the strategy I have pursued herein has been to pivot towards the physical sciences to show that when it comes to anti-emergentist philosophical presentiments in the life and mind sciences these might be based on an outmoded understanding of the physical and what it entails. In fact, the very qualifier 'just' in the 'just physics' locution that is frequently appeared in this chapter ought to be dropped. Identifying the physical basis of an emergent phenomena—cognition, agency, organismality, and so on—need not vitiate its causal status.

It is still an important empirical question in these cases how the higher relates to the lower, and much remains to be elucidated, experimentalised, and explicated; but the philosophical mysteriousness that is often seen as part and parcel of the emergentist literature should be attenuated. Again, my goal in this work is not to *resolve* very interesting and pressing issues that surround how agential or mental phenomena relate to the structural and self-organizing processes I've identified, but rather to open a philosophical space where we can more productively communicate about the nature of this relation that does not encounter immutable dichotomies. In the next chapter, I continue the discussion and draw the argument closer to the far-from-equilibrium processes. This linking is required because organisms and agents exhibit a kind of ineliminable activity that inheres within the soft matter and active matter properties of these systems. It is the *structural-material* and the *energetic* dispositions of these systems that ultimately does the heaviest lifting for an account of naturalised agency. With some of the structural details on the table, we can now turn to a discussion on active matter.

Chapter 3

Just Physics? Rethinking Emergence and Causation in Physics and Developmental Biology

Chapter 4

The Extraordinary Liveliness of the World: Active Matter with an Eye Towards Agency

4.1 Introduction: From Passive to Active Matter [55]

> [The] appellation 'living things' assumes a category error: life is not a property inherent in things so much as things are instantiations of organizational states that arise within a larger context of life.
> - Smith & Morowitz, *The Origin and Nature of Life on Earth* [2022: 11]

This chapter builds off the previous one to further the main line of argument of this thesis: namely, that we need to return to our assumptions and understandings of materiality to better align our intuitions in this domain with a range of physical and biological advancements—coming from within disciplines that are themselves progressively elucidating the multilevel, heterarchical, and highly active nature of material processes and their contribution to characteristically minded features such as agency and goal-directedness. While the previous chapter only made reference to the *active*—that is, energetically driven—nature of biological processes, I now turn to foreground these features further as it is ultimately, I submit, the material-energetic situation of a system that is key to determining its agential status from within the biogenic approach. In fact, one could say that the thermodynamic and energetic situation of the system is what makes an agent an agent, and materiality matters insofar as it allows for the maintenance and sustaining of these flows of energy. This not only connects to a rich body of literature seeking to naturalise goal-directedness with the language of far-from-equilibrium thermodynamics (e.g., Moreno & Mossio 2015), but also provides a direct link to an exciting body of literature in the physical sciences called Active Matter Physics [AMP], which will be a primary focus of this chapter. This will encourage a pivot towards a revamped understanding of materiality as not only highly active, plastic, and emergentist in nature; but also emphasises the relational nature of these processes, that is, it highlights what in the philosophical literature is called a relational ontology [RO].

[55] Part of this chapter is based on forthcoming work in *Synthese* entitled "Rethinking the Nature of Matter: Naturalising Teleology and the Kantian Legacy".

Briefly, AMP is a discipline within Condensed Matter Physics [CMP] that focuses on systems composed of entities—called 'units'[56]—that are capable of self-propulsion or energy conversion (Needleman & Dogic 2017). These entities are often referred to as 'active particles'[57] and can be synthetic micro- or nanomotors (such as Janus particles), self-propelled colloids, or living organisms such as bacteria or cells. AMP aims to understand the principles underpinning a range of collective behaviours and emergent properties thereof. It is in this sense that the current argument builds off the work set out in Chapter 3: we have already canvassed the weak and strong emergences explored in Condensed Matter Physics and Developmental Biology, and it is now time to link the discussion with the productivity and inherent activity of biological and certain non-biological media.[58] One of the most exciting conceptual implications of this line of research is the elucidation of principles of *self-organization.* Indeed, the distinction between *self-assembly* and *self-organization* is a crucial one both for AMP and the approach to agency explored in this work, making it helpful to better specify what is involved in the passive and active matter distinction.[59]

We have seen this distinction before (Chapter 1, 'metaphysical preliminaries'), but to recapitulate quickly the difference comes down to how each process can be specified thermodynamically: "Although self-organization implies a nonequilibrium process, self-assembly is reserved for spontaneous processes tending towards equilibrium" (Halley & Winkler 2007: 10). The difference in thermodynamic processes also implies a different way in which these processes can be quantified and measured (Eckbart 2021; Bowick et al. 2022). For one, the very language that we must use to describe the behaviour of active self-organizing

[56] Though what constitutes or qualifies as a 'unit' can be quite underspecified and covers a diverse range: from colloidal molecules, to individual bacteria, and all the way to whole multicellular organisms like individual birds in a migrating flock.

[57] Siriam Ramaswamy, who has contributed extensively to the development of Active Matter as a distinct category, describes the conditions for it as follows: "To qualify as active, a particle needs enough structure to couple nontrivially to the free-energy input; in practice surprisingly little structure is required, and surprisingly small single-particle departures from pure thermal movements, to yield impressive collective nonequilibrium phenomena. The grand aim of the active matter paradigm is twofold: to bring living systems into the inclusive ambit of condensed matter physics, and to discover the emergent statistical and thermodynamic laws governing matter made of intrinsically driven particles." (2017: 3).

[58] It is tempting to understand this as plugging into the *Naturphilosophie* point of the *natura naturans*: furnishing a philosophy of nature that grasps materiality in its productivity, its individuating, not (or not exclusively) in determinate products. See Chapter 2.

[59] We have already thoroughly discussed the idea of self-organization, its relevance for biological autonomy and agency, and the role it tends to play in debates on naturalised teleology.

systems is different from the one we would select for a self-assembling process: the former requires the mathematical language of nonequilibrium thermodynamics whereas the latter can be described according to principles of statistical mechanics. In other words, one deals with dissipative structures and the other does not. In addition to their different mathematical languages, the two also imply different suites of concepts for making sense of their activity—many of which relate to the discussion so far on reductionism, holism, and emergence. We will consider some examples to get a preliminary grip on the distinction and to better set the stage for how a distinct category of 'active matter' contributes to the discussion.

Self-assembly involves the physical association of molecules in a medium into a steady and energetically favourable state, typically in the absence of an external energy source. As Nicholson describes, self-assembly "is driven by local stereospecific interactions between the aggregating 'building blocks', which remain unchanged throughout the process" (2019: 111). Moreover, because the emergent properties of the resultant structure are determined by the properties of its parts, processes of self-assembly need not involve the wider context or other emergent features that could otherwise be at play as they tend towards equilibrium. The common examples of self-assembly are formation of viral capsules or the helical structure of DNA (Alberts et al. 2008). Crucially, the parts themselves remain almost entirely unaffected by the process of self-assembly, regardless of what new properties the resultant structure confers. Self-organization, contrastively, departs from this in important respects: "Much of chemistry [relevant to self-organization] concerns interactions (reactions) where the products are not simple assemblages of components but are distinct new entities" (Halley & Winkler 2007: 13). Self-organization, then, requires not only a different kind of dynamical[60] process to sustain it, but also the possibility of converting old matter into new matter in order to effectuate this pattern of activity.

Self-organizing processes, then, imply nonequilibrium conditions involving both external energy input and a manner of converting, storing, or utilising this energy to generate an internal chemical process.[61] Looking to Halley & Winkler (2007) again, we see that self-

[60] There is an interesting point of equivocation here in the literature on the dynamical nature of these processes. Often, sustaining dynamical activity is equated with self-organization or otherwise active processes. But the qualifier dynamic is a technical one and can apply to 'simpler' processes of self-assembly (Halley & Winkler 2007: 11).

[61] This is correspondingly what makes self-organization a *self-individuating* process as it implies the maintenance of some kind of boundary wherein an increasingly complex internal chemistry (i.e., a metabolism) may occur. We explore the relevance of this to agency further in Chapter 5.

organization necessarily reflects an interplay between internal and external sources of order, and ultimately identifying an environment—in the biological context—and a medium—in an engineered or laboratory one—is required when drawing a distinction between the two. Stated differently, self-organization refers to the ability of these active units to interact collectively in a nonlinear manner to generate a dissipative structure; one that, for a time, is able to fend off the otherwise pervasive tendency towards equilibrium. Examples of these processes—processes which are increasingly appreciated for their active matter nature—include the actomyosin network (Popov et al. 2016) and microtubules (Sanchez et al. 2012). Pealing back the details of the latter slightly, microtubules are architectural features crisscrossing the cell and constantly polymerising and depolymerising through successive cycles of GTP hydrolysis to maintain their steady state (Nicholson 2019: 111). This animates their skeletal description, as Fletcher & Mullins describe, because "the cytoskeleton is not a fixed structure whose function can be understood in isolation. Rather, it is a dynamic and adaptive structure whose component polymers and regulatory proteins are in constant flux" (2010: 485). This will play into one of the ontology-related questions of this chapter (and, to stress again, the importance of understanding emergence as a relational phenomenon) that tries to establish a *relational ontology* for better understanding processes relevant to agency.

One might question why something like a relational ontology is required for a better grasp and understanding of the agential properties of interest in this work. Philosophically, relations have tended to be viewed as secondary to what would otherwise be primary ontological categories like 'substances'. But agency is an intrinsically embedded, interactive, emergent, and relational feature of the systems explored in this work. As will become clear later (Section 4.4), substantiating, as it were, the relevance of a philosophically nuanced notion of relationality is integral to a more robust philosophical understanding of agency as it reworks several assumptions that I have tried to problematise in this work. While the majority of the background here was covered in Chapter 1 ('Metaphysical Preliminaries'), I dedicate a fuller discussion to the question of relationality herein as we are not to understand agential dynamics and the emergence on which it is based *without* understanding how relations are metaphysically robust operations in the world that generate, indeed support, the emergence of new structures: something we have already hinted at in the previous chapter.

It is here that the concept of active matter as a distinct category of material organization enters the discussion.[62] While not every nonequilibrium process or system is an active matter one, the term picks out a distinct category of internally driven systems capable of self-propelled motion and the ability to chemically convert fuel and free energy into work (Kaufman 1993). To connect to some of the previous discussion, it is the designation of an active matter system that enables the differentiation between those systems that can be said to be, or come to be, agential ones and those that are not—at least insofar as they pertain to the biogenic approach and studies in Basal Cognition.

Stated differently, while tornadoes and dust devils are nonequilibrium, they do not satisfy the designation of active matter because their manner of self-driven activity is not internally generated; nor does it exhibit the internal differentiation allowing for a complex chemistry or internal mode of self-maintenance (e.g., a metabolism) to occur. In fact, we might even say of these systems that they do not have proper 'selves' at all. This is interesting because the (in)ability to draw the boundaries around systems which can properly be said to be 'goal-directed' and agentic and those which are not (i.e., organisms versus dust devils) presents a conceptual dilemma for many accounts engaged in the naturalising teleology literature (as Cusimano & Sterner (2020) argue and which I explore in Harrison (forthcoming)). I submit that, equipped with the active matter framework, we can better identify something of a demarcation between agential biological—and potential nonbiological—systems and those that do not exhibit this pattern of activity.

The conclusion of this chapter ends on a philosophical note—one that pertains to our manner of ontologising in science and theory and the kinds of entities we posit in order to explain the origin and genesis of complex behaviours. A long-standing prejudice, one which we overviewed in the first chapter of this thesis, is that what is primary (in both explanatory and, what often amounts to similar things, causal terms) is substances and essences which are usually construed in isolation of their involvement in networks of activity that can subsequently confer causal powers. That is, whatever causal powers do emerge at a specific level of material organization would themselves be based on a more fundamental causal reality. This is an ontology that vitiates relationality and interactivity in favour of simpler constituents. In contrast to this view, I submit that the exciting work and discourse

[62] It is important to preface this discussion by stating that active matter, while a nonequilibrium process, does not exhaust the range of nonequilibrium systems.

proceeding from the domains of condensed and active matter physics promotes an alternative construal of material and causal processes that foregrounds and highlights *relations* as what are central for understanding the emergence of complex forms in physics and biology. It is the incessant activity and directionality that pervades material reality that leads Karen Barad to argue that this "makes manifest the extraordinary liveliness of the world" (2007: 91).

It is precisely this 'liveliness' that I try to capture in the latter half of this chapter by introducing and defending a *relational ontology*—one that explores self-organization in characteristically non-living processes. What this is meant to capture is the gradation and progressive shading into, and out of, the living from the non-living.[63] This might be a surprising statement to make, considering up until now I have been trying to draw direct lines between self-organization, living activity, and the agential or proto-agential capacities emblematic of life. In fact, the naturalisation strategy pursued herein is crucially dependent on some dimensions or facets of this self-organization being a more general phenomenon than is commonly assumed. Stuart Newman, for instance, has identified 'generic' and 'biogeneric' (Newman 2019) material processes, where 'generic' here represents more general features or laws governing physical behaviour (at the mesoscale matter can behave at once like a liquid and then at another time as solid, which is found both in engineering applications and in embryogenesis); and 'biogeneric' include the aforementioned generic processes that are further amplified, enhanced, or augmented by the mechanisms honed and cultivated by evolutionary processes. Importantly, and to pick up on a point in the previous chapter, the ability of either to occur at all depends crucially on bi-directional interactions between physical scales and the behaviours that can occur at them—further foregrounding the point of *relationality* for the argument concluding this chapter.

This chapter will be structured as follows. In section (4.2) I overview the basics of the AMP framework and provide some common examples to illustrate the difference between active and passive matter systems. Section (4.3) relates the discussion of AMP more directly to considerations of intrinsic purposiveness and agency, closing the line of argumentation

[63] As SMith & Morwitz (2022: 2) write, "Life inherited the laws of geochemistry, and grew out of geochemical precursors because some of those laws required the formation of a new state of order qualitatively unlike any of the lifeless states of matter". It is interesting that in their *magnus opus* on the origins of life they are constantly identifying the physical and chemical precursors that antedated the emergence of life and that, with its emergence, become internalised, incorporated, and embedded into the very basics of living processes and their amplification in higher-level forms. As they write "The chemical activity at the rock/water interface is closely connected to the chemical activity *within* living systems

started in Chapters 2 and 3. It will be argued that AMP encourages the return to the notion of matter and a general reformulation of materiality as reductive, brute, and passive. Section (4.4) then explores the implications the foregoing discussion has for our understanding of relationality and emergence. I conclude in section (4.5) by setting the stage for naturalised accounts of agency, the topic of the following chapter.

4.2 Active Matters

To reiterate, AMP is a subdiscipline within CMP that focuses on the chemical and physical interactions at play within self-organizing dissipative structures. Perhaps one of the most surprising and exciting insights to come out of AMP is the identification that non-living systems can display certain patterns of thermodynamic activity normally seen as excluded of the living systems. This section overviews some of the basic ideas animating the AMP framework and canvases both biological and nonbiological examples.[64]

The physicist Siriam Ramaswamay describes the nature of active matter as follows: "Active matter are driven systems in which energy is supplied directly, isotropically and independently at the level of the individual constituents—active particles—which, in dissipating it, generally achieve some kind of systematic movement" (2017: 2). This is meant to additionally differentiate between how systems both receive energy (e.g., field driven in the case of ferromagnetism) and how they are able to convert that energy for systematic or directional movement (that is, to perform work). With this in mind, Ramaswamay identifies the key features and goals of AMP. First, to qualify as active a particle or unit needs to be structured nontrivially to free energy input—such that chemical or mechanical energy can be harnessed to drive work cycles. Interestingly, as Ramaswamy notes (ibid.: 3), in practice impressive nonequilibrium phenomena can emerge due to surprisingly small departures from thermal movement. The main goal of the active matter paradigm, then, is to bring living systems "into the inclusive ambit of condensed matter physics", consistent with the previous chapter, and to identify the emergent statistical and thermodynamic laws that underpin systems of intrinsically driven particles (ibid.).

[64] This point also helps me to defend against certain interpretations of bio-chauvinism, that is, the belief that it is biological materials only that are capable of exhibiting the properties associated with mind and agency (Searle 1984). While I do emphasis the importance of materiality, this is not supposed to come at the neglect of a range of possible nonbiological or exobiological systems which could be called goal-directed in a minimal or prototypical sense—something that AMP is actively exploring.

Examples of these kinds of systems span the living and non-living domains. As an example of the latter, platinum-tipped nanorods immersed in a solution of hydrogen peroxide will catalyse the breakdown of the solution. This in turns creates polarised flows at the interface of the rods which allow them to propel through the medium. So long as free-energy consumption and constant movement are maintained, the nanorods are capable of sustaining directional movement—albeit, of a much simpler kind than what we find in the examples below. Moreover, the nanorods lack the kind of internal differentiation necessary for more elaborate and flexible forms of movement. A more complex example we will discuss below is that of an active motile oil droplet (Eckbart 2022). Through processes relating to surfactants, chemical composition, and a suitably composed environment, oil droplets composed of nitrobenzene can exhibit proto-directional activity that enables them to continuously 'search' gradients. While no doubt lacking the kind of internal complexity and saturation of existential needs associated with agency in the biological domain[65], these model systems have been taken as exhibiting precursor dynamics from which more complex goal-directed behaviour can emerge (McGivern 2020), a point to which we will revisit shortly.

In addition to identifying the thermodynamic nature of the system, researchers in AMP often identify the material and spatial properties of both the macroscale system and their components. Examples of these properties are hysteresis (a memory effect that refers to a material's ability to go through a selective and irreversible change when a transformation occurs), anisotropy or polarity (such as a Janus particle exhibiting dipolar sides that lead to differential chemical reactions that can continuously drive far-from-equilibrium activity), toughness (the amount of energy a material can absorb before fracturing), and elasticity (similar to toughness, this refers to the amount of energy a material can handle), among others (Tripaldi 2022).[66] It is when there is material coordination in the form of self-organization that we can begin to talk about active matter systems: the actomyosin skeleton (Needelman

[65] That is, they lack *intrinsic purposiveness* but represent intriguing steps on the way to the forms of teleological systems that would have emerged early in evolution (Smith & Morowitz 2022).

[66] Although I cannot explore the suggestion further herein due to space limitations, it is interesting to consider how material sciences resituates what would commonly be taken as 'secondary properties' of perception into the fabric and configurations of materials themselves. This is intriguing because properties of 'brittleness', 'elasticity', and 'toughness' are often presented as subjective or observer-dependent properties *par excellence*. In taking active matter physics and materials science more seriously, I suspect that whether this is the case is much more an empirical question than the pedigreed philosophising on the topic would suggest. More generally, it encourages a reconsideration of the entire 'primary' and 'secondary' qualities of perception and the assumptions that scaffold the debate. It is therefore interesting to consider how the reshaping of the organic and the inorganic, the teleological and the non-teleological, has ramifications for other dichotomies such as the 'subject' and 'object'—and ultimately, mind and matter.

& Dogic 2017); Julicher et al. 2018), cellular activities (Fodor & Marchetti 2018), collective swimming (Wioland et al. 2018; Kempf et al. 2019), and even macroscale organisations such as avian murmurations and herds of animals (Cichos et al. 2020).

At first blush, this appears to be a heterogeneous assortment of systems. However, all of them are characterised by self-propelled units that interact to form collective far-from-equilibrium behaviours. Moreover, these self-propelled units (motor proteins, cells, individual organisms) are themselves highly competent, active contributors to group dynamics, being able to respond flexibly to local transformations and yet are constrained by higher-level contextual factors that mediate and constrain the degrees of freedom of individual units (see Green & Batterman 2017; Greeen 2018, 2021; Ebrahimkhani & Levin 2021; Batterman 2021). This brings us back to the discussion that opened this chapter: namely, how AMP furnishes resources for better identifying *self-organizing* features on the way to more fully fleshed agency.

As we have seen, this is precisely one of the goals of AMP and materials science: to reveal the nature of the specific material embodiments that enable endogenously driven self-organizing activity—something not evidenced in, nor made possible by, the pell-mell list of materials commonly invoked to undermine the case of naturalised teleology (such as candle flames, as seen above, or Swiss Cheese in Putnam's more colourful example). Much of the discussion in the literature, then, turns on notions of *passive* and *active* matter. This distinction primarily refers to the way in which energy is utilised, transformed, or transferred within and through the materials in question. For example, passive materials are studied for their load-bearing capacities and other mechanofunctionalities that do not require energy input and rely on chemical and/or mechanical potential only. Contrastively, active materials receive significant attention and relevance within the life sciences because of their ability to harness chemical reactions and transform energy input into mechanical responses (Egan et al. 2015). As Bowick et al. phrase it: "The defining property of an active system is that the energy input that maintains the system out of equilibrium, whether truly internal or created by contact with a proximate surface, acts individually and independently on each 'active particle'" (2021: 1).

Speaking to the transformation of energetic potential into mechanofunctionalities, they continue by saying that "once the chemo-mechanical processes that convert fuel into motion are integrated out, the dynamics of such active entities breaks time reversal symmetry in a local and sustained manner" (ibid.). In other words, what is critical about the active properties of biological materials is their ability to (1) constrain individual 'units'[67] that can independently, but in a concerted manner, maintain the activity of the system; and (2) introduce nonlinearities and time-breaking symmetries into the dynamics that break the balance we might otherwise find in Newtonian mechanics and that therefore generates self-sustained flows of activity and cyclical currents. Active matter systems thus embody a unqiue property of living systems: the ability to convert energy inserted at the molecular scale into organized motion and behaviour at a higher order of scale (Bowick et al. 2021: 2).[68] In part, then, whether something qualifies as passive or active comes down to the framework within which it can be understood and investigated. In the case of active matter, this can only occur within the framework of nonequilibrium physics and dissipative structures (Bechinger et al. 2016).

Having introduced the difference between active and passive matter systems as explored in AMP and materials science, it is worth pausing for a moment and further elaborating what makes active matter unique for many of the processes and capacities of significance to life. Thus, Marchetti and colleagues write that active matter

> [Is] composed of self-driven units… each capable of converting stored or ambient free energy into systematic movement. The interaction of active particles with each other, and with the medium they live in, give rise to highly correlated collective motion and mechanical stress. Active particles are generally elongated and their direction of self-propulsion is set by their own anisotropy rather than by an external field. (2013: 1144)

What Marchetti et al. highlight here is that the units that comprise an active matter system must independently convert and maintain energy transformation at an individual level in order to contribute to group behaviour; and, moreover, that spatial and physical properties (isotropy, nematics, polarity; see figure 4) influence and bias the directionality of constituent units—highlighting the importance of the shape and spatial properties of a unit for making sense of the specific material embodiment in which it is embedded. We will return to this point further below. For now, it will help to work in reverse by looking at a paradigmatic case

[67] Though what qualifies as a 'unit' in active matter physics can be quite vague, liberal, and open-ended—covering anisotropic rods through to an entire organism.

[68] Importantly, however, while AMP has focused primarily on this "defining property" of life, it does not seek to sequester this insight within the field of the life sciences; rather, principles of AMP and biological properties have been exploited helpfully in the domain of embodied robotics (Appiah et al. 2019), cognitive science (McGivern 2020; Brancazio 2022), and computer science/Artificial Intelligence (Kaspar et al. 2021). The implications for some of these disciplines is explored in (Harrison et al. Forthcoming).

that is commonly referenced in debates on emergence and naturalising teleology: that of the Rayleigh-Benard cell.

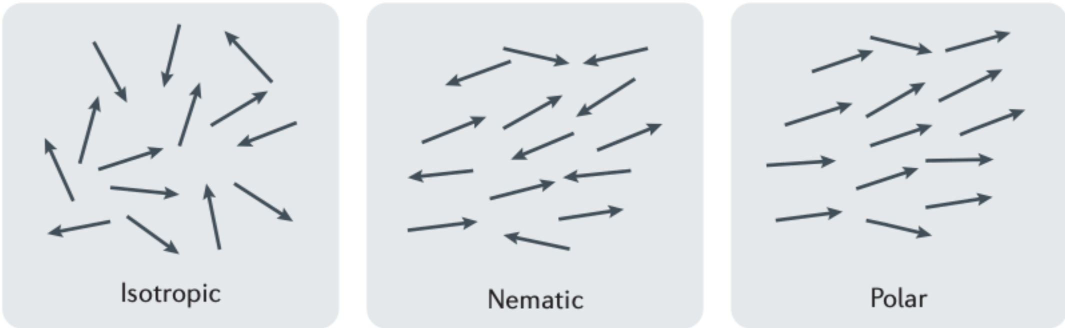

Figure 4. The different orientations and directions in which active particles can be ordered and which generate different properties at the macroscale accordingly. Marchetti et al. (2013).

Rayleigh-Benard cells are an example of non-equilibrium activity. Their activity is familiar to anyone who has casually placed cool oil onto an evenly heated pan. The abrupt contact of the oil with the hot surface creates a bifurcation that constrains the activity of constituent oil molecules; that is, the pan—operating as an external energy source—places constraints on the degrees of freedom available to the individual oil molecules. Whereas oil at room temperature consists of disoriented or unoriented molecules, the degree of freedom of each unit becomes more constrained when wound up in the hexagonal structure of the Rayleigh-Benard cell: individual molecules become highly correlated so long as the energy input is continuously maintained. Due to their minimal nature as far-from-equilibrium dissipative structures, the Rayleigh-Benard cell has commonly been referenced to highlight "the precursor dynamics" from which end- or goal-directedness "endogenously emerge" (Juarrero 2015: 512). It also exhibits a degree of robustness, remaining stable under reasonably small perturbations (Batterman 2022: 224).

Despite their relevance as models for nonequilibrium processes and even as models for teleological dynamics, "this out-of-equilibrium steady-state flow *requires* that thermal energy be introduced to the [pan] exogenously" (2022: 224). While its constituent molecules are constrained in a manner that drives the Rayleigh-Benard cell's unique directionality, each convection roll is composed of passive molecules and is driven out of equilibrium by energy input provided at an external macroscopic boundary (Needleman & Dogic 2017: 1-2). In terms provided above, the Rayleigh-Benard cell does not qualify as a case of an active matter system because constituent molecules entrained in the wider collective do not independently or actively harness energy to drive emergent behaviour; rather, the emergent hexagonal order is driven by an external macroscopic boundary and is in no way inherent to the properties of the medium itself.

This therefore contrasts with many materials that are implicated within the living domain, many of which are comprised of innumerable of active molecules that interact nonlinearly to create mesoscale self-organising structures, such as the slime mould plasmodium (Angel et al. 2021)) or tissue structures (Levin 2021).[69] Common examples span the living and non-living, including self-propelled oil droplets (Hanczyc & Ikegami 2010; Hanczyc 2011; Cejkova et al. 2014), microtubule networks (Sanchez et al. 2012), cytoplasmic flow (Moligner & Manhart 2018), and the eukaryotic cytoskeleton (Brugues & Needleman 2014). The key difference between the Rayleigh-Benard cell discussed above and the materials that constitute these active systems can be elaborated as follows: "The cellular cytoskeleton, cells, and entire tissues are driven away from equilibrium by the continuous motion of thousands of constituent nanoscale molecular motors, protein-based machines that transform chemical energy into mechanical motion" (Needleman & Dogic 2017: 2). In addition to highlighting the relevance of these processes to biology, researchers in AMP seek to highlight how biological materiality is based on a more 'generic' (Newman 2011) activity that is pervasive in the physical world—placing biological activity (in evolution and development) on a more graded scale. Given that the return to the notion of matter I explore and encourage requires a reconceptualization of this inherent activity that is not confined to the living domain but rather represents a scaling up of these processes in biologically relevant ways, it is worth elaborating an example of these kinds of dynamics that can be posited as "precursor dynamics" (Juarrero 2014) from which goal-directedness and agency endogenously emerge. The

[69] It is important to clarify here that organisms are composed of both *passive* and *active* materials (again, understood in terms of load-bearing capacities and the ability to transform energy into motor outputs). For instance, Ren and colleagues remark that plant actuation relies on both passive and active properties. They write: "Active systems (biochemical-energy-based actuation) activate and control the response by moving ions and altering the permeability of membranes. Passive systems (humidity-driven activation) are mainly based on dead tissues designed to undergo predetermined changes under environmental conditions" (2021: 2). Thus, the current differentiation between active and passive systems is not meant to demarcate living and cognitive processes from non-living or noncognitive ones *per se*, but rather to highlight how certain materials, by their very nature, can come to be implicated in a wider organisation capable of sustaining life and mind. What matters is organisation, and active and passive materials must be played off against one another to reach the form and function of a specific system's niche. The main point is this: for too long the question of organisation has been treated as orthogonal or independent of matter and materiality; but whether certain materials can enter, be sustained or sustain themselves in, and contribute actively to the self-organising causal structure on which life and mind depends is an empirical question that requires us to look more closely at the specific nature of the material embodiments of the system in question. By deferring to materials sciences and AMP, it is my goal to place philosophical approaches to teleology vis-à-vis life and mind within an empirical testbed and research program that can realistically constrain the possibility space for mindedness in nature.

example we will look at is the one mentioned above on the self-propelled oil droplet and its relevance in the 'minimal cognition' literature.

4.3 The Extraordinary Liveliness of the World: Self-Propelled Oil Droplets and a Renewed Understanding of Materiality

Although my interest in this thesis is more on naturalised forms of agency—and agency insofar as it is an antecedent concept or condition for cognition—it is difficult to extricate discussions of minimal agency from the more longstanding debates on 'minimal cognition'. It is in the latter domain that discussions of active matter have gained particular prominence—while also proving more controversial (see Lyon 2020). We have already discussed in the Introduction that one way, I submit, to dispel some of the controversy is to shift the goal posts away from ascribing cognition to these more basal or physical processes and instead reframe the discussion around a kind of active materiality that is widely found in nature and which ultimately acts as a precondition for a minimal kind of *agency*—one that must be in place in order for the aims of Basal Cognition and the biogenic approach to better substantiate their claims concerning cognitive and agential systems.

In this section I introduce the active motile oil-droplet as an exemplar case explored in the domain of AMP and explore its relevance to the naturalisation of teleology through the minimal cognition, as this is the arena where it has received the most discussion. In this vein, Hanczyc & Ikegami write, "The fundamental basis for cognition may already be present in simple non-living physical systems that possess a limited suite of properties also found in living systems" (2010: 233). To explore this suggestion, they present the motile oil droplet (first formulated and explored in the Vicsek model (Vicsek et al. 1995)) which is understood as an active matter system exhibiting many features and behaviours tantalizingly similar to the proto-cognitive ones found in biology (cf. Lyon 2015), such as navigating chemical gradients and a minimal kind of self-maintenance. Unlike the Rayleigh-Benard cell—the usual paragon of nonequilibrium self-organization—the motile oil droplet exhibits a degree of internal chemical complexity that allows it to not only remain out-of-equilibrium but to navigate chemical gradients in ways that keep its internal chemistry persisting.[70] The

[70] To contrast again with the Rayleigh-Benard cell, the motile oil droplet relies on sufficient chemical precursor being present in the medium but it does not rely on direct energy input or continuation through an external boundary like the pan in the case of the Rayleigh-Benard Cell. This further iterates the point that the manner in which the two systems are driven is of a critically different kind.

significance of this is that "[Cognition] could then be traced down to physiochemical phenomena, such as thermodynamic fluctuations in open nonequilibrium systems" (Hanczyc & Ikegami 2010: 233). Before proceeding to these conceptual implications, it will help to briefly explicate how the system is organized.

Motile oil droplets are active material systems consisting initially of an oil droplet in water and containing surfactants, a compound that lowers the interfacial tension between two different substances (in this case: between the oil droplet and the water). Researchers then introduce a chemical precursor (here, oleic anhydride) into the oil (nitrobenzene) that hydrolyses into more surfactants when it contacts the water at the oil-water interface.[71] This reaction powers both the droplet's movement and enables sustained activity so long as the precursor is present. As Hanczyc & Ikegami describe: "By embedding a catabolic chemical reaction in [the] oil droplet body, we have determined some of the conditions necessary to establish an interactive loop that involves the global movement of the system in a wet chemistry model" (2010: 233-234). Moreover, the distribution of the precursor is uneven throughout the droplet and circulated via a cyclical convective flow that establishes a chemical gradient and is created by the droplet itself as a product of the chemical reaction (ibid: 234). It is thus not driven exclusively by an external boundary condition, but by the energetically constituted and internally sustained dynamics of the oil-droplet system.

The convective flow itself brings new precursor into one pole of the system while controlling the release of products on the opposite pole. The differential distribution of precursor as scaffolded by the convective flow creates a recursive feedback loop that enables the continued maintenance of the oil-droplet's activities. Hanczyc & Ikegami note that this generates a minimal form of 'sensory-motor coupling', specifically highlighting the functional role of its interfacial properties:

> All the complexity of this self-moving motion ultimately depends on the oil-water interface and the integration of the sensor and actuator into the same dynamic structure. The interface is not a hard-shell container, but a soft and flexible boundary under tension that can interact physically and chemically with the local environment. … Therefore, we consider the interface to be the sensor of the system. (2010: 235)

[71] This and the following description of the oil droplet system builds on Hanczyc & Ikegami (2010).

While there are many interesting dimensions worth highlighting in the motile oil droplet example, I finish by highlighting one last feature of relevance for its taxic behaviour. Recall from earlier that AMP often highlights the spatial configuration of units and systems that endow them with their directionality (as we saw above with the case of isotrophy and other forms of polarity). This applies equally to the oil-droplet system: the geometry of the interface shape can control sensitivity to the environment – and geometry induced fluctuations can be the source of fluctuation in movement (ibid.: 236). In other words, the unique spatial properties and shape of the oil droplet matters for the kinds of dynamics it is capable of sustaining. For instance, the oil-droplet can assume either a spherical or horseshoe shape (as determined by occupying a specific spatial range), with the latter exhibiting straight directional movement and being subject to forces that do not predominate at the spatial scale where the spherical shape is more favourable.

I would like to draw a cluster of implications of taking the motile oil droplet as a better case of non-living precursor dynamics than those that commonly feature in philosophical debates on naturalised teleology. First, as I have mentioned elsewhere (see Harrison, forthcoming), the language of self-organization—and the notion of *closure of constraints* found in the Organizational Account of biological autonomy—is often posited as critical to the identification of agential, intrinsically purposive systems. But without a more explicit identification of the energetic situation of the system then the otherwise formal language of closure of constraints can be descriptively applied to a range of systems (Cusimano & Sterner 2020), many of which we would not *prima facie* include in a theory of agency. As such, one might wonder how we get from tornadoes and dust devils to biological agency: what is doing the important work to bridge the gap and, hence, make agency and closure of constraints both a naturalised and robust phenomena and not—as a standard view would have it—the result of limitations of a specific scientific framework at this time? I think the seed of an answer is found in work on AMP: that we must be sensitive to the material and contextual conditions of self-organization such that we identify the kinds and types of materials; the spatial properties thereof; and the physical properties that occupy certain scales in order to properly demarcate non-agential, proto-agential, and agential systems. Something like this matter-first (including energy) perspective is found in early iterations of the Organizational Account (see Exterbaxia & Moreno 2005), for instance, but later work seems to prefer the formal language of closure of constraints that is relatively neutral when it comes to the details of material instantiation.

Second, the case of the motile oil droplet highlights the significance of spatial properties and scale for understanding the constraints placed on development and, ultimately, more cognitive capacities such as sensorimotor coordination, learning, memory, and goal-directedness. The reason this is the case is because cognition or agency can often be seen as presented independent from their physical instantiation—even in accounts that would be more amenable to the view laid out herein such as autopoiesis.[72] Again, pulling from Toscano[73] we see that "It is symptomatic of the shortcomings of the autopoietic account that it eventually results in the claim that the autopoietic production of a system qua self-referential network of relations can be decoupled in principle from its [developmental] and evolutionary conditions of realisation" (2010: 59). By Varela's own admission, "The organisation of a machine implies matter, but matter does not enter into it as such" (1989: 42). If what I am suggesting is correct, then engagement with the AMP and Materials Sciences literature ought to compliment the naturalisation drives associated with, e.g., the Organizational Account and autopoietic approaches to life and mind while simultaneously avoiding concerns of 'biochauvinism'[74]; and it does so by identifying the large, but not radically open-ended, space in which the dynamics associated with agency can emerge. This would keep out certain examples that plague the debates on naturalised agency (tornadoes, dust devils, and digital computers[75]) while still allowing for certain artificial and many exobiological systems to qualify.

Third, the motile oil droplet exemplifies what theoretical physicist Karen Barad calls 'the extraordinary liveliness of the world' (Barad 2007: 91) that titles this chapter. In her book

[72] What I mean is that eschewing reference to materiality is not just associated with throwback or otherwise revamped cognitivism but can even be found in contexts that have long emphasised the importance of embodiment and corporeality.

[73] I reviewed some shortcomings of autopoiesis in Chapter 1 of this work.

[74] Let me unpack this latter concern a bit, as it is something of an elephant in the room with the specific materialist bent I have assumed in this work. Arguably the main reason those frameworks are neutral with respect to materiality is to attain a kind of theoretical *generality* that allows them to cover a range of possible systems: from artificial and constructed ones, to potential exobiological ones that have completely different architectures from those on Earth. That is, eliding materiality is meant to avoid concerns of *biochauvinism*, a concern we have already reviewed. However, one of the striking things about AMP is that it better specifies the kinds of proto-agential and agential systems that emerge naturally without being committed to a specific *kind* of matter. The AMP focus is on materiality and the potentialities, dispositions, and affordances certain material organisations afford or purchase a system—we can place boundaries on the 'space of possible agents' without thereby implicating ourselves in chauvinistic or partial views of those systems.

[75] The case of digital computers is more complicated than the others, especially since the fields of A.I. and M.L. have long moved past many scenarios involving purely digital computation. We will explore these cases more in the penultimate chapter of this work.

Meeting the Universe Halfway, Barad explores the agential interactions involved both in *diffraction*, a physical process that identifies how waves propagate and interact with physical media or barriers (such as in the famous double-split experiment), and how processes of measurement represent an interaction between socially constructed, experimentally specified set ups and the phenomena to be investigated—such that the thing being measured is co-determined by the measuring process itself. Her point in identifying this is to highlight the interactive dynamics that come with physical processes—processes themselves which are incessantly active, interactive, and therefore relational. It is this fundamental activity that leads Barad to posit, in contrast to a more entrenched view of materiality and physicality, an 'extraordinary liveliness of the world' that better animates an otherwise passive and brute view of nature.

My goal here is to not necessarily advocate for this wider understanding of physics—although I gestured at something close to it in the previous chapter and it compliments directly the view presented herein—but instead to identify a special case of this "extraordinary liveliness": one in which active matter represents an augmentation and amplification of a more general and fundamental activity; and that, once in the form of ordered, self-organizing dynamics, can represent a step on the way to the unique forms of biological activity and agency we find in the more 'scaled up' domains of biology and cognitive science (cf. Levin 2019). Properly speaking, then, there would not really be a 'passive' materiality that conforms with the philosophical materialism of a passive and inert world. Be that as it may, the designation of active matter still holds significance insofar as it identifies the processes by which materiality can be swept up in flows of energy to maintain and generate novelty and complexity. The case of the motile oil droplet provides a toybox example of this and helps to get a grip on why furnishing an active matter approach contributes to our understanding of naturalising forms of agency. To see why, it will help to dovetail the discussion of this and the previous chapter to outline what in philosophy is called a *relational ontology*. Drawing out the significance of active matter and outlining this relational ontology will prove insightful because one of the main arguments in this work is that agency, on any naturalised account worth its salt, will be able to identify not only how agents have causal influence on the world, but also how that influence itself is inextricable with the networks and processes of interactions in which it finds itself enmeshed—a discussion that we will turn to next.

4.4 A Relational Ontology for Naturalising Agency

In order to proceed with the naturalisation strategy I have suggested in this work, it is now important to pivot towards relationality and interactivity as robust ontological categories that do important philosophical work. To see why, it will help to briefly revisit the emergence literature. Emergence and relations often go hand-in-glove both in a positive and negative manner, meaning that furnishing an account of emergence often invokes an ontological category of relations; and antirealist attitudes towards emergence often vitiates relationality ontologically.[76] Seager (2012: 183), for instance, describes emergence in terms of levels of reality and that these levels are relational. For Seager, these 'high-level descriptions' apply not only to the topic of minds and consciousness—the subject of his book—but to thermodynamical, informational, evolutionary, and psychological phenomena, and through to 'more concrete structures' such as tectonic plates and chemical kinds. Crucially, though, "From the viewpoint of the SPW [Scientific Picture of the World] it is evident that all these features are nothing more than patterns of events which are explanatorily salient to us" (ibid.). These patterns—again, ranging from the agential and sentential to the thermodynamical and psychological—are "rather like the clouds that appear in the shape of animals as we watch them" (ibid.). This transforms into what Seager calls a "generalised epiphenomenalism"—these features only have significance in *relation* to meaning-making creatures like us, and, like all relations, are causally impotent and explanatorily redundant with respect to a scientific world view.

Speaking further to this point, the ability to draw relations tends to be framed as a product of scientific theories drawing the lines between dots (facts) that put things into relief for beings like us—highlighting the epiphenomenal, mind-dependent nature of emergence. As Taylor writes, the "best way to understand the concept of emergence is as the unavailability of a certain kind of scientific explanation for an observer or observers" (2015: 654). This picks up on comments from Kim we saw earlier in this work, namely, "we interpret the hierarchical levels as levels of concepts and descriptions, or levels within our representational apparatuses, rather than levels of properties and phenomena in the world" (1999**).** The views implied in these anti-realist positions is some kind of physicalism wherein substances, essences, or "bits of matter and their aggregates in space-time exhaust the contents of the world" (ibid.). Rosenberg, as we have seen, also eschews primacy of relations because they imply bringing

[76] In this regard, reductionism tends to imply an elimination of relationality—what matters to the reductionist here is to show that relations are epiphenomena of brute, fundamental physical facts (See esp. Seager 2012, Ch. 8).

phenomena with causal properties into the world that do not inhere within the entities that constitute them, "something no mechanist… should want to countenance" (2020: 61).

We could keep going. Arguably, these attitudes towards relationality and emergence form the backbone of much past and present Western Metaphysics, which has tended to have a penchant for substances and essences at the cost of identifying how those essences come to be wrapped up in a world of relations (Dupre 2014). I have already canvassed reasons why we need not think the 'scientific picture of the world' countenances the austere reductionism and anti-realist positions often associated with it[77]—and the sciences of CMP and AMP have done more than others to loosen the grip this view of nature has on philosophical discussions. And nowhere is this truer than revising our understanding of 'bits of matter and their aggregates' as exhausting the world and obviating the need for invoking relations as a robust ontological category. Following Bickhard & Campbell, we must reject "a deeply entrenched assumption… that what basically exists are things (entities, substances). Our best physics tells us that there are no basic particulars, only fields in process. We need an ontology that gives priority to organization, which is inherently relational" (2011: 33).

Discussions of 'relationality' as just explored often appear overly philosophical in expression and, in a discussion that explores CMP and AMP, might appear as unnecessary. To reiterate, this turn towards ontology is important because it captures a set of deeply held assumptions and attitudes towards emergence. In other words, because self-organization is inherently an emergent phenomena it implies a degree of relationality that is not adequately captured within a 'substance ontology'. The intent of this section, then, can be seen as repositioning certain ontological categories (namely, relationality) for philosophical discussion that could 'demystify' emergence from within philosophical naturalism, which, as the above adumbrated literature shows, has overwhelmingly platformed a view that takes relations as 'observer dependent' and, in an important sense, 'epiphenomenal' or somehow 'less real'.

With this in mind, I submit that AMP (in tandem with the various other scientific domains explored in this work) are furnishing the resources for better ontologising in this regard—not just for a better understanding of the emergence of complexity and novelty in physics and

[77] For what it is worth, Seager acknowledges that the scientific picture of the world he advocates is a *metaphysical* and not *epistemological* view (2012).

biology; but for the ambitions of the current work on naturalised forms of agency.[78] Recently, Nick Brancazio has advocated for a similar position that takes relations (interactions) as a distinct and robust ontological category that does service for our understanding of agency: "The study of active matter systems demonstrates the utility of using a minimal approach to agency in studying interactive agential dynamics in more complex systems by treating interactions as an ontological category" (2022). To get a grip on this, and how it will eventuate in an account relevant to the discussion of agency, it will help to return to the Vicsek model that we briefly referenced in the previous section on the motile oil droplet.

The Vicsek model is intended to capture the phenomenology of collective and flocking behaviour of active particles. First introduced by Tamàs Vicsek and collegaues (1995), it was originally intended as a minimal representation of certain types of emergent phenomena observed in groups of animals or particles. In the model, individual particles move in a two-dimensional space and align their velocities with those of their neighbour within a certain radius. The model assumes that each agent's velocity is updated at discrete time steps based on average velocity of its neighbours, with some added noise to account for fluctuations. These simplifying assumptions or rules captures the tendency of particles to align their motion with neighbouring particles and precipitates in collective motion.

The Vicsek model illustrates how local interactions among particles can lead to the emergence of global patterns and coherent collective behaviour. By studying the model, researchers can explore the fundamental principles behind the emergence of order and self-organization in systems of self-propelled entities. Moreover, one key insight of the Vicsek model is that the alignment interaction alone can drive the system through a phase transition. At low noise levels, the particles align their velocities, which results in ordered motion. However, when the noise exceeds a critical value, the system transitions to a disordered, fluid-like state, where the particles move in random directions (Matsuda et al. 2019). As Brancazio writes, "The

[78] Again, pulling from Harold & Morowitz we see that major features of the biosphere, the evolutionary process, the fine-tuning of metabolisms, and the origination and diversification of species all depend on a radically interconnected biosphere characterised by the emergence of new orders and modes of constraining and channelling energy—absence of interconnection, even if it were thinkable, would simply miss what the biosphere and biology is all about. Consider this when they write, "We characterise life as a 'confederacy' of different sources of order, many of which we argue have independent origins within different domains of chemical or physical processes, or planetary conditions. The robustness of the full suite of living regularities results from a parallel appeal at many levels to boundary conditions and constraints of physical laws. The function that distinctively biotic dynamics performs uniquely is to *interconnect* [sic] these members of the confederacy into webs of mutual support and interdependence. A view of life as an integration of multiple disparate sources of order may explain how the emergence of life on Earth could have been at the same time an outcome of quite ordinary events, yet one that depends in detail on its planetary context" (2022: 14).

model… shows that flocking behaviours can arise from self-propelled particles observing very basic alignment rules. With just the trademark motility of individual units of active matter systems and a mechanism for alignment, Vicsek models demonstrate that a number of different spatiotemporal patterns of activity can be achieved with simple modulations of population density or slight alteration of the alignment mechanisms". What is important from the perspective of coordinated dynamics is the interactions between various active particles that, absent global coordination, can generate order and complexity in a relatively 'bottom up' manner.

While this might initially sound reductive in nature, it is meant to highlight the sheer amount of flexibility and plasticity bottom-up features alone have for generating form—when we combine these with the downward acting effects explored in the past two chapters, then it better conveys the emergentist picture I paint herein.[79] Crucially, the ability of these more 'basal' processes to generate form *in tandem with* emergent and contextual features (such as the mediating influence of bioelectricity, which we will explore shortly) all depends on the role that active particles *interacting* (due to alignment between those particles) in ways that generate complexity 'over and above' what the individual particles are capable.

4.5 Conclusion

I would like to conclude this chapter on a more general note, one that ties the considerations put forward herein to the discussion of Basal Cognition that opened this thesis. Recall that Basal Cognition deals with organisms that are thoroughly saturated in a predominantly 'physical' reality. That is, a reality in which the complexity of basal behaviours is imbricated with the complexity of physical and chemical processes on which they depend. My goal in identifying this behavioural-cum-physical nexus is to introduce a middle ground wherein biology and physics are in *conversation* with each other, reformulating the confrontational and antagonistic positioning of the two sciences that is often assumed (e.g., that physics in biology threatens a reduction of the latter to the former). By introducing a middle ground where biology and physics are in proximity, we also introduce organisms occupying these scales as needing to deal with problems that the relevant physical level presents. That is, the physics

[79] The point about reductionism here is pertinent because based on the previous discussion, it might have been thought that I am advocating for an unrealistic emergentist picture in which bottom-up assembly (self assembly) is obviated by the appearance of higher-order, organizational features. In fact, biological processes imply a complex intertwining of both 'passive' process of self—assembly and the active ones identified herein.

sets the problem space and to understand the problem space we must also understand that physicality.

In Basal Cognition, then, the *behaviour* and the *physicality* of these model systems cannot be dissociated—at least not without generating a lopsided vision of their behaviour. What I am proposing here, then, is that in trying to understand the supposedly cognitive nature of, e.g., *E. coli* or *paramecium* we are *also* asking questions about how the supposedly cognitive (or agentic) behaviour relates to those *physical problem spaces* and the *materiality* as interacting with, and *inhering within*, those physical spaces—that is, within the constraints that the tyranny of scales imposes. Introducing physical sciences techniques, concepts, and approaches into biology then is not meant to threaten the autonomy of the life sciences—even less is it intended to propose a kind of reduction of the latter to the former. Rather, the conversation between the two sciences is supposed to be bidirectional. From physics to biology: it highlights the instances in which allegedly cognitive behaviours are bounded and constrained by the material affordances of their cognitive endowment *in interaction with* the problem space that the physicality introduces (e.g., the tyranny of scales problem discussed previously). From biology to physics: that concepts normally confined and restricted to the life sciences—top-down causation, emergence, and perhaps agency—have conceptual purchase in the physical sciences. In effect, then, what we see in the dialogue between the two disciplines is an *elevation* of mechanism to the status of the organism, rather than organism reduced to the level of mechanism.

To this end, Stuart Newman has commented on the importance of understanding this physical nestedness: "The forms of the earliest multicellular organisms… were more like certain materials of the non-living world than are the forms of their modern, highly evolved counterparts, and that they were certainly moulded by their physical environments to a much greater extent than contemporary organisms" (2003: 178). What Newman captures here—specifically with the language of moulding—is the idea I have tried to capture in this chapter with 'tyranny of scales': such that understanding organismic processes benefits from an understanding of physical sciences methodologies, specifically the non-reductive approaches provided in Active and Condensed Matter Physics. But introducing the moulding effects present at scale—and drawing a line between these physical behaviours and the agential dimension of basal organisms—is meant to also serve the purpose of highlighting the relevance of *physical* properties of the organism for its cognition: features such as shape and composition, as well as aggregative properties that emerge from self-organization, have a more determinative role in their behaviour, which reinforces the point that their cognition—if we choose to call it that—cannot be dissociated from these more mundane properties. This is a point that Thompson expressed in *On Growth and Form* when he writes:

> [The world of man] is ruled by gravitation… [But in the] world where bacillus lives, gravitation is forgotten, and the viscosity of the liquid, the resistance defined by Stokes' law, the molecular shocks of the Brownian movement, doubtless also the electric charges of the ionized medium, make up the physical environment and have their potent and immediate influence on the organism. The predominant factors are no longer those of our scale; we have come to the edge of a world of which we have no experience, and where all our preconceptions must be recast. (Thompson 1942: 77).

The point that will be increasingly relevant in the remainder of this work, as it is still quite common to see researchers reference molecular activities with machine analogues—machines that occupy the scale *more familiar* to us and thus shot through and through with preconceptions about the behaviours of materials. Once we descend to the level of cells, assumptions often associated with 'mechanism' simply do not hold the weight they once did.

Chapter 5

A Road Map for Naturalising Agency

> You cannot even think of an organism… without taking into account what variously and rather loosely is called adaptiveness, purposiveness, goal-seeking, and the like
> - Karl Ludwig von Bertalanffy, *General Systems Theory*, 1969.
>
> Organisms are fundamentally purposive entities, and [yet] biologists have an animadversion to purpose.
> - Denis Walsh, *Organisms, Agency, and Evolution*, 2015

5.1 A Philosophy of Nature for a Relational Understanding of Agency

In Chapter 2, we saw that a critical insight of Kant's teleological formulation in the *Third Critique*—and the fecund and prolific discourse that followed in its wake—was that *nature* appears differently depending on the methodologies we develop to investigate it; the questions we ask to frame it and the technologies at our disposal for furthering insight into it. One of the main goals of this thesis has been to show how questions and concerns of agency as a robust ontological category stem from similar considerations and often boil down to philosophical presuppositions for how we understand "things in the broadest possible sense of the term hang[ing] together in the broadest possible sense of the term", as Sellars puts it (1962). What this comes down to is a kind of naturalism worth having—one that is not arbitrarily or dogmatically maintained to force nature to conform to the image we have of it, but concertedly investigates the physical, biological, and mind sciences in their capacious and expansive purview. If naturalism is, roughly put, the metatheoretical commitment that we defer to the natural sciences when we ontologise (that is, when we posit what is or is not in the world), then much comes down on what we might call, following Sellars, the scientific image.[80] We have already seen how for many researchers (Kim 1995; Taylor 2012; Seager 2012; Rosenberg 2014; Craver 2014), the scientific image vitiates the need for incorporating agency as a causally robust category when understanding the physical and biological sciences. At best, what we have is a kind of *as if* teleology that captures "patterns of events which are

[80]In the previous chapter we saw that many naturalists (seeking to vitiate relationality) also advocate for a Scientific World View (SWV) that determines the legitimacy and limits of philosophising.

explanatorily salient to us" (Seager 2012: 133). And what this amounts to is the erection and maintenance of a divide between this 'scientific image' and a view of nature that includes intentional, directional, agential, and goal-directed features of agents and subjects (ourselves included).

We have already seen how the division can have innocuous and nefarious forms depending on how we flesh them out[81]—here, my point in revisiting this is to merely highlight that language and explanatory frameworks, ironically, are imputed the majority of the power when it comes down to identifying psychological features of the world: from a more minimal agency explored in this work to the rich psychological capacities that characterise Metazoans (Godfrey Smith 2018). That is to say, we construe agential ascriptions as psychologistic in nature because we work with a scientific image of matter and materialism that depicts materiality as devoid of the features or causal structures that would be required for its 'naturalisation'. The main point of this thesis has been to argue that, at least vis-à-vis agency, the "notion of matter actually found in the practice of science has nothing in common with the matter of materialists" (Dewey 1958: 74). Again, in Chapter 2 we saw that in order for intrinsic purposiveness to attain proper scientific status it was necessary to identify an immanent understanding of teleology—such that matter can contain an 'inner animating principle', something which was anathema to Kant and which still trails a whiff of odium when spoken of in philosophy of mind and science today. In fact, even in domains where the materiality of autonomy, intrinsic purposiveness, and agency is seen as important for their legitimation, it is still possible to find scepticism regarding the importance of matter itself in understanding what sets these systems apart from other, non-agential systems—such as tornadoes or digital computers. I submit that, following Karen Barad, what this evidences a 'deep mistrust of matter': "There is an important sense in which the only thing that doesn't seem to matter anymore is matter" (2007: 132-133).

Barad's goal in her book, *Meeting the Universe Halfway*, is to reintroduce materiality as an important element in the construction of social and human reality—a project which seems orthogonal to the one pursued herein but which, in key respects, recapitulates the worries I identified with naturalism. Writing against a tendency to impute the social construction of reality to discursive, linguistic, and cultural practices while leaving to the side the inherent activity and potentialities of matter, Barad remarks: "Why are language and culture granted

[81] For Sellars part, the two were mutually enriching and necessarily in communication.

their own agency and historicity, while matter is figured as passive and immutable or at best inherits a potential for change derivatively from language and culture?" (ibid.); that is, a view that displays a "deep mistrust of matter, holding it off at a distance, figuring it as passive, immutable, and mute, in need of the mark of an external force… to complete it" (ibid.). It is in this vein that she promotes an agential realism based on a *new materialism* that captures the inherent interactivity and interpenetrations between meaning-making subjects and material processes of emergence and diffraction (ibid.: 139). Given that some form of materialism or naturalism has been the default position in analytic philosophy for decades, the attitude towards this new materialism in the social sciences might be seen with some amusement—a kind of "welcome to the party". But this would be to understate the conceptual import of new materialism and to overstate the scientific status of contemporary philosophical materialism. Barad speaks further to this point when she writes: "The primary ontological unit [from the perspective of agential realism] is not independent objects with inherent boundaries and properties but rather *phenomena*. … Phenomena are ontologically primitive relations—relations without pre-existing relata. The notion of *intra-action* (in contrast to the usual 'interaction', which presumes the prior existence of independent entities or relata) represents a profound conceptual shift" (2007: 139)—a conceptual shift I think is needed for the ways we theorise agency.

This picks up on a point that closed the previous chapter: namely, that what matters from the perspective of naturalising agency—furnishing the agential realism Barad speaks of—is initiating a conceptual shift away from the kind of exteriority characterising 'bits of matter and their aggregates' (and an associated causal model) to one in which relations and their 'intra-actions' are factored in and, indeed, what becomes primary when accounting for the agential status of a self-organizing system (and associated components).[82] The distinctive move I make herein is to tie insights of the material sciences to these ontological considerations to show how a revamped understanding of materiality helps clarify the interpenetrations that exist between agential or proto-agential activities and physical ones, picking up on the idea that "the mechanism that underlies the formation of ordered phases is mutually reinforcing interaction among many small-scale, individually stochastic degrees of freedom known as collective or cooperative effects" (Smith & Morowitz 2022: 24). It is now time to tie these considerations more explicitly to the question of intrinsic purposiveness which guides this argument—the very 'animadversion' Denis Walsh speaks of when he

[82] There is a tricky question here because at times the literature posits agency at a cellular level and above, while others agential interactions refers to how active particles interrelate.

remarks: "Organisms are fundamentally purposive entities, and [yet] biologists have an animadversion to purpose" (2015).

What I would like to propose is that agency represents a higher-scale ordering of lower-level events that operates in accordance with a goal—where that goal is something immanent to the self-organizing dynamics of the system. Crucially, this depends on the thermodynamic ordering of the system such that its very activity keeps it in the out-of-equilibrium state. Agency in this regard is the gross behaviour of the self-organizing system that constrains matter and energy flux to generate a more ordered state and further that ordered state into the future. In the language of CMP: "Whereas disorder is generic, sufficiently coherent interactions to produce order are rare, and for this reason the plausible mechanisms to produce any form of order that we observe as robust and stable are limited" (Harold & Morowitz 2022: 20). While nothing in this description implies agency, I submit that the distinguishing feature separating intrinsically purposive systems (the key distinguishing qualifier for agency) from others is the immanence of their dynamics to their materiality. From the perspective of the biogenic approach, this suggests that this is the hallmark of a kind of bio-agency requisite for work in Basal Cognition. Moreover, it foregrounds the fact that the essential ingredients of agency are its relational and strongly emergent (in the minimal sense of Chapter 4) dimensions—dimensions that mark out agents as ontologically robust. For now, I bracket considerations of non-biological agency (which might not exhibit the hallmarks of intrinsic purposiveness) that might be found in Artificial Intelligence and advanced Robotics systems, as well as agency of communities and non-contiguous eusociality. I follow Godfrey-Smith in suggesting that the natural emergence of such systems absent the fine-grained details of metabolism and biological systems is improbable, but that once such systems (i.e., humans) exist their appearance might be possible. [83] This suggestion is picked up in the following chapter and not considered further herein.

To make concrete the suggestion that intrinsic purposiveness is a naturalised form of self-organization antecedent to the kind of agency we find in biological systems, I pick up on recent work by Potter & Mitchell (2022). Recall from Chapter 4 that "to justify agent causation… systems need to exhibit a causal power that irreducibly inheres at the level of the

[83] George Deane (personal correspondence) has suggested that an interesting limitation of contemporary Large Language Models, or LLMs, is that their power is derivative or at least heavily contingent upon our human agency to compile the data and training sets (to say nothing of their maintenance). This suggests that a more robust kind of agency might be required for the systems-level autonomy generally aspired to in the pursuit of Artificial General Intelligence (AGI).

whole system (i.e., it is nonreductive) while still maintaining that the causal power is instantiated in, or realised by, the system's physical constituents (i.e., it obeys physicalism)" (2022: 1)—something Timothy O'Connor calls an 'ontologically primitive causal power', one that is causally dependent on microphysical events and structural states yet is something ontologically primitive nonetheless (2016: 195). Relating this to Barad's point above, what is critical with respect to agent causation is establishing how the emergent feature is inextricable with its material basis while at the same time maintaining a causal locus that is not 'pushed around' by its constituents. This chapter, then, represents a roadmap for how we can demystify the physical basis of agency that does not imply any special 'stuff' that sets biological agency apart (i.e., it does not commit an error of biochauvinism) from other modes of material becoming in the world, while also vitiating reductionist worries about the causal status of the emergent. To do so, I explore the following eight features as jointly necessary for a minimal form of naturalised agency: (1) Thermoydnamic Autonomy; (2) Persistence; (3) Endogenous Activity; (4) Holistic Integration; (5) Lower-scale Indeterminacy; (6) Some version of robustness and 'multiple realizability'; (7) Historicity; and (8) Normativity.[84]

The structure of the remainder of this chapter will be as follows: In section (6.2) I survey some of the literature on agency, focusing on the anthropogenic approach, which delineates agency along a more anthropocentric dimension, and contrast this with the biogenic approach. Section (6.3) proceeds to the 'roadmap' of naturalising intrinsic purposiveness. Section (6.4) ties the discussion together and puts the further emphasis on materiality that I identify as a missing piece from much of the discussion in naturalising debates. Section (6.5) concludes with some observations on how the present discussion is not an abstracted view independent of the material sciences and sciences of minds, but is actively being operationalised in the domain of soft robotics and certain approaches in bio-inspired Artificial Intelligence to arrive at more adaptive and flexible architectures that might be required for some of the grander ambitions of these fields, e.g., Artificial General Intelligence [AGI]. Before proceeding to this, however I want to make a quick note on what this chapter is and is not about, as it will help to simultaneously situate this work within the wider discourse of naturalised approaches to agency while also differentiating the strategy herein from those projects.

[84] This list was first proposed by Potter & Mitchell and I take it as suggestive of a roadmap for naturalising teleology

5.1.1 The Wider Debate on Naturalising Agency

> Because of the widespread mechanistic distrust concerning the notion of purposiveness, we do not possess the conceptual and mathematical tools required to appropriately incorporate true organismic agency into models of evolutionary dynamics. This is why we'd rather pretend the phenomenon does not exist, rather than taking it seriously.
>
> - Johannes Jager, 2021.

The work and research that went into this thesis is based on a pedigreed and established tradition in the life and mind sciences, many great examples of which I have overviewed. Evidently, the present work is situated within the wider web of naturalising the purposive and agential modalities of organisms via the notions of recursive causality and the thermodynamic regimes associated with it. These are systems that *must* act in a denumerable set of ways in order to be the kind of systems they are. Other systems may transiently exist out of equilibrium but cannot transduce or constrain energy for the end of their own preservation. Emblematic of this perspective is autopoiesis (Thompson et al. 1993), certain species of 4E cognition (Barandiaran et al. 2016), and the organizational account of biological autonomy (Moreno & Mossio 2015). In addition to these projects, with their explicit focus on minimal systems and modes of interaction, my project also owes a debt to the philosophical debates on naturalised phenomenology—crucially, because the very 'animadversion' that underpins the attitudes towards teleology in the life sciences underpins a wider worldview and scientific image that dispels consciousness, sentience, and experience from the armature of legitimate ontological categories. Indeed, the two are so interrelated that a fuller treatment would highlight how their respective maligned status (at least, historically, though each have of course gone through their renaissance; see Ginsburg & Jablonka (2019)) stem from a similar source: namely, the construal of physics and materiality, the 'stuff' supposedly investigated by the natural sciences, as devoid of the propensities and potentialities involved in the emergence of mind and life—where the latter is epiphenomenal to the former. My project, then, benefits from the wealth of literature that has jointly encouraged reintroducing agency, mind, and sentience as defensible categories in scientific practice.

Be that as it may, the current project differs in important ways from the above—though with crucial points of contact when and where necessary (as I have tried to sketch at appropriate moments during the argument). There are two differences, in order of importance, that I see the present account as providing that are either not fully articulated in the above theories: or, worse, actively missing. The first is the kind of 'metatheoretical' project—the *philosophy of nature*—associated with interrogating the nexus of ideas that configures the debates in which we then theorise. As we saw in the previous chapters, background or pre-theoretical understandings of causation, mechanism, and emergence tacitly or explicitly mediate the kinds of inferences we draw and the conclusions at which we arrive. Revisiting these assumptions accordingly reformulates the strategies available to researchers and theoreticians for better grappling with the physicality of agential systems and the philosophical implications of their renewed 'reality'. The second part of the project that represents a difference in emphasis with previous proposals is its more explicit focus on the materiality of the energetics that have otherwise been used as the underpinning of naturalised approaches to teleology. While this chapter keeps the discussion at a relatively high level of abstraction, the previous two chapters and the following one (Chapter 6: 'Soft Robots and Living Machines') puts into focus the material sciences with their concerted investigation into the potentialities afforded by materiality and its diverse configurations and disposition for self-organizing.

This chapter, then, is more focused on the ways in which we individuate a kind of minimal agency that will be for a large part restricted to the biological domain (that is, bio-agency). I will only introduce concrete examples when necessary, and I am not interested in cataloguing various instances or examples of agency in the biological world—nor the related 'floor and ceiling' problem associated with identifying the most minimal agent at the bottom and where, if ever, it breaks down as it scales up (e.g., from multicellularity to eusociality to tool-using and culture making beings like ourselves (i.e., can a washing machine be said to be agentic due to its participation and involvement in a web of utility and meaning dependent on human agents?)). It is important to reiterate at each point that in making this a 'naturalising' argument it is primarily a demystifying project, one concerned with how we arrived at a picture of the world that is missing key features including agency and minds. The nature, robustness, breakdown, scaling-up, breadth, etc. of agency in its empirical context forms a set of highly interesting, complex questions that will require interdisciplinary engagement from across a range of disciplines and I do not feign concrete proposals in this chapter. With this in mind, I will turn to two different notions of agency: an anthropogenic and a biogenic approach, where the former begins from a higher-level understanding of agency and works down; and the latter establishes a more minimal account of agency on which my approach depends.

5.2 What Is Agency?

The reader will have certainly noticed I have been cagey on providing a definition of agency. I do not wish to contribute to the already expansive range of definitions in the literature, and instead in this section I will overview a series of definitions or approaches I take as more or less amenable to the present perspective and then, in turn, show how the approach I have opted for here tries to substantiate or 'materialise' them. This section introduces an 'anthropogenic' account of agency before proceeding to more 'minimal' versions explored within the biogenic approach.

When one thinks of agency colloquially, one might think of what is sometimes called 'full-blooded agency' (Ferrero 2023)—applying to rational agents with privileged epistemic insight into their own actions. This might then come to include a moral dimension, such as accountability, and explicability, such as being able to explain and assess the grounds for one's actions (see Ferrero 2023 for an overview of this form of agency). This more robust—'full blooded'—account of agency is often the assumed one in the literature and its ascription to anything beyond the human domain generates much of the 'animadversion' that Welsh spoke of above. It could be called, following Pamela Lyon, the *anthropogenic approach*: "assume human cognition [or agency] as the paradigm and work 'down' to a more general explanatory concept" (Lyon 2006: 11). Exemplary of more anthropogenic approaches as applied to evolutionary biology would be that of Samir Okasha (2016).

Okasha's work on agency in evolution (i.e., organisms *as* agents) comes down to a 'key concept in human agency' (2016: 11). This concept builds off that of Kennett & Matthews (2003: 307), who write that "Effective agency … requires a *unity-of purpose* both at a time, in order that we may eliminate conflict among our motives and do one thing rather than another, and over time, because many of the things we do form part of longer-term projects and make sense only in the light of these projects and plans". Unity-of-purpose, then, exhibits two of the crowning features of human agency: (1) "a person's goals and intentions should cohere with each other in the sense of being mutually reinforcing, or at least not clearly inconsistent"; and, (2) "their actions should tend to further their goals, that is, they should be instrumentally rational" (Okasha 2016: 19). With the anthropogenic unity-of-purpose, Okasha then concludes that "an analogous unity-of-purpose is necessary in order for an evolved organism to be treated as agent-like" (ibid.). While Okasha insists (2016: 20) that talk of agents is not 'idle metaphor' and "that agential thinking in biology… can be a powerful tool for understanding adaptation" (2016: 10), it is sometimes not clear whether the unity-of-purpose idea, as specified along anthropogenic conditions, can be applied to the diversity of life. Indeed, it is not even clear that such a stringent requirement applies to us. Not just in psychiatrically interesting cases do we see unity-of-purpose break down, but perhaps within so-called neurotypical instances it might be questionable that such a definition holds. Moreover, the contrast to this unity-of-purpose account of agency is "the most minimal notion of agent… [which] is simply that of an entity that *does* something, or behaves" (ibid.: 12).

I submit that this 'most minimal notion of agent' is not an agent at all—and that if the options available to us are stretched between a 'mere doing' and a full-blooded agent, that it is difficult to account for or accommodate the distinctive modes of activity and agency expressed in the biological world. Lastly, despite defending the utility of agent-talk in biology, Okasha's approach often teeters on the 'as if' teleology emblematic of a Kantian approach—which could also be said to begin from the human—and 20th century biology – the biggest difference being that he defends an ineliminable role for agency parlance in developmental and evolutionary biology. But what I will now argue is that the *biogenic* approach furnishes an account of agency that is more defensibly called 'minimal' (i.e., it is not a 'mere doing'); and that it can more robustly apply to a wider domain of creatures and contribute to our understanding to more complex forms of agency and mind found in humans and other Metazoans.[85]

5.2.1 Minimal Agency

As we saw in the Introduction to this work, the biogenic approach takes a different tact when investigating several concepts and capacities normally restricted to humans, namely, cognitive capacities and agency. In contrast, then, to a view of agency that begins from the human and works 'down', the biogenic approach "starts with the facts of biology as the basis for theorising and works 'up' to the human case by asking psychological questions as if they were biological ones" (Lyon 2006: 11). Or, as Lyon writes—this time with Michael Levin, Fred Keijzer, and Detlev Arendt, all of whom have independently ventured versions of the thesis—"the cognitive sciences should at last join the rest of the life sciences in the way they

[85] The extent to which the biogenic and anthropogenic approach relate and complement each other will not be explored further in this work.

approach their quarry" (2021: 1). What this comes down to is "three essential steps" (ibid.): fist, begin with the smallest and simplest organisms that exhibit the capacity or function of interest. Second, in those organisms "*identify principles* from observed and measured patterns of genetic, epigenetic and behavioural interactions" (ibid.). Third is understanding ways in which these principles *scale up* to more complex organisms and then observe where the crucial points of (dis)similarity reside. Jointly, they suggest that this could represent a 'Copernican revolution' in our understanding of biological organization and behaviour (as we explored in Chapter 1). One domain in which this approach could provide clarity, as we have already suggested, is that of agency—and when I speak of starting with the *smallest and simplest* organisms[86] (see Figure 1) that display the phenomenon of interest, it is here that we speak of *minimality* and minimal agency (Brandiaran et al. 2009; Moreno 2018; van Hateren 2023).

There is a long history in the cognitive sciences and philosophy of mind exploring minimal instantiations of agency—at least since Maturana & Varela's formulation of Autopoiesis (1979). Writing more recently, Weber & Varela (2002) argue that autopoiesis naturalises intrinsic teleology due to its establishment of mutual generative relations between organismic components—that is, between parts and wholes—which is a throughline to Kant's concept of a *Naturzweck*, or natural purpose (cf. Di Paolo 2005). Weber & Varela, armed with this concept of natural purpose and autopoiesis, go on to argue that this continuous auto-production (hence, autopoiesis, its etymological roots) leads to an "insaturation of a point of view" (2002: 116) and ensures a form of individuality or autonomy that "*ipso facto* [is] a locus of sensation and agency" (2002: 117). This is the general strategy for most naturalising projects on minimal agency, beginning with the substantiation of Kant's intrinsic purposiveness before proceeding to show the ramifications that follow from it (see Chapter 2; Wolfe 2022). While I certainly see this strategy as promising and follow some of it for my approach that comes at the end of this chapter, sticking with the Kantian formulation without targeting the wider nexus of ideas dogging teleology may pose something of a difficulty for naturalising agency, as I have tried to show.

For now, it is important to note that minimal agency has diversified much since the formulation of autopoiesis and assumes many different names: such as biological agency (Kauffman & Clayton 2006); proto-agency (Fulda 2016); and biological autonomy (Moreno &

[86] From hereon, we will say 'systems' to keep open the possibility that non-living, exobiological, non-biological, or artefactual-designed systems can also be said to exhibit some kind of agency.

Mossio 2015). While different in their specific formulation, this literature collectively posits that the causal regimes supported in far-from-equilibrium, self-organizing systems with an internal space for complex chemistry (e.g., a metabolism) instantiates a minimal form of agency because it simultaneously introduces new features: namely, normativity—the ability of things to go right and wrong for the system. While the literature can get quite complicated and crosstalk is rife, discussions of minimal agency are often inextricable from questions of the appearance of normativity and teleology in nature. Collectively, accounts of this minimal form of agency eventuate in a standard definition revolving around perdurance of system identity due to active self-maintenance and continuous (within bounds) productions of parts that come to perpetuate a system's existence. In this regard, a recent definition that Sonia Sultan and colleagues (2022) have offered is fairly representative of the philosophical and theoretical motivation here: "[Agency] is the capacity of a system to participate in its own persistence, maintenance, and function by regulating its own structures and activities in response to the conditions it encounters". In a more sloganistic expression, agential systems are those that *act on their own behalf* (Kauffman & Clayton 2006).[87]

The inclusion of basal organisms within the remit of cognitive science is meant to show that their behaviour and distinctive decision-making capacities are "best regarded within the framework of cognition rather than mechanism" (Ball 2023: 7). Minimal *agency*, then, under the biogenic approach picks out a set of behavioural capacities, thermodynamic tendencies, and interiority (an insaturation of a point of view) that is believed to be of more general distribution in the biological world. Intriguingly, and as a corollary, there have been suggestions that this kind of minimal agency might even be exhibited in non-biological systems, such as we briefly explored in the previous chapter (for review, see Moreno 2018). With this in mind, we may now turn to the question: What would it mean (or look like) to *naturalise* agency?

[87] There are strong internalist overtones to lots of the discussion on autonomy and agency here, so it is important to stress that autonomy does not imply exclusion or being cut off from the environment (as earlier iterations of autopoiesis seemed to imply; see Di Paolo 2005). Indeed, because many of these proposals have explicit thermodynamic motivations, openness and continuous exchange with the environment is often built into the formulation of identity here. As van Hateren (2023: 92) writes, "[Autonomy] does not mean independence, but merely the capacity to influence the conditions under which the system exists. An autonomous system may still strongly interact with its environment … but it is not just pushed around".

5.3 Naturalising Agent Causation: Eight Criteria

I have already suggested that to naturalise agent causation, systems must express a causal power that exists uniquely at the system level (i.e., it cannot be decomposed into smaller 'parts' without losing something unique to the systems-level), while still adhering to naturalism in that the causal power is instantiated in, or realised by, the systems constituents (Potter & Mitchell 2022: 1). A great deal of the present work has set out to show the possibility of this, and now is the time to more concretely present how this could be accomplished. We have already seen that emergence—as explored in Condensed and Active Matter Physics—need not adopt the stringent eliminativism regarding emergence that is commonly assumed in philosophical physicalism. We have also briefly explored the role that emergent features play in constraining the activities of lower-level happenings—both thermodynamically, in terms of constraining energy consumption and channelling (bio)energetic processes to produce work that perpetuates the system—and in terms of the structural features that are ineliminably part of higher-orders of scale and have associated qualities that, quite importantly, have 'no reality' at scales below them—such as with stiffness and stress in the ECM. Having worked to dislodge the incompatibility, then, between these concepts and a sufficiently updated philosophical physicalism, I now explore in more detail what else is required for the above adumbrated 'minimal agency' to be conceptually coherent and scientifically tenable.

Following Potter & Mitchell (ibid.), naturalising agency can be boiled down to a core list of ingredients that are jointly necessary for naturalised agency: (1) Thermoydnamic Autonomy; (2) Persistence; (3) Endogenous Activity; (4) Holistic Integration; (5) Lower-scale Indeterminacy; (6) Some version of robustness and 'multiple realizability'; (7) Historicity; and (8) Normativity. This is a tall-order, to be sure, and each could (and has) taken its own paper to defend. I can offer only a sketch in this chapter, with my overall goal being more of a philosophy of nature reworking of the assumptions and prejudices that lead to a broad prohibition on agency and teleology in scientific practice. Be that as it may, a more concrete strategy is now required, and it is in this section where I dedicate the most space to how the concept of agency can be operationalised to display its nonreductive nature while still being consistent with our naturalistic approaches to a domain of study. Let us take each in turn.[88]

[88] The section to follow will be a modification of structure and content from Mitchell & Potter (2022). Although their treatment is already sophisticated, this section provides more information and clarification where it is needed. In any case, the (8) features they have identified have, indeed, been taken as quite constitutive of accounts of naturalised

1. Thermodynamic Autonomy

We cannot ignore the thermodynamic situation of a system when it comes to determining the agentic status of that system. This is the lesson I tried to stress in the previous chapter. By explicitly looking at the materials and units that come to form a system and how those units can independently—but through collective coordination—transform energy into mechanical output, thereby continuously reproducing conditions of existence required for self-maintenance, I submit we can more properly distinguish between famous edge cases (such as tornadoes) and systems that are agential (e.g., organisms).[89] Arguably, the emphasis on thermodynamics is the constitutive component of almost all accounts of biological agency. Moreno & Mossio, for instance, write: "Autonomy… is essentially grounded in thermodynamics. … [These] are dissipative systems dealing in a constitutive way with a thermodynamic flow that traverses them" (2015: 6). That is, establishing the minimal level of *autonomy* (not agency per se) requires a thermodynamic grounding where dissociations between 'matter' and 'form' do not make much sense.[90] "Rather, an adequate understanding of biological organisation should reconcile form and matter, insofar as many fundamental features of biological organisation make sense… only in relation to the conditions of their realisation in nature" (Moreno & Mossio 2015: 7).[91]

---

agency—though some features receive more attention than others. For my purposes, as I will stress at the end of this chapter, I am mostly concerned with (1)-(4) and (8).

[89] It is important to stress again that the thermodynamic situation is not sufficient to establish agency, as this section has already made clear: to be an agential system on my account, it must have some kind of intrinsic purposiveness with inherently endogenous activity. As far as we know right now, all such systems are *active matter systems*: meaning they are composed of, or constellated out of, active units, parts, and materials that can endogenously transform that energy for continued self-maintenance. Understanding the thermodynamic situation, then, is quite critical for understanding whether something is a 'mere doing' or exhibits minimal agency.

[90] A quite common contention in the history of the life and mind sciences is that the materiality and energetic situation of a system is negligible for understanding its gross organization and behavioural profile. Researchers have often eschewed such details in favour of simpler, more parsimonious computational models (see the 'algorithmic chemistry of Fontana (1992); or even early formulations of autopoiesis where "matter is, of course, part of the formulation, but does not enter into it as such" (Varela 1993)]. In these cases, "the aspects related to energy and matter … are assumed to be negligible in order to understand the principles of biological organisation" (Moreno & Mossio 2015: 7).

[91] We have already seen [read back through to check for consistency] that the distinction between matter and form was a central feature of Kant's philosophy—and the inability of matter to express form, or, as Kant puts 'contain an inner animating principle' that grounds the possibility for its own becoming, is constitutive of the 'unstable middle position' set up by Kant's transcendental philosophy. Schelling, for his part, sought a way to collapse the water-tight distinction—relegating their use to a product of *reflection*, e.g., our ability in thought to dissociate matter from form for the sake of analysis. Agreeing with Moreno & Mossio, then, there is no doubt "such an abstract approach has proved productive for scientific research" but such a "watertight separation between 'matter' (i.e., the material basis…) and form (the 'abstract' organisation) can be misleading when studying living systems" (2015: 7).

Thermodynamic autonomy is critical to my view of naturalised agency due to the establishment of a boundary condition and set of constraints that enable a sophisticated internal chemistry that has three interrelated features (though, as I will stress, the second has received considerably less attention). First, thermodynamic *closure* (what Moreno & Mossio call *closure of constraints*) enables a metabolism that constrains the flow of energy in the environment to be put towards the production of parts that can then feedback into the barrier—such as is postulated with autocatalysis or, in the living domain, the cell membrane and its carefully maintained, semipermeable boundary. Keith Farnsworth expresses the centrality of this as follows:

> If there were no physical boundary between an agent and its surroundings, then any external force would act unaltered throughout the agent and any force initiated by the agent would act equally on the internal and external environment, so no distinction in action could be made between the agent and its surroundings. If no distinction can be made, then no measurement could be made to tell us whether an action arose from within or beyond the agent. (2018: 12)

But we have already seen that active matter systems are *processes* that do allow for such a degree of measurement; and it seems quite critical that the ability to *discriminate* between actions endogenously generated and actions exerted 'from the world' is central to the establishment of a minimal agent—further speaking against Okasha's 'most minimal form of agency' that qualifies as a 'mere doing'. 'Mere doings' belong to objects, and, following Denis Walsh (2018), such 'object theories' are inappropriately applied to organisms and related systems (cells, networks, etc.). To borrow from Walsh, these self-organizing exemplars of agency spring onto us an 'ontological surprise' and require a different methodology and descriptive/explanatory armature for comprehending their activity and becoming. To reiterate, quite key to this is maintaining closure of constraints such that the system can continuously maintain itself until it eventually tends towards thermodynamic equilibrium (read: ceases to persist).

The second feature of thermodynamic autonomy is that, often times in relation to the first discussion point, the emphasis is on similarity and similitude: the thermodynamically maintained boundary continuously produces parts that come to reconstitute the system. But what we see, at least in the case of biological agency, is a much more radical kind of productivity: that is, the ability to create *new* kinds of materials that allow the system to proactively respond to a given environmental or endogenous situation. It is the ability to generate such novelty that is arguably one of the most distinctive features of living agents and their internal chemistries—whether through metabolism, epigenetic changes, or through selectional mutations that allow bacteria (for instance) to rapidly produce new materials to counteract hostile circumstances (Lamb & Jablonka 2011).

Thus, thirdly, the thermodynamic autonomy creates a physical barrier that provides the organism with a level of protection from external influences (Potter & Mitchell 2021: 3). Mitchell & Potter elaborate, stating that "rather than being subjected to every physical and chemical influence it encounters, the system is shielded from such forces, allowing it to regulate its internal processes. This is a vital move towards agent causation as it implies that external events do not definitively control the system. The upshot of this is that the external forces impinging on the selective membrane or boundary do not, by necessity, lead to a perturbation or alteration of behaviour (hence a degree of causal insulation: they are not just pushed around). Accordingly, this opens up a space for other factors to have influence on underlying dynamics, one factor of importance to the present argument being agent-level or systems-level causation. The ability to incorporate these *agent-level* features is precisely what I tried to show in the previous chapter—and is one of the reasons why notions of emergence and top-down causation are important to naturalisation projects. The recursive causality enabled by thermodynamic autonomy leads naturally to the next component of agent causation: Persistence.

2. Persistence

Agents are temporally extended entities with the ability to alter their current states in relation to past and present experience for the sake of perdurance and persistence—for as long as the system can be continuously maintained and parts can be restored (this, naturally, varies widely across the biological world [find interesting citation that demonstrates this]). Moreover, persistence is a phylogenetic and ontogenetic component of agency (see Walsh 2015), allowing agents to be extended both in developmental and evolutionary time. While the term 'temporal thickness' is often reserved for discussions of consciousness (Metzinger 2004) or, more minimally, having a mind, there is a straightforward sense too in which agents can be characterised by their temporal thickness (as we will see especially in (7) below).

As we saw with Egbert earlier[92] and the notion of *existential needs*, there is a clear difference between, on the one hand, tables and rocks that passively persist due to the absence of an outside force; and on the other, self-organizing systems that must act in ways so as to maintain their perdurance. To not do so would be tantamount to dissolution, and would not make for a very long lasting dissipative system—certainly not one that can then be swept into the open-ended evolution required for life on earth. Linking with (1), then, the difference in how *agents* persist and how *objects* persist is of a *fundamentally different* kind: "The difference lies in the fact that the way these two classes of systems exist is fundamentally different: rocks (and other non-dissipative entities) are merely passively stable, whereas dissipative structures are constantly falling apart and yet persist thanks to processes of repair, replacement, or reconstruction" (Egbert 2021: 3).[93] In contrast to the former, dissipative structures are *processes* (cf. O'Malley & Dupre 2006) and are therefore characterised by the distinct formal and mathematical language needed to describe them (far-from-equilibrium thermodynamics rather than statistical mechanics), the activities they can achieve, and their mode of measurement, namely, having 'rates' of energy dissipation that open up to quantification and measurement to see how viable they are as dissipative structures. To reiterate, "No equivalent measurement [exists] for passively stable systems as their existence is not a process in the same way that it is for dissipative structures" (Egbert 2022: 2).[94]

Agents, then, maintain stability through a constant and controlled constraining of energy and internal flux via "recursive chemical reactions" and "self-sustaining causal loops" (Potter & Mitchell 2021: 4). That is, component parts work to constrain other parts within the bounds necessary to keep the whole system going. Organisms thus exhibit dynamic, holistic patterns that regenerate the constraints required to keep themselves organised (i.e., alive). Along with a point made at the end of (1), this point of persistence further pushes against a reductionism that would see the causal powers reduced to the level of interacting components. It is the *pattern* of activity at a higher-order of scale that constrains and enables the constrained flow of energy for productive output. I will stress this point further in section (5.4), but for now I

[92] See Chapters 1 & 4.
[93] I stress this point as the difference is key to (3) to follow. I leave a philosophical analysis of the difference until then.

[94] There is a nice point of contact here between Egbert's emphasis on rates and processes being key for differentiating the two types of entities (e.g., objects like rocks and proto-agents like cells), and Farnsworth observation in (1) above when he writes: If no distinction can be made [between the inside and outside of a system], then no measurement could be made to tell us whether an action arose from within or beyond the agent" (2018: 12). In each case, there is a nonmysterious (indeed, quantifiable) way in which we can distinguish between the kinds of activities we see as necessary for agency and those of other (passively stable) entities.

turn to how (1) and (2) enable the third thing required for an account of naturalised agency: that of endogenous activity.

3. Endogenous Activity

That an agent on this view requires endogenous activity should be apparent from the previous two requirements for being an agent. The establishment of a barrier that enables thermodynamic autonomy (1) and persistence (2) also tends to establish an internal domain that 'buffers', in some causally important sense, the system from its environment. The ability for internal chemistry to maintain this self-differentiation (or, rather, self-individuation) is one of the key features of endogenous activity. As Potter & Mitchell (2022: 4) note, persistence might enable the system to stand apart from the physical world enough to be a distinct causal loci within it; but endogenous activity is what allows the system to evade being determined by its environment.

Borrowing from some of the more metaphysical language used in Chapter 2 (Origins of Organization), what is key to endogenous activity is that it allows the causal powers to inhere within the system itself and not be subject to the *nexus effectivus*—that is, *efficient causality*—that might characterise 'mere mechanism'. Recall from Emmeche et al. (2000) that *efficient causality*, undoubtedly the dominant mode of causal thinking in the life sciences, is the principle of a temporal cause-effect relation involving an interactional exchange that connects entities at one given level to entities at another" (2000: 4). For Kant (in the *Critique of Judgment* in particular), it comes down to the principle of mechanism that predominates in the physical world. For Jaegwon Kim (1999), this comes down to every physical effect having a sufficient physical cause. The establishment of endogenous activity is meant to 'insulate' the system from this deterministic notion of causality.

Thermodynamic autonomy makes it so not every change in the environment or in a solution has an *ipso facto* cause within the system. However, physical barriers are not sufficient for causal autonomy because the actions and forces that do causally impact the behaviour of the system could still be deterministic. To bring about causal autonomy in a sense relevant for agent causation, then, it is the inherent *endogenous activity*, as explored in the chapter on AMP (Chapter 5), that allows the system to act spontaneously and in non-predictable ways in relation to the environment. As we have explored in the previous chapter, organisms are constantly having to do work to sustain their activities. It is thus the pattern of dynamic activity—autonomy in combination with persistence and endogenous activity—that demonstrates a novel ontological locus to have causal influence in the environment. I believe this is, in part, what Walsh has in mind when he speaks of nature springing an 'ontological surprise'. Let us pause for a moment on this ontological surprise, as it is critically one of the things that makes organisms, *qua* agents, distinct entities in the world and in need of a different descriptive language and explanatory armature to accommodate their activities.

Walsh adopts the idea of an ontological surprise from Hans Jonas, whose seminal *Phenomenon of Life* text set the stage for much latter 20th century and 21st century discussions on the place of intrinsic purposiveness in nature. Jonas writes, "In living things, nature springs an ontological surprise in which the world-accident of terrestrial conditions brings to light an entirely new possibility of being: systems of matter that are unities of a manifold… in virtue of themselves, for the sake of themselves, and continually sustained by themselves" (1966: 79). Walsh, building on this idea, promotes a methodology that resonates closely with the view put forward in this work, something he calls a *methodological vitalism*: "[My] metaphysical position is that organisms constitute a special category of entity; they are natural agents. The methodological proposal is that because organisms are agents, a genuine understanding of the difference they make to the world requires a battery of theoretical concepts and exploratory modes that do not apply to the study of non-living things" (2018: 167). This fits well with a point made at the start of this section: in order to justify agent causation then a distinct entity with irreducible causal activity is present—that is, it breaks from an understanding of *efficient causality* that would apply to *objects*. Walsh calls the state of being an object *objectcy*: "Objectcy consists in the fact that the elements of the domain remain unaltered … unless they are influenced by external sources of change" (2018: 175). Moreover, Objects, at least standardly understood, are not capable of self-movement or propulsion. "The principles we call upon to explain their changes—for example laws of nature, intial conditions, the space of possible configurations—are *exogenous* to the objects" (ibid.).

The establishment and maintenance of *endogenous* activity, conversely, supplies the relevant dynamical activity for not simply being determined, or 'pushed around', by its environment. It quite literally requires a different scientific language to accommodate and account for its activity: that of far-from-equilibrium thermodynamics and active matter physics (see Chapter 5). In fact, Walsh argues that in order to account for agents in our scientific ontologies, we must switch from *exogenous* theories of causality to one centred around *immanence*—highlighting a point I make in Chapters 2, 3 and at the end of Chapter 5: namely, that according to the formulation of intrinsic purposiveness and what would be required for it to be considered 'naturalised', that we must formulate a concept of materiality and causality wherein matter can contain the ground for its own determination—that is, that intrinsic purposiveness is *immanent to the system*; not descriptively and exogenously ascribed thereof as part of heuristic convenience.

The point here is that endogenous activity equips the system with the relevant degree of causal autonomy: external perturbations or situations may serve as conditions for an agent's actions and operations; but they do not *determine* it. Stated differently, the systems actions are ultimately generated from *within* (are immanent to) the system (Potter & Mitchell 2022: 5). The difference between how objects and agents behave, then, reinforces the need to adopt an alternative methodological and descriptive approach. But in order for agent causation to be more robustly defended, then demonstrating how the dynamical pattern is the *holistic integration* of the system is also required. This will be our next point of discussion.

4. Holism

We have seen throughout this work that one threat to understanding agents as robust causal happenings in the world is the contention that they can be reduced to a set of parts and their interactions. If agents, on this view, can be understood as a shopping-list of ingredients "where the ingredients are mechanistic entities and their properties, and the recipe amounts to the organization and sequence of activities these entities perform" (Baetu 2015: 105), then it is not clear what invoking the holistically integrated agent-level adds to the discussion. Indeed, it instead appears that the agent can be left out of an explanation without much loss.

The idea that organisms can be decomposed into their constituents like this has a long standing heritage in the machine-analogy of the organism (Nicholson 2015). To see why holism does not play a constitutive role in the case of machines, consider that the parts relate to each other in a machine *extraneously*. That is, they are *extrinsically purposive* in the manner we explored in Chapter 2. This is in contrast to the view of holism—which so perplexed Kant and generations of biologists that followed—that posits an *intrinsic purposiveness* at the level of the integrated system: its parts exist because of, and for the sake of, the persistence of the whole, self-organizing system. Even though both machines and organisms act towards the attainment of specific ends—e.g., both are *purposive systems* (and there is a sense in which machines have some form of 'existential needs' too)—organisms exhibit the intrinsically purposive organization that we explored in Chapter 2. In the case of extrinsically purposive systems, their telos is constructed or produced for them and it does not serve interests that are, generally, other than its maker or user (Nicholson 2014: 163).Organisms, as we have seen, are intrinsically purposive in the sense that their activities are directed towards their own maintenance and organization, and this is something immanent to the material organization of the system: "The intrinsic purposiveness of organisms is grounded on the fact that they are self-organizing, self-producing, self-maintaining, and self-regenerating" (Nicholson 2014: 163).

The key point here is that in organisms—due to their holistic nature—the parts are neither causally independent of, nor temporally antecedent to, the whole they constitute: Instead, they exist in a state of mutual interdependence, relying on each other for their creation, upkeep, and rejuvenation. This is what is key, then, in defending naturalised agency for a biological ontology, as it enforces the importance of a causally integrated system where, as Kant puts it, the parts exist for the sake of the whole and the whole for the sake of the parts.[95] Holism, then, is what establishes the intrinsically purposive nature of the systems in question. And being intrinsically purposive in this material, bioenergetics sense is constitutive of being an agent on the proposed account.

The point about holism can be further articulated by providing the example of chemotaxis in a bacterium. Bacterial chemotaxis is a well-studied phenomenon involving transmembrane structures and chemical pathways for detecting food substances (Potter & Mitchell 2022: 6). The transmembrane system is connected to a flagellar system, and once activated allows the bacterium to navigate chemical gradients for richer concentrations of nutrients. But even this seemingly simple pathway is heavily modulated by ineliminable contextual cues concerning the metabolic state of the whole organism. The upshot is this: if contextual cues (such as metabolic states) can modulate the behaviour in response to the same environmental circumstances or inputs, then it is not entirely correct to isolate the chemotactic pathway as the *determinant* of the behaviour. Rather, it is a constellation of properties and existential needs across the system that determines and influences how the bacteria will behave. Existential needs are irreducible to consitutent mechanisms and causal pathways *within* the system, and

[95] Stuart Kauffman similarly defends the significance of holism here (a 'Kantian Whole'), which can be defined as a system in which "the parts exist for and by means of the whole, and the whole exists for and by means of the parts" (2013: 609).

instead reside at the agent level. Bacterial behaviour—indeed, organisms generally—do not make sense outside of their drive to pursue existential needs, and this is a holistic feature of the system.[96]

Defending holism also goes some way in blocking a species of reductionism we saw at the beginning of Chapter 4 ('Just Physics?'): that of horizontal reductionism. Horizontal reductionism is the postulation that the causal powers relevant to a systems activities inhere not at the level of the whole, but to a specific subset of components within that system. Because organisms cannot sensibly be decomposed into a 'list of ingredients' that themselves causally instantiate the dynamical pattern independent of the whole, the agent cannot be said (without losing something of significance) to be reducible to this set of parts (that is, it is not machine-like in a manner picked out by the machine analogy above). Lastly, the emphasis on interconnection and interrelations between wholes and parts places a further, metaphysical emphasis on *relationality* as determinative for organismic identity: what an agent *is* inheres in the relational pattern of activity, and relations here are inherently emergent (Campbell & Bickhard 2013). We will return to this point at the end of the chapter.

5. Lower Scale Indeterminacy

Holistic integration—covering features (1)-(4)—appears to block certain versions of reductionism: specifically, horizontal reductionism. However, it could still be the case that there is still some causally fundamental level 'below' the level of the agent that is truly acting in these cases—that is, *vertical reductionism* still might follow. However, I have given reasons throughout this work to be sceptical of causal fundamentalism, but I have not articulated how blocking causal fundamentalism relates to the position on agency promoted herein.

Recall that vertical reductionism is the principle of an almost Laplacian view of nature wherein lower-level happenings would be the location of causal powers. Emergent features would be causally epiphenomenal on this view. Evidently, this would present a threat to agent causation. Thankfully, there are already reasons native to the supposedly most 'fundamental' theory—that of quantum field theory—that suggests this is not the case. As Campbell &

[96] Intriguingly, Potter & Mitchell, in their discussion of holism, remark that biological systems are a 'Kantian whole'. These are systems so "deeply interconnected that it doesn't make sense to decompose them it into its component parts, because the true essence of the system exists in the relation between those parts" (2022: 7).

Bickhard (2011: 34) argue, "Our best physics tells us that there are no basic particulars, only fields in process".[97] This shifts our ontologising away from discrete loci occupying supposedly fundamental levels to an ontology that gives priority to *organization*, an inherently relational property. In addition to highlighting the importance of an ontology that prioritises organization and relationality, it also suggests something important about reductionism.

More perspicuously, the evolution of quantum fields is not wholly predetermined in a manner that conforms with standard philosophical formulations of physicalism (which might conform closely to a Laplacian view; see Gallagher 2012). It is now well acknowledged that quantum fields involve fundamental indeterminacy, but it is also acknowledged that this is not enough to safeguard agent causation (Potter & Mitchell 2022; Ellis 2012). What matters in the case of organisms—as we saw especially over the past two chapters—is the emergence of structures that bias the indeterminant unfolding of these processes into more constrained and determinant directions: In living organisms, this occurs primarily because of the iterative, step-by-step influence of natural selection and the continuous learning that individual organisms undergo throughout their lives (Potter & Mitchell 2022: 9). This is what is meant, partly, when we speak of natural selection naturalising extrinsic teleology; and ontogeny—or a theory of the organism more generally—being required for the naturalisation of intrinsic purposiveness: that there exists structures that can scaffold, constrain, and guide the material and energetic processes to their determinate outcomes. Establishing this also correspondingly raises the issue of goal-directedness and the actions required for pursuing goals.

6. Multiple Realizability

I briefly discussed certain elements of multiple realisability in the Introduction—specifically, its tendency to malign the material part of cognitive processes in favour of more information-theoretic (and quite immaterial) software specifications (see Polger & Shapiro 2016 for an insightful overview). Despite my hesitation in endorsing multiple realisability as it is commonly understood in the cognitive sciences (implying substrate-independence of cognition), it is clear that multiple realisability has some (more refined) role to play in the life

[97] Campbell & Bickhard's account of emergence and downward causation will be increasingly essential to the present chapter at the end.

and mind sciences (as Chirimuuta explores (2019))[98]. In biology, the multiple realisability of functions is often called 'robustness' (see Wagner 2005, 2014; for a philosophical articulation, see Koskinen 2019). Briefly, robustness is a concept in biology that captures the functional invariability of biological processes despite considerable environmental perturbations and metabolic fluctuations. This plugs nicely into the previous (5) point because it emphasises the importance of a functional level that scaffolds and guides lower-level processes. For the sake of the discussion of agency, what it comes down to is the way in which hierarchically structured goals become sedimented in ontogeny—guiding and constraining the flow of information and energy an organism can use to pursue its goals. Stated differently, the goals of the system need to exhibit some insulation from lower-level happenings—and indeed, the ability to ignore or eschew a lower-level description is quite critical to establishing goals at an agent-level as causally meaningful—and at least some of the system's causal power must inhere within higher-level organization (Mitchell & Potter 2022: 9).

To see why, we will look at the example of bioelectricity explored in BC. Bioelectricity has been extensively explored in the work of Michael Levin and colleagues at the Levin Lab (where it has been explored for its implication in a range of biological processes: such as cancer, ontogenesis, and has formed the basis for a more general theory of biological development and morphological coordination (2019, 2022). Indeed, all cells are electrically active, and one of the key points of Basal Cognition is that neurons (and nervous systems) exploit evolutionarily ancient mechanisms to pursue goals at higher orders of scale (Levin 2019). The developmental *goals* of a system depend on macroscopic patterns. Complex animals acquire a particular body plan during development which is robust in the face of genetic mutations, in utero fluctuations, and other cellular disturbances. Indeed, Levin and colleagues have demonstrated with numerous model organisms that scrambling early-stage embryos can still lead to largely normal and species-typical morphologies. The usual story for this is that a genetic program determines the species-typical body plan of the organism, but a set of important experimental and theoretical work has sought to situate genes in their *material* and *bioelectric* context: "Recent advances in developmental bioelectricity have shown how endogenous dynamics of resting potential changes modify transcriptional cascades and thereby instruct axial patterning, organ determination, and size control, as well as guiding the behaviour of individual cells" (Levin 2019:5).

[98] I explore the implications of this version of multiple realizability for the mind sciences in the following chapter. With my colleagues Wiktor Rorot and Urte Laukaityte, I have explored this issue extensively in the domain of soft robotics (see Harrison et al. 2022).

In Chapter 4 ('Just Physics?'), we looked at the material influences of the ECM and tissue complexes for carcinogenesis. Interestingly, carcinogenesis has also been a primary area of study for BC and agential approaches to the organism—with Sonnenschein & Soto (2016) arguing that it is now imperative to cultivate an agential theory of the organism to capture the specific cellular dynamics involved in a diverse range of biological processes, not just when things go wrong in cancer (see also Soto & Sonnenschein 2011). One dimension of carcinogenesis that we did not explore in Chapter 4, but which has been implicated in the development of cancer, is the *bioelectric* context: "In carcinogenic transformation… cells become isolated from the physiological signals that bind them into unified networks… in the absence of global cues, they revert back to their unicellular past, when their behaviours were aimed at optimizing the future of just one cell" (Levin 2019: 6). In physiological networks, cells are often electrically connected via gap junctions that generate global oscillations regulating individual cell behaviour—and a breakdown of this coherence is precisely what we see in certain instances of cancer transformation.

The point, when it comes to multiple realisability, is that the maintenance of agency—the ability to pursue goals in a unicellular or multicellular context and execute behaviours that ensure its persistence—must be buffered from lower-level happenings in certain instances for this process to unfold. If the causal powers inhere at the level of the agent (in this case, the bioelectric patterning that regulates cell-cell interactions) then attributing developmental goals to a system or network-level is not just idle metaphor. The case of cancer demonstrates this, if negatively, by showing how once the robustness that characterises this bioelectric patterning breaks down, the ability both for the cell to pursue goals at a higher-order of scale (the level of the organism) and the ability of the organism to pursue its own goals is hampered.

7. Historicity

With (7), which we can call historicity, we seem to be ascending into an understanding of agency that is already quite elaborate. This appears to require more robust memory systems capable of embedding the agent in a causally thick past that allows it to navigate the present in pursuit of its goals. While this is indeed one thing required for the kinds of agential dynamics ubiquitous throughout the organic world, I intend to convey something slightly different with historicity in this section.

Researchers in extended cognition have long identified the significance of *meaning* when discussing what makes an organism or agent distinct from other kinds of systems. As Evan Thompson remarked, sucrose is only a nutrient in the context of an organism like *E. coli*; a molecule in a solution does not qualify as nutrition independent of the system for which it can nourish (Thompson 2007). Without getting too deep into the complex literature on *naturalised meaning* or *semantics* (a heady enough topic as it is), we can say, provisionally, that meaning is a *relational* feature of an *agentic* system embedded in an environment that affords it the potential to survive, flourish, and reproduce. The agent, as a physical system, consists of structures that have been selected for, and which have developed ontogenetically, that crystalise or embody these relational features of an environment (Allen & Stotz 2012). It is the physical structuring of the system that endows it with this relationality that I pick out with *historicity*: agents are fundamentally historical beings, and that history unfolds on both phylogenetic and ontogenetic timescales.

The focus in theoretical and evolutionary biology has undoubtedly been on the extrinsic purposiveness that natural selection provides (see Dresow & Love 2023). But it is increasingly appreciated that this component of evolution unfolds in developmental time and can be, and often is, modulated by learning and the behaviour of the system in question (see Stotz & Allen 2012). The extent to which history matters for a system, then, is the extent to which that system can be said to be capable of having 'experience' in a developmentally and behaviourally meaningful way[99]. This emphasis on history, then, could be reframed so as to outline the centrality of basic learning and adaptivity for being an agent.[100]

---

[99] Here, my use of experience connotes the ability to integrate environmental stimuli via signal transduction in a way that can lead to long-term, adaptive changes within that system. It does not imply conscious experience, or even that there is 'something it is like' to be that system. To reiterate, it is the capacity for the environment to become internalised and imprinted on a system such that when it is coupled with that system's own endogenous activity, it is able to modify structures to adapt to specific ends.

[100] Indeed, the lack of adaptivity in earlier formulations of autopoiesis turned out to be a critical conceptual lacuna, as Di Paolo argued (2005).

8. Normativity

One of the most essential things about being an agent—and arguably what almost all articulations of naturalised agency are geared towards in one way or another—is that of *normativity*: the capacity for things to go rightly or wrongly for a system. Taking examples surveyed thus far, we could say that there is nothing 'going well' for a tornado zipping across a field before eventually dissipating. Even the motile oil droplet with its primitive internal chemistry lacks the sophistication for there to be something like 'normativity' intrinsic to it. Machines can be said to be normatively guided because they're the products and artefacts of intentional systems like ourselves. Their normativity is thus extrinsic in most cases. What interests me in this work is the kind of *intrinsic* normativity present in living systems. As we will explore in the following chapter, organisms have real 'skin in the game'—they are not extrinsically related to their outcomes or fates in a way that, say, a digital computer might be. Much of the theoretical nugget I wish to get at with this work on agency could be framed in relation to Haugland's point from earlier: what makes it such that organisms paradigmatically 'give a damn'? Agency is one of the wedge issues I use to get at this.

Similar with meaning above, the literature on naturalised normativity is fraught and dense. My goal in this chapter has been more modest and evocative: in highlighting the thermodynamic, material, and physiological basis of agency, what I have been arguing is that there is a deeply physiological sense in which normativity can be understood: because the parts of organisms are not extrinsically related but rather are geared towards the maintenance of a whole (what we might call, following Merleau-Ponty, the 'structure of behaviour')—and it is the inhering of this normativity within self-organizing dynamics that this insaturation of a normative standard can occur.

5.4 Reinserting Matter & Materiality into the Debate

The eight features above are meant to be a collective roadmap of what qualifies as agential—there are surely other ways to parcel out agents according to the needs of researchers and their disciplines (for instance, learning algorithms are often called 'agents' in machine learning literature but do not exhibit the kind of agency I discuss here; moreover, these agents being considered *as if* agents would not necessarily present problems in the way treating organisms as such has generated in the past). Again, my interest is in a kind of agent that represents a distinct ontological category from the extrinsically purposive teleology of, say, digital computers, tools, and artefacts. The agents that concern me are those that can be said to exist on their own behalf. As sloganistic as this sounds, the above discussion, in tandem with the previous two chapters, is meant to show that these systems exhibit distinctive causal dynamics; are amenable to certain kinds of analysis and methods of quantifications; and require different conceptual resources for better explaining their activities. If we link the naturalising agency debate with the discussion of intrinsic purposiveness, then, I believe what we have begun to identify is the *immanence* of agency within diverse constellations of materiality. This is one of the main points I have tried to stress in this work thus far: if purposiveness (and, hence, agency) can only be imputed extrinsically under the Kantian scheme—relegating non-anthropogenic versions thereof to the domain of description or a descriptive gloss—then under the *Post-Kantian* perspective, with its concerted rethinking of matter, the emphasis shifts to how *materiality itself* enables the dynamics and organizational features seen as critical for agency.

It is this sense of agency that I want to capture with the above eight criteria: these are agents with real skin in the game—agents for whom not acting, not behaving in viably requisite ways, cease to persist. Jonas' insistence that organisms, as agents, spring an ontological surprise, one where "systems of matter that are unities of a manifold… in virtue of themselves, for the sake of themselves, and continually sustained by themselves" (1966) is an existentially and philosophically meaningful demarcation of a domain of systems that require different strategies for understanding them. This is important, as we will see, not just for currently living systems or understanding how the evolutionary process has unfolded; but in determining what is sometimes called 'the space of possible minds'. As we will see in the following chapter, cultivating an agential theory that focuses on the material processes of actual agents—and the constraints and physical limitations mediating them—can equip us with fecund conceptual resources for developing autonomous systems both in Soft Robotics and Artificial Intelligence (as explored in Harrison et al. 2022). Furthermore, although I do not explore the possibility further in this paper, a more robust agential theory in biology and cognitive science might aid exobiological research—a discipline that, at least in the theory-arm, is no less concerned with spaces of possible minds (see Powell 2020 for a comparable approach).

I have already suggested that a more concerted engagement with the materiality implicated in the dynamics relevant to life and mental processes is missing from most discussions on the place of mind and agency in nature. The kind of materiality that interests me is expansive and broad in scope: from the piezoelectric and elastic materials being explored in engineering settings (Shah et al. 2021), to the chemical forces germane to the maintenance of metabolisms—such as oligomers and other such large molecules that enable a kind of complexification seemingly required for open-ended evolution and the emergence of increasingly complex bodies. Coupling these considerations on materiality with the thermodynamic and existential specifications of the above eight criteria, then, provides fecund conceptual resources and testbeds for better understanding the self-organizing and interactive dynamics pertinent to agency.

5.5 Conclusion

One is struck when reading work in Materials Sciences by the emphasis that many developments, inventions, and discoveries in human culture exhibited a sensitivity and appreciation of the potencies of matter. Something like this appreciating for an active materiality and its relevance for agential dynamics is going through a renaissance in the physical and biological sciences—and I would like to conclude this chapter with a look at one of the most exciting domains of materials sciences where the biogenic approach to agency is becoming increasingly appreciated. Laura Tripaldi notes: "The best way to exploit the self-organizing capacity of a material system is to renounce all top-down control and allow its relational structure to emerge in all its complexity" (2021: 85). This is very evocative, but renouncing all top-down control is probably an overstatement. Ebramkhani & Levin (2021) propose that it is engineering principles in combination with our understanding of biological materials that can play a role in what they call 'guided self-assembly' (though whether self-assembly is the right appellation here is something I tried to cast doubt on in the previous chapter). Moreover, it overstates the causal direction in self-organization—not leaving much room for the agential dimensions I have explored in this work.

Nevertheless, Tripaldi is certainly correct when she writes that "atoms are not inert bricks to be assembled one by one; matter is endowed with its own inorganic will, which, depending on the approach we choose to adopt, can act either as an inertial force opposing our manipulations, or as a startling intelligence that acts to our advantage" (2022: 85). It is in probing the depths of matter and better conceptualising spontaneous self-organization in physical phenomena and agential dynamics in the biological domain that we can further both the sciences of matter, and potentially develop artificial, synthetic systems that exhibit the remarkable flexibility and adaptivity that exemplify the living world. As it happens, this is precisely what we find when we look at the disciplines of soft robotics and neuromorphic computing—representing an intriguing coalescence of theoretical considerations in biology (which serves as the inspiration) and the ambitions of creating autonomous and flexible machines in robotics and A.I., a topic to which we now turn.

Chapter 6

The Biogenic Approach and the Making of Soft Robotics and Bio-Inspired Artificial Intelligence[101]

6.1 A Biogenic Approach to Minds and Agency for Bio-Inspired Artificial Intelligence[102]

As we saw in the Introduction to this work, standard approaches to understanding cognition—and the wider goal of recapitulating it on simulated platforms or in the field of robotics—have tended to neglect the importance of the materiality of the body and its relevance for constraining, enabling, and mediating cognition. This contention is centred right at the origin of the cognitive sciences and is typically framed in terms of *multiple realisability.*[103] It is often argued, then, that cognition is a species of software that, in principle, is instantiable in any back-of-the-envelope set of materials so long as they are "suitably organized" (Putnam, 1975).[104] Although few authors would defend this version of multiple realisability (MR) today (see Polger and Shapiro, 2016 for a state-of-the-art discussion of the philosophical literature), the belief that the materiality of cognition is mostly a "hardware" problem with the truly interesting explanandum being cognitive "software" that sits above still permeates much of the theoretical and philosophical literature. However, as we have seen throughout this work, much turns on what it means to be "suitably organized" and it is by no means clear that any pell-mell set of materials could instantiate the complex dynamics on

[101] The majority of this chapter is reworked from an earlier piece of mine, "Mind the Matter: Active Matter, Soft Robotics, and the Making of Bio-Inspired Artificial Intelligence" published in *Frontiers*. My co-authors are Wiktor Rorot & Urte Laukaityte. The contributions to the original work break down as follows: I oversaw the project for the Diverse Intelligences Summer Institute (DISI 2021) and organised the totality of the project. In the original paper, Wiktor Rorot wrote section (6.2) on Multiple Realizability, and I have reworked that section in this thesis to reflect my own voice. Urte Laukaityte helped with overall strategizing, input, and editing of the paper. All other sections were predominantly, if not exclusively, written by me as lead-author.

[102] Any redundancies found in this chapter are due to its repurposed nature from Harrison et al. (2023) and, although some details might be repetitive, helps maintain narrative consistency as a whole. Any redundant materials found herein can be framed as a preface for how the previous discussion might arrive at some understanding of cognition amenable to the goals and aims of robotics and A.I.

[103] A term so far alluded to but which has lacked proper theoretical specification.

[104] Putnam's memorable example of 'Swiss cheese', even if later recanted, is a common allusion in this chapter. I have removed repetitive quotes for the sake of reader's ease. The background to this debate was explored in Chapter 1.

which cognition depends. In fact, building on the material from the previous chapters, what we see is that materiality imposes a filter on the kinds of physical states that can endogenously and autonomously sustain ordered states out of equilibrium (Harold & Morwitz 2022: 37)—such that greater sensitivity to the material and energetic conditions of the agential system will need to be factored in when determining whether it exhibits the degree of internal complexity associated with the intrinsic purposiveness that has been the focus of this work.

This chapter, then, telegraphs the argument of the previous chapters into the domain of robotics and so-called bio-inspired artificial intelligence [AI]. to advocate for a harder pivot towards the materiality of cognition. The goal of this argument is to address a cluster of interrelated issues both in the philosophy of AI—specifically as it relates to the MR thesis—and in engineering paradigms in the actual sciences of A.I. and robotics. Stated differently, I take the insights explored in the previous chapters on Basal Cognition, Active Matter, and Material Sciences and argue that the separation of non-mental, living "hardware" and cognitive[105] "software" has grown increasingly suspect. What I am claiming is not (or not simply) that the body matters to cognition (a view that tacitly supposes the body in service of a more or less higher-order, more or less unified, cognitive subject), but rather that the body itself—at varying levels of organization—exhibits cognitive and agential capacities through and through: from cellular activities entrained to regulating morphology, development, and intercellular communication; to tissue complexes and system functioning; through to more baroque appearances of cognitive sophistication encapsulated in cephalopod, arthropod, avian, and mammalian brains—Darwin's "endless forms most beautiful" (*Origin of Species*). In a slogan expressed elsewhere (Levin, 2019, 2020; Levin and Dennett, 2020), this is cognition *all the way down*, not just proprietary to a unified subject. Making sense of the theoretical commitment behind this claim and how it eventuates in the development of intelligent machines is the main goal of this concluding chapter. It is thus worth clarifying at the outset that what I am interested in here is Autonomous Robots (AR), i.e., autonomous, embodied systems capable of recursive self-organization, goal-directedness, and agency—the ability to flexibly and actively select goals relative to its "existential needs" (Froese, 2016; Egbert,

[105] The work up until this point has focused predominantly on notions of agency. The paper that forms the content of this chapter focused more on the term 'cognition' and will be explicated accordingly as the chapter proceeds. The reason I focus on cognition herein is because this was a work in cognitive robotics that took inspiration from Basal Cognition and was written before the majority of this work was produced. However, the two concepts interrelate (see Chapter 1) insofar as an agential dimension is required, or so I have argued, for the genuine kind of cognitive endowment that interests me in robotics and A.I.

2022) and remain the kind of system it is (Man and Damasio, 2019). The key here, as I see it, is to understand how the matter matters to being this kind of system.

In the spirit of Basal Cognition explored in Chapter 1, the picture being worked against is one of a neurocentrism that cleaves neuronal (and cognitive) activity from the living, developmental, and morphogenetic processes for which nervous systems originally evolved (see Lyon, 2006; Van Duijn et al., 2006; Keijzer et al., 2013; Newman, 2016, 2019, 2022; Levin, 2019, 2020; Fields and Levin, 2020; Sims, 2020, 2021; Fields et al., 2021; Jekely, 2021; Lyon et al., 2021; Wan and Jekely, 2021). It is this sense in which the tacit commitments of MR—the in principle cleaving of active, living processes and cognitive ones—deserve a reconsideration. As Peter Godfrey-Smith remarks, philosophers and cognitive scientists tend to operate with a "picture in which living activity is a kind of non-mental substrate, and then evolution lays a computer—the nervous system—on top of the merely living, after which cognition and subjective experience result" (Godfrey-Smith, 2016a: 496). This can be seen in the very structure of the cognitive sciences and its lack of (explicit) emphasis on the life sciences. That is, while biological perspectives have influenced theorising about the mind [e.g., autopoiesis (Varela et al., 1993; Weber and Varela, 2002) and enactivism (Di Paolo et al., 2017)], they have not furnished real competitive alternatives to more mainstream cognitivism and computationalism [see Meyer and Brancazio (2021) for an insightful discussion]. To this day, it is common to see neurons and the brain—the "stuff" of cognition—almost wholly abstracted from the life processes in which they are embedded.

Here, I cast doubt on the (un)happy divorce between material and cognitive processes by suggesting that looking toward recent developments in soft robotics (Man and Damasio, 2019; Blackiston et al., 2021; Bongard and Levin, 2021; Kaspar et al., 2021; Kriegman et al., 2021), active matter physics (Hanczyc and Ikegami, 2010; Needleman and Dogic, 2017; McGivern, 2020; Egbert, 2021), and basal cognition research (Lyon, 2006, 2015; Van Duijn et al., 2006; Newman, 2016, 2019, 2020, 2021; Levin, 2019, 2020; Bechtel and Bich, 2021; Lyon et al., 2021) complicates any cleaving of cognition from its living, material context. In light of recent empirical advancements, I submit that now is a good time to revisit our philosophical assumptions regarding the MR of the cognitive and suggest that a more promising path in the development of AR and Artificial Intelligence (AI) is to take the materiality of cognition *more*, not less, seriously—a position explicitly disallowed in standard understandings of the MR thesis. The argument herein thus consists of two interlocked moves: first, I identify a set of assumptions that structure the debate on MR and that generate strong intuitions regarding the mental-physical interaction that have historically discouraged taking the materiality of cognition seriously. Second, I propose a path to AR that explores a more thoroughgoing, "radically embodied" approach: one that does not see the body as a "non-mental" substrate on top of which cognitive software (the nervous system) is placed, but instead depicts cognition as a more fundamental feature of cellular (read: living) activity and self-organizing processes in far-from-equilibrium conditions that are then *scaled up* in appropriate ways to arrive at more sophisticated multicellular animals.

At this point, it is worth being explicit about three things. First, there is a strong connection between living and cognitive processes—consistent with much of the literature on the so-called life-mind continuity thesis (Maturana and Varela, 1980; Thompson, 2010; Sims, 2021). *Prima facie*, this would seem to undermine our goal of constructing AR, as it would suggest some of the prototypical cognitive behaviours we see in certain soft-bodied robots and active material systems (examined in the sections below) cannot qualify as such due to their non-living nature. This problem can be ameliorated, however, by adopting a conception of cognition that depicts the living and developmental side of the process as a more general feature of self-organizing systems in far-from-equilibrium thermodynamic states that must act in a denumerable set of ways to remain the kind of system it is. Simply put: I accept here a view of "life" which does not presuppose particular material foundations (e.g., carbon-based), but rather takes it to be an organizational feature (cf. Moreno and Mossio, 2015; Harold & Morowitz 2022). Under this view, cognition can be seen as tailored for the homeostatic processes that underpin goal-directed, autonomous, and agentic behaviour (see Pezzulo et al., 2015), and a non-living system would fall closer to the cognitive the more it embodies such dynamics. This brings me to the second consideration, namely, that the lynchpin for the discussion of cognition turns around the notion of "existential needs" (see Chapter 1) and can be explicated, following Lyon et al. (2021), in relation to the set of sensory and information processing mechanisms organisms have for familiarising themselves, valuing, and interacting actively with the environment in order to meet the existential needs of survival, persistence, growth, and reproduction. In the literature, this is often called *basal* cognition, as it refers to a set of mechanisms and capacities with highly ancient, highly distributed origins. I earmark this for now and return to it below for a more nuanced discussion.

This focus on existential needs depends on recent research advocating for taking the materiality of our embodiment further than mainstream embodied cognition has commonly done (cf. Muller & Hoffman 2017). I thus place a premium on the very processes, goals, and demands of a living body that are normally elided from more theoretical meditations on the cognitive. As we saw in Chapter 1, a similar approach has been proposed by Man and Damasio (2019). They suggest a transition away from the hard parts that typify traditional roboticist approaches to fragile, vulnerable, and soft materials characteristic of organismic embodiment. The fundamental innovation introduces homeostasis and risk-to-self as the warp and weft of cognitive embodiment: "These machines [what I have called AR] have physical constructions—bodies—that must be maintained within a narrow range of viability states… Rather than up-armouring or adding raw processing power to achieve resilience, we begin the design of these robots by, paradoxically, introducing vulnerability" (Man and Damasio, 2019: 449). Indeed, similar to Man and Damasio, I believe a shift from embodied (*simpliciter*) AI to homeostatic and precarity driven AI is the key requirement for the coming generations of AR. This puts more emphasis on the material *processes* and material situation than simply focusing on embodiment full stop.

The second notion I depend on has already been mentioned: that of precarity. Tom Froese has argued that the nature of our embodied precariousness (risk-to-self) is essential for agency and the problem of meaning.[106] He writes, "The precariousness that is intrinsic to all organismic, and therefore also of all mental, existence is the original reason why things matter to that individual being" (Froese, 2016: 34). That is, organisms are cognitive agents with meaningful engagements with the world because of, and not in spite of, their fundamentally precarious nature. Importantly, this can be also expressed in terms of values and value-realising, which some believe to be the main force driving and organizing action in cognitive agents (cf. Hodges and Baron, 1992; Hodges and Raczaszek-Leonardi, 2021). Precarity is the minimal form of valence, hence enabling cognition and agency (cf. Lyon and Kuchling, 2021). Recall that if "the problem with AI", as John Haugeland put it, "is that it doesn't give a damn" (Haugeland, 1998), then what I want to explore herein is how an active matter lens focusing on specific material reconfigurations that enable systems to maintain themselves in far-from-equilibrium conditions can make headway on this most defining of problems for computer science: autonomous robots that might one day 'give a damn'.

[106] We might call this a species of the frame problem (McCarthy and Hayes, 1969): why would an artificial agent come to care about its existence and actions on which it depends?.

The structure of this chapter will be as follows. Section 7.2 introduces the Multiple Realisability thesis and advocates for the account of functionalism that is relevant for both the original articulation as well as the fine-grained functionalism for which I advocate. This allows the present argument to link up with contemporary strands of functionalism in a way that still takes fine-grained details of the system seriously. Section 7.3 briefly recapitulates some of the AMP material from [Chapter 4] to emphasise the focus on materiality that is relevant for new soft bodied robotics trends. Section 7.4 then shifts towards the main focus of this chapter, which is soft robotics and the bio-inspired nature of these novel systems that exhibit increasingly flexible and plastic behaviour—advancements that could represent a stopover on the way to more general and genuine artificial intelligences. Section 7.5 concludes with some philosophical observations about the AI literature and how the biogenic approach explored in this work might plug naturally into the ambitions of these researchers.

6.2 Multiple Realizability, Functionalism, and Fine-Grained Functionalism

Before exploring the biogenic approach and its relevance to main, it is important to highlight some of the entrenched philosophical assumptions of contemporary AI and robotics approaches that I believe impede the development of ARs.

The methodology of contemporary AI research is built on the philosophical programme of functionalism in philosophy of mind. Functionalism was formulated in the second half of the 20th century in response to the issues surrounding the physicalist mind-brain identity theories dominant at the time. In the late 1960s, Hilary Putnam advanced a novel line of thought that sought to establish that mental states (and properties) are *functional* states. From the onset of this view, functions—understood as causal mappings between sensory inputs, other internal states, and behavioural outputs—were construed in broadly computational terms. This allowed philosophers to disentangle cognition from its neurophysiological, material basis and argue that cognitive processes are "General" (Polger & Shapiro 2016: 15), that is, shared across species, and in fact that cognitive functions can be realised by entirely distinct types of systems—not only differently organized animal brains but also a variety of non-biological systems. A special case of interest was the then-developing digital computers, which seem under this view to be well suited for realising psychological processes. This is the idea that has come to be known as the *multiple realisability of mental states* (hereafter, simply MR).

As Chirimuuta (2018) observes, functionalist theories of mind and the concept of MR hold a unique status in philosophy as views to which a "near majority of philosophers have subscribed to, and for more than one generation". However, in an important book, Polger & Shapiro (2016) argue that the view of MR explicated above is increasingly untenable and misaligned with the empirical evidence accrued over time. The main thrust of their argument is aimed at the tenuous distinction between inherent, superficial variation in the biological world and deeper differences which are in fact responsible for MR. This in fact tours out to be damning regardless of whether one assumes that mental states are multiple realisable functions (the ontological, objective stance) or whether one argues that they can be *explained* as multiply realisable functions (the epistemic or subjective view). Their points are targeted at what can be called MR 1.0 (Chirimuuta 2018), and, as a result, they call for a rejection of traditional functionalism.

The functionalist account has suffered from other important theoretical criticisms as well, among which we may highlight the dual objection that functionalism is either (1) too liberal or (2) too chauvinistic. According to the liberality concerns, functionalism is too generous with what it can ascribe mentation to—recapitulating concerns I have covered in previous chapters in the intrinsic teleology and agency debates. In fact, an important argument in this vein comes from Putnam himself, who later in his life rejected the computational theory of mind. Putnam (1988) proves the theorem that "[e]very ordinary open system is a realisation of every abstract finite automaton", which would lead to an uncontrolled expansion of systems that I should consider as realising cognitive functions—contrary to our experience with the world. The opposite argument has been initially suggested by Ned Block. Block (1978) argues that any version of functionalism that avoids liberalism by opting for some set of physical specifications falls into biological chauvinism and hence denies mentality to creatures that I would ordinarily consider as such. His reason for the claim is the thought that one could always conceive of some system that would fail to meet the physical constraints and yet intuitively seem to possess psychological states.

However, despite the problems with traditional functionalism, the conviction that MR is an important feature of the cognitive remains widespread among researchers. In fact, it plays a significant role not only in the study of cognition but in the life sciences at large. This leads Chirimuuta to propose that instead of rejecting MR altogether, we need to carefully update this notion to account for the role that function parlance plays in biology—beyond the species of traditional functionalism that is the source of Polger & Shapiro's critique. Moreover, we saw in the previous chapter that some species of MR seems not only scientifically warranted, but required for a better understanding of the agential dynamics that have been the focus of the discussion so far. Rather than jettison the concept *in toto*, it would be better to incorporate what we know about the necessity of certain fine-grained functional details and materials into how we theorise on the distribution of minded creatures.

Pulling again from Chirimuuta, I posit a MR 2.0 that is grounded in biological considerations. MR 2.0 departs from 1.0 by emphasising the importance of material details when demarcating possibly minded systems while also trying to grant sufficient breadth to the space of possible minds. Indeed, describing functions in terms amenable to MR is an important assumption for working scientists and allows them to make decisions with regard to what physical properties can be safely ignored in their experiment—thereby reducing the complexity of the problem to be studied. This conception of MR, Chirimuuta suggests, allows us to maintain that "[i]t can both be true that the material from which the nervous system is built (i.e., living, metabolizing cells) is crucial to their function *and* that those functions are multiply realised" (Chirimuuta, 2018: 411). In particular, this view lets us appreciate that the Heraclitean nature of biological material—its ability to preserve integrity through "continual turnover of matter and energy" (Chirimuuta, 2018: 411)—is crucial for understanding the functioning of cognition [as Godfrey-Smith (2016a) also argues], but at the same time a roadblock to the success of purely reductive methodologies, as this constant shifting obfuscates functionally relevant patterns which occur at a meso-level of description.

Interestingly, MR 2.0 is compatible with the idea of fine-grained functionalism advanced in Godfrey-Smith (2016a). While Godfrey-Smith explicitly rejects the idea of MR, his arguments can be seen as targeted at the less plausible MR 1.0 and approximates what Godfrey-Smith otherwise calls "coarse-grained functionalism". The two are distinguished by the level of organization that they focus on in identifying and characterising the relevant states and processes. Godfrey-Smith accepts a multi-layered view of reality and concedes that "[t]here are reasonable coarse-grained senses of 'learn' and 'perceive' in which anything with the right coarse-grained functional profile, including a robot, does learn and perceive"

(Godfrey-Smith, 2016a: 501). However, he moves on to argue, the systems that we know to be cognitive and proto-cognitive, i.e., a variety of organisms, have an entirely different fine-grained make-up. Not only is it important that in living systems "the information processing side of its activity is integrated with the metabolic side" (Godfrey-Smith, 2016a: 502) but the small spatio-temporal scale at which cellular metabolism occurs has several unique characteristics. In particular, the cells are full of a *molecular storm* with "unending spontaneous motion [...]. Larger molecules rearrange themselves spontaneously and vibrate, and everything is bombarded by water molecules, with any larger molecule being hit by a water molecule trillions of times per second." The ubiquitous electrical charge is just one form of energy present, as chemical, kinetic, and electrostatic energy are constantly transduced into one another. Each part of the cell is subject to forces stronger than it can exert and causality is best perceived as "biasing tendencies in the storm, nudging random walks in useful directions" (Godfrey-Smith, 2016a: 485–487). Cellular metabolism arises from this material volatility and constant flux and, as Godfrey-Smith underscores, principles governing it remain crucial for the processes that constitute cognition, due to their co-evolution.

While the exact dependence of the mind on these low-level processes remains an open question, Godfrey-Smith argues that fine-grained functionalism can account for the failure of traditional functionalist approaches to understanding and engineering minds. Consider a machine—a computer—or a cyborg; even if it has similar coarse-grained functions, it will be lacking the fine-grained functions which depend on the living (i.e., far from thermodynamic equilibrium) organization of biological organisms. It may be capable of "sensing" or "learning", but these terms, or so Godfrey-Smith argues, are broad and coarse-grained, such that they do not rely on a similarity between the fine-grained functional profiles of sensing machines and sensing humans. Reality is multi-scaled and so focusing only on the scale of such coarse-grained properties will not yield the kind of understanding of cognitive processes we need to build intelligent artificial machines.

For Godfrey-Smith this view leads to a rejection of MR altogether but that is the case only for the traditional conception I call "MR 1.0". "The finer-grained features are not merely ways of realising the cognitive profile of the system. They matter in ways that can independently be identified as cognitively important", he argues (Godfrey-Smith, 2016a: 503). He indicates the inherent historicity of neurons—the change in their functional profile resulting from their own activity—as an example. This argument paves the way for Chirimuuta's upgraded notion of MR 2.0, which would hold that fine-grained functions and the material basis of cognition need to be centred in their own right, but could still, at least in principle, be multiply realised. Interestingly, a related point has in fact been a source of criticism for Godfrey-Smith's view raised by Brunet and Halina (2020), who discuss the existence of molecular machines—computers which preserve some of the low-level characteristics indicated by Godfrey-Smith—as an argument for the possibility of developing artificial sentience, which Godfrey-Smith appears to deny. However, given the discernible compatibility of Chirimuuta's MR 2.0 with Godfrey-Smith's fine-grained functionalism, it is more useful to consider Godfrey-Smith's rejection of contemporary approaches to AI to be concerned solely with their focus on coarse-grained functions.

To this list of grievances with regard to the traditional functionalist assumptions underpinning the current AI frameworks I may add one more, namely, that the coarse-grained functions they try to realise *in silico* are inherently highly complex. These are usually specific to a human way of engaging with the world, loaded with folk-psychological ideas, and disjointed from their evolutionary and developmental trajectory. In result, they are disconnected from the various scaffolds that biological intelligences use for the same purpose. This means that when trying to implement a particular psychological function in a computer AI researchers face a much more difficult problem than the one that evolution faces. If this consideration is correct, a "molecular computer" of the sort examined by Brunet and Halina (2020) would not be sufficient to be deemed a promising candidate for sentience, as developing the requisite coarse-grained functions on this platform would constitute a similarly difficult problem as in the case of silicon-based computing.

This is because Brunet and Halina's view relies on the implicit distinction between the computational "hardware" of molecular computers and cognitive "software" (problematised in the introduction and discussed in greater detail in Section How fine-grained functional details matter to cognition). They are interested in the possibility of designing "universal Brownian circuitry capable of extracting useful computation from nano-scale fluctuations" (Brunet and Halina, 2020: 233) and instantiating cognitive processes on top of this circuitry. But that means that their understanding of the functions of cognition remains coarse-grained and hence disjointed from the properties of fine-grained functions. As a result, an AI researcher working on this platform would face a problem as difficult as when working on computing platforms employing standard, von Neumann architecture. The necessary missing step, consistent with fine-grained functionalism, seems to be the use of competent, intelligent parts in the manner suggested by work within basal cognition (e.g., Levin, 2019). I will explore this view in detail in Section 'How fine-grained functional details matter to cognition'.

It is important to note that while both Godfrey-Smith and Chirimuuta leave their claims about the relevance of materiality for cognition at a pretty abstract and general level, we believe that several interconnected research fields—particularly active matter physics and soft robotics that are the focus of the current Chapter—allow for substantiating these claims further. Notably, doing so lets us draw some initial hypotheses about what "suitable organization" presumed by fine-grained functionalism could consist in and how metabolism may fit into this picture. I turn to the discussion of these disciplines in the next section.

6.3 Active Matter and Soft Robotics: Understanding the Material Underpinnings of Cognitive Sophistication

In the previous section I overviewed some of the contemporary literature on MR, specifically regarding the cognitive. I referenced the fact that some of the basic pretenses of MR 1.0 seem to have grown increasingly suspect in light of empirical advancements in the cognitive and life sciences. Indeed, the crux—for our argument—is the condition that the material configurations instantiating cognition be "suitably organized", a requirement that is a lot more stringent than proponents of MR 1.0 would allow. To this end, I began to suggest that recent developments in the areas of soft robotics and active matter physics hint that, while dimensions and aspects of cognitive systems can be manifested in alternative media, they do so *insofar* as they approximate the organizational, living, and developmental dynamics of organismic cognition—a position that I have called fine-grained functionalism.[2] Thus, this section turns toward the empirical basis for something like fine-grained functionalism to adumbrate how the material out of which embodied agents are constituted is integral to sustaining the self-organizing dynamics on which cognition depends. The main goal, then, is to suggest how a more thoroughgoing, "radically embodied" approach to AR and AI supplies the requisite tools to advance the field toward intelligent, plastic, and adaptive machines (Man and Damasio, 2019; Pishvar and Harne, 2020; Kaspar et al., 2021).

Before continuing, it is worth anticipating briefly why this approach is, or so I want to suggest, a more thoroughly embodied approach than previous iterations of embodied cognitive science. Consider how multicellular agents are themselves constituted out of highly competent, cognitive units (Baluška and Levin, 2016; Levin, 2019, 2020, 2021; Levin and Dennett, 2020; Lyon et al., 2021). In other words, the cognitive cogency of the higher-level (in this case, multicellular) agent depends on the scaling up (see Section How fine-grained functional details matter to cognition) of the cognitive processes—agency, goal-directedness, decision-making, memory, learning—found in the dynamics of constituent (somatic) cells. As we will see (in Section Soft robotics), individual cells are remarkable structures that, due to their regulatory and organizational dynamics, maintain internal milieu viability and their connectivity with other cells in the extracellular tissue complex with precision and flexibility. Reminiscent of 19th century theories of the "cell state" (Reynolds, 2007), our approach thus positions organismic cognition as emblematic of the homeostatic and self-organizing processes that typify all living units. Or, as Man and Damasio put it, "high-level cognition [is] an outgrowth of resources that originated to solve the ancient biological problem of homeostasis" (Man and Damasio, 2019: 447)—hence, cognition all the way down.

As it happens, the building of higher-level cognitive agents out of progressively smaller—but still cognitive—units and parts is precisely the perspective being taken up in the domain of soft robotics and synthetic biology. Ebrahimkhani and Levin (2021) provide a flavour of this style of argumentation:

> One feature of bioengineering at the meso-scale that is unique… is the fact that bioengineers build out of parts that are themselves highly competent, for example, cells that have their own internal homeostatic and signalling systems. Thus, the experiments that are done with biological parts have the potential to help understand how swarm intelligence plays out at the tissue level to solve morphogenetic problems. Such advances… act as an inspiration for novel architectures in machine learning, artificial intelligence, and resilient autonomous swarm robotics.

Indeed, this bio-inspired approach feeds well into current ambitions of developing AR and unconventional computing platforms (e.g., Jones, 2015). The key is how the above sciences emphasise the importance of an *active matter* approach. Rather than the inert, hard, and passive parts traditionally used in robotics, active matter approaches indicate how the very materiality of the system can perform complex feats that obviate the need for overarching or centralised control (Bechtel and Bich, 2021; Kaspar et al., 2021). In what follows, I survey the fields of active matter physics and soft robotics to then return in Section How fine-grained functional details matter to cognition to our fine-grained functionalist take on how the matter matters for life and cognition. By now, it should be clear that in arguing this position I am not being substantialists: it is not this or that type of matter (say, carbon) that is important, but the matter insofar as it can sustain organizational complexity of the right sort.

6.3.1 Active Matter as it Relates to Proto-Cognition

Although we have covered much of the AMP literature in previous chapters, it is important to recapitulate some of the basics as it is largely within the field of AMP that the active nature of robotic materials is being explored (see Baez, 2021). Recall that AMP straddles the intersection of physics and biology and deals with materials and material systems that are intrinsically out of thermodynamic equilibrium. Some examples of these are field-responsive matter, hydrogels, and piezoelectric materials, while active matter systems are those that harness properties of such materials to drive their distinctive non-equilibrium behaviour. These include the actomyosin cytoskeleton (Needleman and Dogic, 2017; Jülicher et al., 2018), cellular activities (Fodor and Marchetti, 2018), swarming behaviour (biofilms, multicellular bodies: Wioland et al., 2016; Kempf et al., 2019), and even macroscale organizations such as avian murmurations and herds of animals (Cichos et al., 2020). Although this might appear to be a heterogeneous set, these systems exhibit the broad commonality that their individual units (motor proteins, cells, individual organisms) are themselves highly competent, active contributors to group dynamics (Needleman and Dogic, 2017).

For example, it is increasingly common to view multicellular bodies as a kind of swarm behaviour (Arias Del Angel et al., 2020), which depends on the intrinsically active nature of constituent cells. Indeed, Arias Del Angel et al. (2020) have commented on how facultative multicellularity in both protists and prokaryotes depends on active, field-responsive, and internally driven physical processes of constituent parts, remarking that the overall organismic form hinges on the interplay of the inherent physical properties and agent-like competency of cells making decisions in a context-sensitive and flexible manner. In contrast, then, to passive systems (e.g., the Rayleigh-Bénard cell) that receive energy exogenously at a boundary condition, active matter systems—of which organisms and certain designed systems are paragons—themselves consist of units that are internally driven (Batterman, 2021). Crucially, Needleman and Dogic (2017) remark that active units are capable of *self-*

organization, whereas passive units can only *self-assemble*. In the context of being "suitably organized" what we see is that not any back-of-the-envelope set of materials can sustain the organization dynamics on which life and, I add, cognitive processes depend—instead, to be suitably organized one must have self-organization, and it is here that an active matter approach is most pertinent.

Seeing how the dichotomy of active and passive structures underpins much of the literature in AMP, it is worth explicating further what marks out the former exactly. In their influential review of AMP, Marchetti et al. (2013) write that active matter systems consist of the following features:

> They are composed of self-driven units… each capable of converting stored or ambient free energy into systematic movement. The interaction of the active particles with each other, and with the medium they live in, give rise to highly correlated collective motion and mechanical stress. Active particles are generally elongated and their direction of self-propulsion is set by their own anisotropy rather than fixed by an external field. (Marchetti et al., 2013: 1144)

The key distinction I wish to draw out here is that being an active matter system relies on two features: (i) the energetic nature of the constituent units (actively converting ambient energy as opposed to being driven solely by energetic contributions at an external boundary conditions) and (ii) their inherent shape (anisotropy) influencing the systematicity or directionality of how energy is used—"geometry of [its] interface shape can control sensitivity to the environment" (Hanczyc and Ikegami, 2010).

More recently, active materials have been exploited in soft robotics (Ebrahimkhani and Levin, 2021), computer science (Jones, 2015), and AI (Kaspar et al., 2021) as a way to overcome the many resource constraints that have long plagued the fields. The key point can be expressed as follows: "The cellular cytoskeleton, cells, and entire tissues [as exemplary active materials] are driven away from equilibrium by the continuous motion of thousands of constituent nanoscale molecular motors, protein-based machines that transform chemical energy into mechanical motion" (Needleman and Dogic, 2017: 2). Intriguingly, this is a point that has echoes in Section Traditional vs. fine-grained functionalism in our discussion of fine-grained functionalism and the relevance of spatial *scale*: "Metabolic processes in cells occur at a specific spatial scale, the scale measured in nanometres… In that context and at that scale, matter behaves differently than how it behaves elsewhere. … There is unending spontaneous motion that does not need to be powered by anything external" (Godfrey-Smith, 2016a: 485). At larger, more coarse-grained scales, these complex and systematic processes do not occur. Already, then, we come to see how fine-grained structural details matter for sustaining self-organizing dynamics at a wider variety of scales.

Of course, what is central to the discussion of fine-grained functionalism is the connection between these active material processes and prototypical instances of cognition, such as goal-directedness, memory, learning, agency, systematic directionality, and so on. As it happens, recent work on active materials has begun to show the variety of ways in which some individual—and sometimes multiple—capacities are present in non-living systems, a discovery that has led some to speculate that AMP is revealing not only the physics of life (Popkin, 2016), but the physics of cognition as well (McGivern, 2020). To wrap up the discussion of AMP, then, we make a more direct connection to work on basal cognition and the concept of existential needs introduced above.

Capacities of non-living active matter systems that have been particularly illuminating are those of autonomous movement, environmental sensing, coordinated action, and problem solving (McGivern, 2020). The ability to accomplish these feats importantly depends on the material situation of both the system in question (swarming nanobots, self-propelled oil droplets) and the environment where it finds itself. In self-propelled oil droplets, for example, researchers introduce internal convection currents that create a bifurcation between systematic internal activity and its viscous medium. The droplet's movement is driven by a convective flow that has an uneven influence on the inside of the droplet, which helps create a feedback system between its internal dynamics and the medium external to it (so-called Marangoni flows; see Hanczyc and Ikegami, 2010 for a review). Although this is a simple system, Hanczyc and Ikegami (2010) suggest that it serves as a model system for understanding the origins of chemotaxis in unicellular organisms, as the droplet must continuously navigate gradients to find the chemicals that sustain its internally driven dynamics: "The system becomes sustainable by circulating the reactants and products effectively as organized by the convective flow" (Hanczyc and Ikegami, 2010: 236).

As we can see, it meets both conditions of active matter listed above: (i) its constituent particles are internally energetically driven in that they tap into reservoirs of (Gibbs) free energy available within the system, and (ii) its droplet shape results from the inherent properties of the oil (its viscosity and surface tension) and its relation to the medium in which it is embedded; moreover, the geometric configuration of the oil droplet actively contributes to the distinctive capacities it exhibits. This is an intriguing model system for our purposes, as it places a premium on active materials and constituent units that spontaneously self-organize and, given the right guidance and influence from designed experimental parameters, can sustain itself for significant periods of time. Although not elaborated here, the case of oil-droplets also underlines the way in which the inherent shape ("geometry-induced fluctuations"; Hanczyc and Ikegami, 2010) of an active unit determines locomotion and, in bacteria, chemotaxis. Moreover, a mechanical pushing of the cytoplasmic sol of a cell (as in the social amoeba *Dictyostelium*) elicits directional and coordinated motion (see Dalous et al., 2008; Boussard et al., 2021). Thus, the properties exhibited in active matter systems, such as oil droplets, highlight the material basis for capacities found throughout the living domain.

AMP shifts our focus on the study and development of minimally cognitive systems (that is, systems that exhibit prototypical features of cognition such as directional locomotion, memory, or learning) in two important ways. First, it does not aim to replicate paradigmatically intelligent behaviour modelled on human activity (playing chess, say) and instead emphasises environmentally embedded behaviour with wide distribution in the natural world (McGivern, 2020). Secondly, and perhaps most centrally for our argument on MR 2.0, "work on active materials is not specifically aimed at computational characterisations of behaviour" (McGivern, 2020: 442), i.e., it does not rely on coarse-grained functions of the medium of interest, but rather builds on simpler, well-established principles from areas such as condensed matter physics, building a bottom-up description of activities of systems of interest. Work within AMP, then, demonstrates how harnessing the physical processes and active materials that underlie organismic behaviour contributes to and mediates the cognitive sophistication we find in the biological domain—suggestive of how such principles can, and are, being exploited in the domain of artificial and designed soft robotic systems, which I turn to shortly.

Before continuing, however, it is worth dwelling on the aspect of AMP that I see as central to the discussion of cognition that forms the remainder of this chapter. Recall from the introduction that our understanding of cognition revolves around the fulcrum of existential needs and how capacities such as agency, goal-directedness, and self-maintenance are the basis for further cognitive sophistication. Matthew Egbert has recently argued that non-biological model systems—such as our humble oil droplet—serve as ideal testbeds for exploring the material and thermodynamic basis of these existential needs: "conditions that must be met for [that system] to persist and… behave in ways that satisfy those needs" (Egbert, 2021: 5).

There are two ways to understand existential needs vis-à-vis any object or system, the first rather banal and the second more critical for the kinds of systems I explore in this thesis. The first is the sense in which, trivially, any object must have existential needs to be what it is. A table cannot be heated above a certain temperature or subject to a certain amount of pressure and still remain a table. But, and this is the more important point, there are crucial differences between what is required of garden-variety non-dissipative objects like rocks, tables, and chairs to be what they are and self-organizing, self-maintaining dissipative systems in far-from-equilibrium conditions. The difference in existential needs for the two types of systems is captured as follows:

> [Non-dissipative entities] are merely passively stable, whereas dissipative structures are constantly falling apart and yet persist thanks to processes of repair, replacement, or reconstruction. This means that existence for passively stable entities is the absence of a destructive event. In contrast, for dissipative structures, existing is a process—and a process that must continue for the system to persist. (Egbert, 2021: 5)

Importantly, processes have quantifiable and measurable rates that open dissipative structures to a study of how viable such a system is, that is, how well it persists despite the tendency to degrade. As Egbert notes, there is no equivalent measurement for passively stable systems: their existence is not a process and does not require the same set of behaviours and activities that active matter systems engage in. I therefore agree with Godfrey-Smith when he writes "macroscopic machines provide a poor model for the material basis of living activity and for the material basis of mental activity in living beings like us" (Godfrey-Smith, 2016a: 489).

Our discussion of AMP furnishes one strand of the argument for fine-grained functionalism, namely, that the fine-grained material and thermodynamic details of living systems matter a great deal more than common assumptions on the MR of the cognitive might *prima facie* suggest. Indeed, organisms are subject to what physicists call the "tyranny of scales"—they are sensitive to, and influenced at, every order of scale, from the nanoscale to the mesoscale, and for multicellular agents like ourselves, the macroscopic scale. These are highly sensitive coordinated structures, and there is no non-arbitrary point below which the physics no longer matters to manifesting the distinctive cognitive capacities that contribute to a living system's survival. Although our discussion of cognition has been minimal in this section, I turn now to soft robotics to see how these insights are being actively taken up in designing intelligent synthetic machines.

6.3.2 Soft Robotics

Soft robotics is a sub-discipline of robotics and artificial intelligence that explores how intelligent, adaptive, and plastic behaviour emerges out of the inherently active, precarious, and soft parts that constitute such systems. It is a discipline that examines how to construct systems that exploit the physical laws and tendencies at play at every level of scale. In other words, it investigates how organisms are embedded and subjected to a "tyranny of scales" that must be accommodated and exploited to meet their existential needs (Ebrahimkhani and Levin, 2021) and how one may apply these insights to the creation of intelligent machines.

Tellingly, Shah et al. remark that the inspiration for soft bodied robots comes from the highly integrated nature of biological cognition: "In these integrated living systems, intelligence, memory, learning, behaviour, and body structure are all intertwined and emerge from the multiscale dynamics of the same robust and highly fault-tolerant medium" (Shah et al., 2021: 1). This is put in contrast with the standard hard (and passive) components that constitute more standard roboticist approaches. Standard approaches have had some success in the form of modular parts that can be added or taken away depending on the task (such as passive conforming grippers and certain algorithms that can re-adapt to distinct tasks). But even in cases where these techniques might achieve certain adaptive ends, they "operate under the assumption that the robot's body is only reconfigured or reshaped due to external forces, and do not explore the possibility of synthetic machines that actively grow, regenerate, deform, or otherwise change the resting shape of their constituent components" (Shah et al., 2021: 2). Contrastively, as we saw above, the field of AMP begins to highlight the way in which organism morphologies and bodies are inherently active structures that respond proactively to changing environmental situations—a form of adaptiveness that depends crucially on the highly active processes that comprise cellular structures, multicellular integration, and cognitive capacities such as goal-directedness and agency. This picks up on a point made earlier: a key feature of organismic cognition and an insight that has been actively taken up in soft robotics is that higher-level cognition relies on constituent parts that are themselves *highly* competent.

Here, I focus on how the concept of existential needs, raised in the introduction, is critical for the creation of artificial machines capable of autonomously selecting actions required for self-maintenance. In other words, in contrast to passively maintained robots that must be externally guided and directed toward goals, tasks, or functions, it is suggested that (i) the inherent vulnerability of soft embodiment coupled with (ii) thermodynamic processes that are required to maintain a system in such a state would endow an artificial agent with the kind of autonomous self-maintenance and self-organization that are important for cognition (cf. Bickhard, 1993). Only then would these designed systems have real "skin in the game" (cf. Bongard and Levin, 2021). To put the matter differently, to design machines capable of autonomous decision-making, behaviours must have *consequences* for how the system can and should act in the world. I have introduced this idea previously in terms of precarity and risk-to-self (in Section Introduction), and with the analysis of active matter above we may further specify the details of what precarity would mean for a machine. Man and Damasio (2019) indicate, in terms intriguingly close to Egbert's paper cited above, that we can understand how feelings emerge from a physiological investigation of life regulation. Feelings, they argue, are not sufficiently approximated by arbitrary reward or loss functions of standard approaches to AI, since the worldly risks and consequences should directly impact the continued existence of the machine. The quality of feeling "is the harbinger of the good or bad outcome relative to survival" (Man and Damasio, 2019: 446). They argue—and I concur—that it is only at the point when the machine can consistently strive for continued existence that true agency may arise.

It is important to note that what I mean here by "life regulation" is not biology-restricted, but rather is the upshot of a far-from-equilibrium system working against the tendency toward dispersal. The suggestion I would like to make here is that to be this kind of system—a system for which there can be situations that matter to it—it must be a self-organizing one constituted by active physical processes inherent to the materiality of the system in question. Soft robotics, then, is in the business of identifying how the material aspects of the body exploit physical laws to expand robot functionality (Shah et al., 2021: 2).

For example, Pishvar and Harne (2020) note that soft robotics incorporates field-responsive smart matter that can induce an internal flux in response to an applied field that tailors material characteristics of the media, influencing its function and behaviour. As they write, "When responding to applied fields, a multitude of internal changes are possible in soft, smart matter" (Pishvar and Harne, 2020: 1). The range of adaptability is thus expanded when one incorporates material properties that are themselves active contributors to overall robot functionality, in contrast to standard hard parts used in robotics, whose adaptability—in the rare cases when they are adaptable—is due to pressure driven forces at an external boundary condition.

More recently, Kaspar et al. have argued that "synthetic matter that itself shows basic features of intelligence would constitute an entirely new concept for AI" (Kaspar et al., 2021: 345). They dub this pivot in AI and robotics the "rise of intelligent matter" and reiterate the point that incorporating active materials into AI and robotics programmes would expand robot functionality "far beyond the properties of static matter" (Kaspar et al., 2021: 345). Examples of such smart, active matter systems include artificial thermoregulating skin (Kanao et al., 2015), emergent swarming activity of concerted nanobots (Wu et al., 2021), and xenobots that sit at the intersection of bio- and artificial engineering (Ebrahimkhani and Levin, 2021; Kriegman et al., 2020, 2021; Blackiston et al., 2021). All authors appear to be in agreement that incorporating the smart and active propensities of soft matter is crucial for achieving autonomous behaviour in the domain of robotics (Pishvar and Harne, 2020; Kaspar et al., 2021)—a suggestion I will turn toward in the next section on the fine-grained functionalism approach to AR.

To wrap up briefly, the fields of active matter physics and soft robotics are working lock and step to uncover the diverse functionality, adaptability, and plasticity inherent to certain materials that remain in far-from-equilibrium conditions. They are thus fields that explicitly consider the thermodynamic situation of the system in question. An upshot of this is that not any sort of material can accomplish the diverse behaviour or cognitive sophistication exhibited in the biological domain. In other words, to be "suitably organized" requires attention to the media out of which the system of interest is constituted: the matter matters for cognition and is not a dimension of robot functionality that can be abstracted out. Indeed, the conclusion I wish to draw from this literature is that *more* attention should be paid to the material basis of cognition than is commonly done.

Given the importance of the above two testbeds for exploring the nature of cognition, it is crucial to explicitly articulate the connections that can be drawn between active matter physics and soft robotics. Our reasoning for progressing from the former to the latter is that active matter physics deals with far-from-equilibrium dynamical systems, writ large, and the materials and material constellations that can sustain self-organizing processes on time scales relevant to the biological world. It is precisely these processes that are then exploited and harnessed in guided assembly to arrive at the sophistication we find in the field of soft robotics (Ebrahimkhani and Levin, 2021). While prima facie it might appear that the two fields can work in isolation, Pfeifer et al. suggest why this is not advisable: "it [is] clear that autonomous agents display self-organization and emergence at multiple levels: at the level of induction of sensory stimulation, movement generation, exploitation of morphological and material properties, and interaction between individual modules and entire agents" (Pfeifer et al., 2007: 1088)[4]. In other words, active matter physics in tandem with soft robotics furnishes not only the empirical testbeds for crafting more sophisticated autonomous agents, but also renders tractable notions of emergence and self-organization that are of relevance not only to bodily maintenance and self-preservation, but also cognition and the behaviour required to keep such systems viable. In the following section, then, I dovetail the pieces of the argument laid out thus far to advocate for an emerging approach to the development of AI and AR that stems from fine-grained functionalism suggested in Section Traditional vs. fine-grained functionalism and the results of AMP and soft robotics discussed in the current section; a paradigm that appreciates the importance of the materiality of cognition.

6.4 Fine-Grained Functionalism and New Material Paradigms for Robotics and AI.

Recall that one thing I wish to explore in this thesis is the possibility of developing autonomous robots capable of prototypical forms of valuing and engaging with the world in a goal-directed manner. To this end, I adopted the notion of precarity (i.e., "risk-to-self") and focused on the existential needs of a system to remain in a far-from-equilibrium state. We saw that a step in this direction requires rethinking some of our basic assumptions regarding matter and its relation to cognition, which remain deeply embedded in existing approaches to the mind and brain sciences, as well as in approaches to AI and robotics. Indeed, rather than a "layered-cake" model of levels that renders higher-level cognitive phenomena as resting autonomously from and "on top" of its substrate (i.e., "hardware"), I set out to complicate this picture by emphasising the (bio)physical nature of the structures that support, enable, and implement cognition. What I want to suggest is that more attention must be paid to the *fine-grained* details of the system when understanding and studying cognition—and then recreating it in alternative media. This is the crux of the fine-grained functionalism introduced in Section Traditional vs. fine-grained functionalism. Although it is common to see biologists and philosophers emphasise the "Heraclitean" nature of biological matter and metabolism, I believe recourse to the fields of active matter physics and soft robotics situates fine-grained functionalist views on a sturdied empirical testbed. Thus, the developments in these disciplines enable an exploration and substantiation of claims about precarity and existential needs vis-à-vis cognition, which so far I have explored mostly in the abstract.

In this section, I weave the threads of the argument together to argue that creating AR capable of valence and goal-directedness requires us to think about the organizational and material dynamics of the embodied system in a more thoroughgoing way. In other words, what I am suggesting is not (or not simply) that the body *simpliciter* matters to cognition, as advocates of sensorimotor coordination have long held (see Van Duijn et al., 2006). Rather, I argue for a multiscale account in which cognitive and agent-like competency is present at nearly every level of a biological heterarchy capable of sustaining the appropriate organization—cells, tissues, networks, the whole organism, and even swarming behaviour of eusocial species (Levin, 2019). In contrast to views that situate cognition as exclusively proprietary to a higher-level organism, I present an account in which the scale and "selfhood" of the cognitive agent are highly malleable, plastic, and vacillatory.

I therefore argue for taking embodied approaches further than is commonly done in two important respects. First, in the multiscale approach just outlined: higher-level systems (organisms or future soft robots) themselves consist of highly active and competent cognitive units. The preservation of cognitive functionality at varying spatiotemporal scales is indeed a crucial aspect of evolvability and robustness in organisms (Levin, 2020). Cognition is then regarded as "an outgrowth of resources that originated to solve the ancient biological problem of homeostasis" (Man and Damasio, 2019: 447). It is construed as an activity of self-organizing and self-maintaining processes fundamental to all living organisms, and that is then appropriately scaled up throughout a biological heterarchy, an idea I explore further below. Second, and here I loop back to fine-grained functionalism, these cognitive capacities depend crucially on the material and (bio)physical details that are standardly abstracted out or relegated to a "hardware" problem. These two central themes are discussed in the remainder of this section.

6.4.1 Scaling Cognition Down—And Back Up Again

Organismic embodiment is characterised by highly plastic and adaptive parts responding in a coordinated manner to wider organism-level goals. Empirically, this increasingly seems to rely on essentially cognitive units and intelligent parts—i.e., cells, tissues, networks—acting in a concerted manner that involves an "inter-penetrating, concurrent operation of numerous layers of cognition within the same living system" (Levin, 2021: 4); that is, it involves cognitive units maintaining some degree of flexibility, agency, and goal-directedness that is executed in local and global contexts. In biology, this is often on display in the morphogenetic and ontogenetic unfolding of the organism—a complex process that requires cognition to be tailored to both scale-specific as well as scale-free needs in regulating organism development. Indeed, the ability of organisms to plastically change shape throughout their life cycle is the current envy of soft roboticists, where developing shape-changing robots is a frontier in the field (Shah et al., 2021). Here, I explore the phenomenon of shape-shifting, as it helps illustrate how morphogenetic and homeostatic goals pursued at each level can give rise to robust and flexible behaviour at a variety of scales.

As suggested, an organism's ability to arrive at complex morphogenetic outcomes depends on the interpenetration of these functionalities at a range of spatial and temporal scales, as well as the elasticity and robustness of a (predominantly soft) medium. This contrasts with standard roboticist approaches that incorporate hard parts ["up-armouring", as Man and Damasio (2019) phrase it] and assumes that bodies are only reconfigured due to external forces, effectively neglecting the active and proactive responsiveness that typifies biological media and serves as the foundation for homeostatic, self-organizing processes. Indeed, hard-clad robots might experience change at the movement of a joint, but none within the stuff that constitutes it. Contrastively, biological and soft robotics systems "[change] shape at all relevant scales, globally and locally" (Shah et al., 2021: 10). What is important here is that this process is effectuated through the nested hierarchical structure in which every level can pursue its own local (morphogenetic) goals. The morphogenetic (shape-shifting) outcome of this process is thus not only materially and physically active, but an expression of the cognitive coordination to be found throughout the organism. This is in sharp contrast with current robotics, which largely uses unintelligent parts (Shah et al., 2021).

We can call these systems exemplars of "coordinated structures", following Kelso (2016), which are endogenously self-organized systems determined by their own dynamics. Indeed, a characteristic feature of such structures is that they do not depend on an exogenous "ordering influence" (Kelso, 2016: 491), and some have remarked on how this form of self-organization is the basis for higher-level features of autonomy, agency, and goal-directedness (Juarrero, 2015). Perhaps unsurprisingly, the requirements for coordinated structures are parallel to the defining features of active matter systems, suggestive of the relevant building blocks for engineering artificial analogs that could come to endogenously self-organize to create novel, agentic, and goal-directed structures. It should be clear that the vision of cognition I have in mind here is one in which the system itself has real "skin in the game", and therefore requires this minimum degree of autonomy (Bechtel and Bich, 2021).

Importantly, this focus on internal coordination echoes the prominent view that cognition—as it evolved—initially emerged in the course of evolution for coordinating cellular metabolisms and ultimately multicellular (more minimally, intercellular) activity, particularly spatial and temporal coordination across parts of the system—a position Keijzer et al. call the "Skin Brain thesis" (Keijzer et al., 2013; Jékely et al., 2015). The primacy of internal coordination hints at the profound relevance of electrical oscillatory activity found in biological bodies (cf. SELFOs, see Hanson, 2021), which has been put forward as one of the central mechanisms of synchronization—an important topic that future work will explore in detail. Furthermore, the path of engineering intelligent systems from self-organizing and coordinating intelligent parts, while perhaps not the only one, becomes a clearly feasible approach for researchers, since we see that in the history of life this trajectory has in fact led to the emergence of cognition.

The shifting local and global coordination of the system (i.e., organism) exemplifies the distributed approach to cognition I have defended herein. It requires that constituent cells and parts maintain certain aspects of cognitive function—memory, learning, agency, decision-making—at least in the service of their own form and function. Levin captures this point well when he writes that "somatic cells did not *lose* their behavioural plasticity … to become parts of metazoan swarms (bodies): they scaled them to enable pursuit of larger goals consisting of creation and upkeep of massively complex anatomies" (Levin, 2019: 5).

Thus, the concept of scaling up, which I have relied on throughout this Chapter, rests on the idea that multicellularity is itself a highly complex and competitive "environment" that requires local and global morphogenetic goals consisting of trade-offs and top-down constraints between small-scale outcomes and organismic level development. I have already suggested that we can understand this principle of scaling up in terms of internal coordination determined by endogenous dynamics but building on the concept of active matter can help elaborate the idea further.

We can identify an appropriately scaled-up active matter system when two conditions are met: (1) the system is not wholly determined by local causes (see Kelso, 2016), that is, the system behaves in relation to non-local causes; and (2) it exhibits goal-directedness as a coherent unit. Internal coordination results from the conjunction of these conditions and, hence, is inseparable from cognition as I explore it here. Our reliance on active matter is motivated by our fine-grained functionalist claims, which I turn to below.

For now, it is important to emphasise that developmental bioelectricity has been identified as the predominant mechanism behind locally and globally coordinated morphogenetic, developmental, and cognitive outcomes—realising the scaling up of parts into wholes. Bioelectrically coordinated and integrated cells (in the form of an organism or a colony of organisms, as in bacterial biofilms) meet the conditions for a scaled up system laid out above in that (1) the activity of cells often results from information about occurrences happening in a distant part of the integrated system and (2) each part coordinates its actions with others, so that the system as a whole exhibits a consistent behavioural pattern (see Arias Del Angel et al., 2020 for an insightful discussion of this in relation to the social amoeba *Dictyostelium*). Thus, the bioelectrical activity that has often been associated with nervous activity is increasingly seen as an exploitation of highly preserved, ancient, and widely distributed cellular functions and capacities (Prindle et al., 2015)—and I extend the discussion to suggest that this itself hinges on more general properties of cellular, biological, and living material dynamics. This is what in developmental biology (Newman, 2019, 2022, 2021) has been called "biogeneric" processes, indicative of how biological functionality in the service of homeostatic, morphogenetic, and developmental goals is an exploitation of general physical principles of viscoelastic media and oscillatory activity. This again draws a strong connection between the "physics of life" and "physics of cognition" at which I hinted at above.

Indeed, if—as this research suggests—neurons are specialised exploitations of bioelectrical mechanisms, it is more fitting to see the nervous system as initially (both *evolutionarily* and *ontogenetically*) more a matter of "pulling the organism together" than as specialised for higher-level cognitive functions (cf. Fields et al., 2020). Again, pulling from Levin, we see that "neural networks control the movement of a body in three-dimensional space; this scheme may be an evolutionary exaptation and speed-optimization of a more ancient, slower role of bioelectrical signalling: the movement of body configuration through anatomical morphospace during embryogenesis, repair, and remodelling" (Levin, 2019: 5).

The truly innovative move in the literature on basal cognition (that is, cognition as situated in more "primitive" organisms and cellular activities), then, is the explicit recognition of the cognitive (or proto-cognitive; Godfrey-Smith, 2016a,b, 2017) nature of the activities identified above. Indeed, examples of memory in social bacteria (Dinet et al., 2021), learning in unicellulars and protists (Gershman et al., 2021), decision-making in acellular and cellular slime moulds (Arias Del Angel et al., 2020; Smith-Ferguson and Beekman, 2020; Boussard et al., 2021) have all been identified in non-neural organisms, and it is known that constituent cells in metazoan swarms (i.e., multicellular animals) actively and adaptively manage their morphology, behaviour, and physiology as needed for survival. Again, this is cognition within and throughout biological bodies and therefore is suggestive of a more thoroughly embodied cognition insofar as higher-order organized wholes are dependent on constituent units and parts maintaining, in certain crucial respects, cognitive capacities of far more ancient origins. The ability of evolution and hence organisms to exploit the material properties of cellular processes to yield coordinated wholes is the current envy of soft robotics approaches that still rely on guided self-assembly to arrive at robot functionality (Ebrahimkhani and Levin, 2021). As I have already hinted, this differs dramatically from extant AI and robotics approaches that do not avail themselves of such techniques, in that goals can be pursued both at the wider level of the whole organism, at the tissue complex level, and the level of cellular homeostasis and intercellular coordination. Indeed, "the ability of each nested level to have its own local morphogenetic goals… contrasts with today's robots, which are largely made of unintelligent parts" (Shah et al., 2021: 10). To conclude, then, I loop back to our fine-grained functionalism claims to highlight the close imbrication between cognitive capacities of interest in the design of AR (agency, goal-directedness, memory, learning, self-maintenance) and fine-grained aspects of the materials that should, I suggest, be the focus of current and coming robotics approaches.

6.4.2 Fine-grained functions of soft materials

Fine-grained functionalism rests on the crucial observation that cognition is not temperature. Allow us to explain. When we approach cognitive systems and try to individuate the functions they perform, we always do so at a particular level of granularity that reflects certain aspects of the observer (their interests, needs, pragmatics, assumptions) and not the cognitive system observed. Philosopher of science Angela Potochnik phrases this as a matter of reading our assumptions of the multilevel nature of the world *into* the phenomena of investigation, imposing an artificial hierarchy on a complex system where there may not be one (cf. Bechtel and Bich, 2021; Potochnik, 2021). Crucially, these different granularities do not simply map onto the distinction between macro- and microstates that some branches of physics find useful. Cognitive functions, such as learning, memory, decision-making, are not macrostates realised by (possibly very different) physical microstates in the way that the same temperature can be realised by various distributions of thermal energy across molecules—even if we may observe these functions in equal part in a variety of natural and engineered systems.

What I mean to say, then, is that non-biological passive materials (in our case, materials that cannot sustain self-organizing dynamics in far-from-equilibrium conditions) will *not do* the same things as soft biological counterparts: "They will be functionally different, not merely different in "hardware" or "make up"" (Godfrey-Smith, 2016a: 501). For functional equivalence, their material structure and organization must occupy specific spatial and temporal scales to endogenously accomplish self-maintenance and self-organization. The main upshot of this view is that fine-grained functional properties of living systems, such as metabolism and recursive self-maintenance, matter quite a deal more than is commonly supposed in debates on the MR of the cognitive. In other words, if cognition depends on suitably organized, endogenously driven internal dynamics of cellular activity and the appropriate scaling up into even smarter wholes, we begin to see the way such details become the foundation of the cognitive—and, by extension, central to the approach to engineering AR that I advocate. There are two considerations with which I would like to conclude.

First, fine-grained functionalism imposes clear constraints on what sorts of materials are capable of instantiating cognition—without falling into the biological chauvinism typically (and erroneously) associated with this type of view. The required platform must be able to sustain self-organizing dynamics in far-from-equilibrium conditions on temporal and spatial scales that make it susceptible to physical forces, constraints, and tendencies that are not found at larger spatial scales—the scale of standard machines to which biological cognition is traditionally compared (Nicholson, 2014). These conditions are met by soft, active materials: a domain of materials science that continues to grow in popularity since its inception in the 1990s. From a physicist's perspective, exemplary soft materials such as "[c]olloids, polymers and surfactants, sometimes also known as 'complex fluids', have one characteristic in common: they involve a mesoscopic length scale between the atomic (~1 nm) and the bulk (~1 mm). On this intermediate length scale, one finds structures such as suspended particles/droplets, macromolecular coils, and self-assembled structures such as micelles and bilayers" (Poon, 2000). The ability to self-assemble into vesicles is especially interesting, as, according to some researchers (see Kauffman, 1993), such structures form a necessary step in the emergence of life, since they allow for the prebiotic system enclosed within to control its interactions with molecules in the environment and, in result, to remain at the boundary between subcritical and supracritical behaviour. As stated throughout, I do not preclude the possibility of non-living cognitive systems. Indeed, crucial points of our argument turn on the blurring of cherished distinctions between paradigmatically living and non-living systems.

What I do want to highlight, however, is that a non-living system is cognitive the more it approximates dynamical features of living activity—that such activities (which I broadly associate with self-maintenance and, eventually, homeostasis *via* metabolic activity) are the fount from which higher-level cognition emerges. The overlap in the dynamics between living and non-living systems constrains the types of materials that can enter a concerted organization able to sustain itself recursively and endogenously. While common reference points in philosophical debates on MR 1.0 consist of cognition being instantiated by (*inter alia*) tumbling beer bottles, frenzied radio signalling between denizens of the Chinese nation, and, of course, Swiss cheese, it is clear from what I have argued that these are *not* the kinds of things that can sustain self-organization endogenously. Matter behaves differently at the scale of objects normally invoked to support intuitions on MR, reaffirming the point expressed above that these materials will not do the same thing as the molecular motors, nanoscale molecules, and field-responsive materials we find at length scales well below that of everyday familiarity. Dislodging our intuitions about the MR of the cognitive, and upgrading from MR 1.0 to MR 2.0, enables us to attentively observe the behaviour of matter at nano and mesoscales to more properly assess how proto-cognitive capacities relate to the frenzied activity of fine-grained features of the system—not treating them as "noise" or obfuscating complexity to be abstracted out.

What I do want to highlight, then, is that an active matter approach oriented around soft materials could begin to approximate these features in non-living systems and media—as the exciting field of soft robotics is beginning to show (Section Soft robotics above). Hence, an attempt to design and engineer an artificial cognitive system in such materials would tend to fall closer to cognition than extant hard-part robotic systems. Soft materials and our active matter lens provide (some of) the resources to better assess the "suitably organized" claim so often made in debates on multiple realisability. As already mentioned, this allows us to resist biological chauvinism worries, while also delimiting the kinds and configurations of systems that can be autonomous cognitive agents, hence neutralising the liberalism charge as well. That is, as stated earlier, the active matter lens allows us to argue *both* that the materiality of cognition matters *and* that the cognitive can be realised in alternative media (Chirimuuta, 2018; Brunet and Halina, 2020).

These considerations bring us to the second important insight granted by the perspective of fine-grained functionalism. It constrains how we should approach the task of engineering AR, defining a feasible—at least so I hope—research strategy. Developing artificial cognition once we have rejected the "hardware/software" distinction renders the concept of Artificial General Intelligence (AGI)—a Holy Grail of present-day AI researchers (explicitly embraced by companies such as OpenAI and DeepMind)—misguided. AGI can be understood as "loosely speaking, AI systems that possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation" (Goertzel and Pennachin, 2007). The view that emerges from the "cognition all the way down" approach is that cognition is not "General" in this sense—

cognition is not a single programme that can be applied to a variety of contexts, in the way that the programme MuZero (Schrittwieser et al., 2020) is able to master a variety of video and traditional games without explicit presentation of the rules. Rather, cognition results from the orchestration of a vast amount of single-purpose, specialised processes that co-depend on each other across spatial and temporal scales—single cells coming together into larger and larger ensembles. These processes undergo constant rearrangements and shifts, balancing on the boundary of criticality, striving to remain far from thermodynamic equilibrium. What “Generality” the system and its parts exhibit results from constant flux, from its Heraclitean nature, where constant change is required to remain in the same place.

Hence, the task of engineering AR must not be approached from the top-down, and not only because of the high computational complexity of coarse-grained cognitive functions (discussed in Section Traditional vs. fine-grained functionalism). Soft materials need to be engineered into simple “proto-cognitive” units which then need to be scaled up into higher-level systems. While I believe that to a large degree appropriate scaling up requires self-organization, the researcher still remains largely in control of this process, as they can influence and shape the fitness landscape of emerging autonomous embodied robots, guiding them toward meta-stable states that they deem beneficial or useful. In fact, to a degree even an external re-arrangement of the emerging self-organized system may be enough to push it in a particular direction, in a manner similar to how surgical intervention into grown tissue makes possible the creation of xenobots (Kriegman et al., 2020).

This would mean that the task of developing AR doing a wide range of things—whether that would be driving cars, repairing spaceships, performing surgeries, or accompanying us at the table—is likely beyond the limits of what is attainable in the lifetime of the current generation of AI researchers. I believe, however, that the strategy remains similar whether one focuses on this kind of blue-sky research, or rather seeks to achieve more proximal goals that are already stated in the literature among the things engineers are working toward. These more feasible applications, specifically in the case of xenobots, include “intelligent drug delivery”, “internal surgery”, identifying cancer or processing of toxic waste products (listed by Kriegman et al., 2020), as well as “cleaning microfluidic chambers” and “environmental sensing” (suggested by Blackiston et al., 2021). The common approach to the development of such machines focuses on what conditions would be required for the system to believe this task to be "good"—not in terms of arbitrary reward functions, but in terms of risks and opportunities, or fitness landscapes.

One way to accomplish this goal may be in parallel to raising and educating a child (cf. Ciaunica et al., 2021). In contrast to standard approaches in contemporary AI research, which may be more accurately compared with operant conditioning, raising a child consists more in creating—and removing—affordances in the social and physical environment of the baby. We create opportunities, control some of the risks, but in the end it is the child that must take up any particular affordance in order to best learn it. We reward, correct, and punish, but most often we do so implicitly, by accident, and to a much lesser degree than in the case of AI systems. These sparse rewards can be taken to serve more to structure the fitness landscape that the child explores, to boost its internal reward and motivation systems, than to provide a reward or loss function that learning can entirely depend upon.

The approach toward AR I suggest is similar. In the—paradoxically—simplest case where we rely on living soft materials as building blocks, we can observe an application of this strategy in the case of the aforementioned xenobots. In a virtual cyborg-like setup, they explore in simulation their expected fitness landscape guided by some simple tasks and then, *in vivo*, the simple self-organized structure is finessed through external means. The resulting living robot is capable of surprisingly complex behavioural feats, as it forages throughout its simple environment on a Petri dish, coordinating its behaviour with others, and—when presented with an opportunity of interacting with "naive" stem cells—replicating into active organisms, similar in form and abilities (Kriegman et al., 2021). Xenobots, in fact, offer an initial hint that an approach along the lines I suggested here is feasible and may well lead to the development of workable AR, even if with only limited applications to begin with.

In sum, transitioning toward AR capable of selecting their own goals requires incorporating "dynamic materials that possess a substantial degree of conformational freedom, mobility, and exchange of nanoscale components" (Kaspar et al., 2021: 353). In other words, developing these autonomous systems calls for attending to the fine-grained functional profile of embodied active materials that do not themselves depend on exogenous control. This is in fact one of the main outstanding problems in synthetic biology and soft robotics, as current model systems are not capable of self-organizing in a coordinated manner across the nano and mesoscales, instead relying on researchers painstakingly guiding the process to a desired state. In point of fact, how organisms themselves are able to develop toward species-invariant morphological outcomes is itself an open question of momentous importance in biology. However, there is hope that exploiting what we already know about multicellular development and the physical principles of self-organization—paying attention to appropriate "scaling up" of intelligent parts into wholes—can help make way on this in the synthetic domain.

Our suggestion, then, follows recent lines of research that emphasise the importance of constructing machines that themselves comprise smart, active, and, in some cases, cognitive parts. Sometimes this is phrased as a matter of "off-loading" computation from centralised computers to the body, though the language of embodied computation is ambiguous and difficult to specify technically [see Nakajima et al. (2015) and Müller and Hoffman (2017) for divergent stances on this]. What does seem crucial for this next stage of designed soft robots is the ability to achieve global coordination in a more autonomous and self-organized manner, something currently out of reach but hopefully not for too long, as that is itself an active area of research.

6.5 Concluding Thoughts

I opened this chapter with the suggestion that to create autonomous embodied robots capable of valuing and engaging with the world in a goal-directed manner requires incorporating several dimensions of biological endowment. In particular, I looked at the notion of precarity ("risk-to-self") and the related notion of existential needs. The proposal here has been that AI and AR come to autonomously value and interact with the world the more they approach biological analogs thereof–and the closer they approximate the dynamics that introduce the possibility of existential consequences for its actions. In other words, the proposal here has been to develop cognitive sophistication within alternative media by incorporating dimensions of vulnerability, precarity, and existential needs that emanate from the system's own internal dynamics with a denumerable set of actions that must be taken for this system to remain in far-from-equilibrium conditions.

To this end, I set out to dislodge several key assumptions embedded in the cognitive sciences that undermine the crucial role materiality plays in instantiating, mediating, and enabling cognitive form and function–specifically by proposing a fine-grained functionalist approach that treats the more minute properties of the system as central for cognitive function. The fields of active matter physics and soft robotics have begun to blur long-held dichotomies between hardware and software, living and non-living, machine and organism, and so on. But rather than reducing organisms to an anachronistic understanding of mechanism or matter, these fields have begun to actively question our understanding of materiality entirely. What we find, then, is not the hard, inert, and wholly passive parts standardly associated with machines and robots—but an inherent activity suffused throughout certain materials that, when brought into concerted, guided, and orchestrated engagement with one another *via* bioelectricity, can manifest and expand machine functionality in a manner unavailable to paradigms that do not avail themselves of these techniques. To construct autonomous robots, then, I propose an explicitly thermodynamic conception of life and mind that expands the domain of both terms. In effect, the view I have tried to articulate is one in which the mind is more material, and the material more mental, than is commonly believed—a view that is more at home in 18th and 19th century romanticist thought and American pragmatism (e.g., Charles Sanders Peirce)[107] than it is with 20th century reductive theories of matter. When we shift our perspective away from one in which higher-level cognition sits across a divide from inert, passive matter to a view in which materiality is already pregnant with the possibility of the mental, I believe we move one step closer to the goal of creating autonomous and hence actually intelligent machines.

[107] Something we hinted at in Chapter 2.

Chapter 7
Conclusion

7.1 "What is 'Nature'?"

I would like to conclude this work beginning with an anecdote. While writing this thesis I spent six months at the Konrad Lorenz Center [KLI] for Evolution and Cognition Research. An institute that has hosted hundreds of researchers, many of whom have made lasting contributions to the field[108], the KLI is an interdisciplinary institute that exemplifies not only the spirit of this work, but conveys a scientific attitude that I hoped to have portrayed herein and which I hope to carry with me into the future. On top of being a scientific space, however, we also had the pleasure of hosting an Artist Resident, Kendall Baker, whose work creatively reproduces material and natural processes through an artistic and creative medium—often incorporating elements found in forests, marshes, or mountains and intermixing them with manmade products to explore, problematise, and interrogate the supposedly watertight boundaries between a hostile 'nature' and a protected 'culture', where the goal of occidental societies has been to impose order and control on the former through the cultivation and means of the latter (as has been productively explored in the works of Descola (2013) and Scott (2020)). Baker, in his attempt to explore what it was he thought scientists were doing when they investigated or manipulated objects or natural phenomena to reveal their secrets (as a long running image of scientific practice vis-à-vis nature would have it), told me that in interviewing the residents and fellows contemporaneous with him at the institute, a consistent theme appeared and re-appeared—a theme best put in the form of a question: "What is 'Nature?'".

It was this question that appeared to me after a week away for a conference. The downstairs of the New Building of the KLI—which is segmented into two buildings, an old and a new, each with a downstairs and an upstairs—had been transformed with what at first appeared to be unrelated and heterogeneous elements: there were strips of wood; still wet detritus picked out of the nearby Danube; draft upon draft of sketches and paintings pinned to the wall; and paper mâché plastered over the exposed stone separating the Old and New buildings. What was once the highly modern space replete with rectilinear lines and sharp, geometric features

[108] I am particularly thankful for the many conversations that proved enlightening and entertaining from Eors Szathmary.

had turned into an experimental field; what was once clean and orderly had acquired a 'wilded' character. And among this confluence of the human and the wild there stood the normally idle and unused chalkboard (I have provided a photo below). This space, too, had also been 'wilded': normally containing the WIFI passwords and nothing else, it had been laced with gnarled pieces of wood and dried bark, with a few sketches papered here and there. And in the midst of all that heterogeneity there was this question, so simply stated, so casually expressed: "What is 'Nature'?" Given the nature of my research, I couldn't help but find the question intriguing. Was it rhetorical? Was it a question that could even be answered? Was it merely ('merely') evocative? A mixture of feelings and thoughts formed as I looked around the recently transformed space and I couldn't help but feel that such a simple, seemingly nonsensical question actually contained a germ of an idea—and I realised I had hoped my work could be a soil for it.

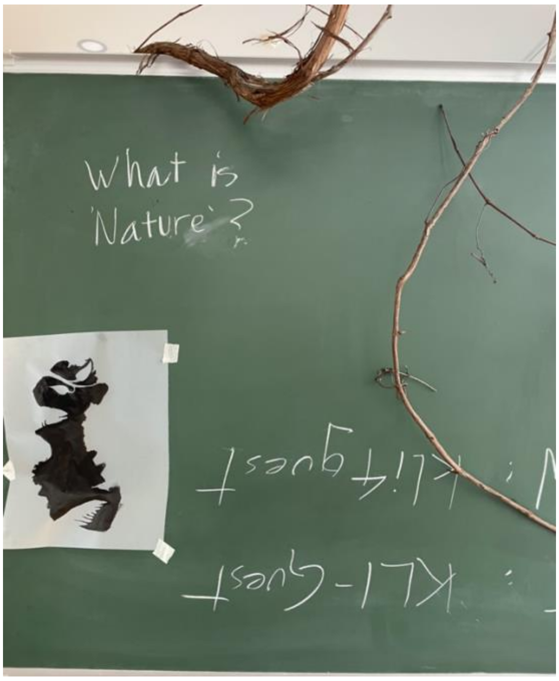

Figure 5. Kendall Baker's scribbling of the question "What is 'Nature?'" on the otherwise sparingly used chalk board at the Konrad Lorenz Institute in Klosterneuberg (Austria).

The question is likely to evoke in a reader a similar range of emotions and feelings that our fellows experienced. While generally diverse, the question was met with broadly dichotomous attitudes depending on whether a fellow was doing more scientific work or more theoretical, philosophical work. The scientists tended to react with some degree of consternation or scepticism; the philosophers were more disposed to romanticisation and appreciation. Plant biologist Sonia Sultan, also Baker's partner, who was a visiting fellow at this time responded saying that the question made no sense: 'Nature' is a provincialized concept dependent on which scientist is asking the question and in relation to their intended domain of inquiry.

There simply was no such thing as 'Nature' outside of that—or, at least, it did not exist in a way such that the question could be answered in a manner amenable to the codified nature of scientific practice. Evidently, my attitude is very different. I did not take the question as *needing* to be answered in a way amenable to a more structured, propositional format, but was meant to evoke deeper questions concerning the nature (as it were) of what is or is not possible (metaphysically) or what there is or is not (ontology).

In truth, the pertinence and relevance of the question is likely to blend both positions: scientists are certainly correct that asking after broad strokes questions of 'Nature' could obfuscate the disciplinarily sensitive conditions in which objects of empirical investigation emerge. It is in a sense similar to this that I believe that Sonia's rejoinder is correct, as what qualifies as natural is as much a product of our modes and methodologies of investigating as the other way around. Conversely, and as I have tried to outline in this work, the question "What is 'Nature'?" gestures towards a wider philosophical project that allows us to engage the sciences and the world disclosed along a metatheoretical access: because scientific practice, scientific seeing, is always situated we must appreciate that there is not an aperspectival starting point from which approaching agency, mind, and subjectivity can proceed.

While at the KLI, Kendall and I had the opportunity to discuss, sometimes discursively sometimes more directedly, about the relevance the question, haphazardly scrawled on the chalkboard, had for the *Naturphilosophen* and other Romantic Biologists—primarily due to the fact that addressing the tensions emerging at the origin of the biological sciences (the nature of recursive causality; the place of mind and agency in nature; the origination of organismic diversity; the relation between form and function (as epitomised in the famous debate between Etienne-Geoffrey Saint Hilaire & Georges Cuvier on the primacy of the one or the other; see Appel 1987), to name a few) required *new ways of seeing nature*, of understanding, analysing, and dissecting it, to truly establish biology as a science. I may have entered this debate through the modern context of the place of agency and mind in the life sciences (what active role do organisms play in the evolutionary process? What status does mental life have in relation to the reductive ideals of certain strands of neuroscience? How can we understand organisms as agents and more than a heuristic sense?); but more than anything else, it was encountering the excitement and optimism of the *Naturphilosophen*, their dogged commitment to the biological and chemical sciences, and the elaborate experimental frameworks that they inspired, that drove me to pursue the line of argumentation I did in this work—and this all circled back, over and over again, to the question: What is 'Nature'?

7.2 Conclusion

This work cannot possibly answer a question (indeed its unanswerability is wound up in the question) of this sort, but I hope to have shown why a consideration thereof is pertinent to many contemporary debates in the life and mind sciences concerning the ontological, explanatory, and scientific status of historically problematic concepts such as agency, teleology, and cognition. Research in Basal Cognition and the biogenic approach represents a novel, if troubled, line of research that situates paradigmatically cognitive processes or capacities in the living context of being embodied, living organism that acts to meet its existential needs. More than anything, this is meant to highlight and foreground the *material*, *physiological*, and *developmental* dimensions of being a cognitive agent—in a reversal of a common mode of thinking that presents cognition or intellection as a later occurrence and novel form of behavioural organization.

Basal Cognition is a promising line of research because it identifies minimal conditions for exhibiting a cognitive capacity of interest. However, the claims made therein can sometimes be difficult to assess, ascribing cognition to active matter systems and even to materiality itself is bound to generate tension within the cognitive sciences—and, if I am correct, represents something of a bad 'PR' move for the research program. Rather than identifying the cognitive nature of certain minimal systems, what I argue needs to be done here is the establishing of a middle ground in which the active and emergent nature of materiality and embodiment can be understood in relation to cognition without the former thereby being cognitive. This is why I have argued for an *agential* approach to some of these debates, as I believe the language of agency is more amenable to the discussion than insisting that the motile oil droplet is *itself* cognitive. It follows from this that agency represents a precondition for cognition—at least in the biological domain, and speculatively in the domain of Artificial General Intelligence, whether in the form of neural networks or embodied robotics.

In this work, I have argued that as important as research in Basal Cognition is developing a *philosophy of nature* that interrogates and antagonises sedimented presuppositions that render certain ways of thinking unintelligible. The key point of effectuating this project is to show that a pervasive teleophobia is *not* (or not simply) the product of a sobering, naturalistic approach to the world ('an austere and demanding view') but emanates from a *metaphysical* belief that itself can be reworked and overcome *in conversation with* how the physical, life, and mind sciences proceed.

To this effect, I have identified *intrinsic purposiveness* as a key component underpinning the forms of agency that I believe set biological (and potential exobiological or artificial) systems apart from other types of entities (rocks, tables, digital computers). We have seen that the original formulation of intrinsic purposiveness comes from Kant, but that as formulated he was unable to naturalise the concept in a way that aided the developing biological sciences of his time. Those who worked further in that direction are the *Naturphilosophen.* Recall from Chapter 2 that Goethe puzzled over Kant's Teleological Judgment and realised early on that the reason for this was that their views of nature were *incompatible*: where Kant saw the predominance and primacy of reductive mechanism, Goethe saw an elaborate world of order and organization that called for an emergentist and processual understanding of development and morphology. This historical episode highlights what I believe we are experiencing today: teleology is once again becoming a defining and central question for the life and mind sciences, but it still encounters at every turn sceptics who complain of untrammelled anthropomorphising and criticise researchers in Basal Cognition for flights of fancy. Resolving the issues between the two camps—and, what should be our bigger priority, furthering the sciences of self-organizing and goal-directed systems—is *as much* a matter of *philosophical* dispute as it is of scientific one. This is why I sought, in a Schellingian spirit, to return to the notion of matter to revive a current of thought wherein the same mysteries that seem inexorable from contemporary science were less 'spooky' (though no less difficult to answer!).

The return to the notion of matter and rethinking of materiality that defines this work is meant to show a few things. First, it vitiates a substance ontology whereby what are primary are things, entities, or isolated causal powers—instead foregrounding a relational and emergentist ontology where organization is pervasive and central. In Chapter 3, I tried to show how even from the perspective of physical sciences, the anti-emergentist positions that thrive in the philosophy of cognitive science do not hold. We must expand the physical within physicalism (and the natural within naturalism) in a way that is more consistent with the vibrant range of disciplines and subdisciplines that comprise the physical and biological sciences. Equipped with these conceptual resources, I believe we now stand at a time where we can better grasp what Dewey meant when he said that the matter of materialists and the matter of working scientists have nothing in common.

In Chapter 4, I wanted to further this line of argumentation to introduce an idea of material organization that is distinct in its modes of operations and the directions of possible development: that of *active matter* and the associated study in *active matter physics*. From the perspective of the biogenic approach, what matters to determining the agential status of an agent is the immanence of its activity within its material organisation: and not any kind of matter or set of relations can do so—an important, and I believe underappreciated, insight coming from this exciting area of research. Chapter 5 sought to draw the strings together to suggest a 'roadmap' for naturalising forms of agency—although the eight criteria we looked at each deserve much more attention than I was able to provide in this work. Nevertheless, I believe the eight criteria identified therein provide some hints as to what a robust agent might look like that goes beyond *as if* agency. This led me to lastly, in Chapter 6, explore the extent to which soft robotics and bio-inspired artificial intelligence could be a sandbox for exploring the forms of naturalised agency promulgated in this work.

Many questions, of course remain—and I am under no illusion that a mystery has been expelled from the heart of modern science. Again, my main goal was to investigate how we came to see agency as unnaturalizable; and propose a direction or avenue of research wherein we can ask legitimate questions concerning the status, nature, and causal capacities of agency—rather than bickering about whether it is a 'real' part of nature. It is. More interesting than remaining stalwartly committed to a view of nature that dispels agency from our ontologies, we should instead work to accept the evidence that organisms are agents—and ourselves, as said organisms, are as well—and then work on the remaining, very interesting questions that investigate the mental-physical interaction. Although there is still much left to do, I hope to have signalled a way in which the gap between the two is less mysterious than has commonly been supposed.

Bibliography

Anderson, P. (1972). More Is Different. *Science* **177**, 393-396 (1972). DOI: 10.1126/science.177.4047.393

Arias Del Angel, J. A., Nanjundiah, V., Benítez, M., and Newman, S. A. (2020). Interplay of mesoscale physics and agent-like behaviors in the parallel evolution of aggregative multicellularity. EvoDevo 11, 1–18. doi: 10.1186/s13227-020-00165-8

Baetu, T.M., 2015, "When is a mechanistic explanation satisfactory? Reductionism and antireductionism in the context of mechanistic explanations", in I. Pârvu, G. Sandu, and I. Toader (eds.), *Romanian studies in philosophy of science*, Dordrecht: Springer, 255–268.

Baez, J. C. (2021). The Joy of Condensed Matter. Nautilus Magazine. Available online at: https://nautil.us/the-joy-of-condensed-matter-9608/ (accessed August11, 2022).

Baluška, F., and Levin, M. (2016). On having no head: Cognition throughout biological systems. Front. Psychol. 7, 902. doi: 10.3389/fpsyg.2016.00902

Batterman, R. W. (2012). The tyranny of scales. In R. W. Batterman (Ed.), Oxford handbook of philosophy of physics (pp. 255–286). Oxford: Oxford University Press.

Batterman, R. (2021). "Multiscale modelling in inactive and active materials," in Levels of Biological Organization in the Biological Sciences. eds, D. Brooks, J. DiFrisco, and W. C. Wimsatt (Cambridge, Massachusetts: MIT Press).

Barad, K. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Chapel Hill, NC: Duke University Press.

Bechinger, C., Di Leonardo, R., Löwen, H., Reichhardt, C., Volpe, G., & Volpe, G. (2016). Active particles in complex and crowded environments. *Reviews of Modern Physics*, *88*(4), 045006.

Bechtel, W. (2014). Cognitive Biology: Surprising Model Organisms for Cognitive Science. In Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. 158– 163).

Bechtel, W., and Bich, L. (2021). Grounding cognition: heterarchical control mechanisms in biology. Philos. Transac. R. Soc. B. 376, 20190751. doi: 10.1098/rstb.2019.0751

Beekman, M., & Latty, T. (2015). Brainless but Multi-Headed: Decision Making by the Acellular Slime Mould Physarum polycephalum. *Journal of molecular biology*, *427*(23), 3734–3743. https://doi.org/10.1016/j.jmb.2015.07.007

Bennett, J. 2009. *Vibrant Matter: Towards a Political Ecology of Things*. Durham, NC: Duke University Press.

Bickhard, M. (1993). Representational content in humans and machines. J. Exper. Theor. Artif. Intell. 5, 285–233. doi: 10.1080/09528139308953775

Blackiston, D., Lederer, E., Kriegman, S., Garnier, S., Bongard, J., and Levin, M. (2021). A cellular platform for the development of synthetic living machines. Sci. Robot. 6, eabf1571. doi: 10.1126/scirobotics.abf1571

Block, N. (1978). Troubles with functionalism. Minnesota Stud. Philos. Sci. 9, 261–325.

Bongard, J., and Levin, M. (2021). Living things are not (20th century) machines: Updating mechanism metaphors in light of the modern science of machine behavior. Front. Ecol. Evol. 9, 650726. doi: 10.3389/fevo.2021.650726

Boussard, A., Fessel, A., Oettmeier, C., Briard, L., Döbereiner, H. G., and Dussutour, A. (2021). Adaptive behaviour and learning in slime moulds: the role of oscillations. Philos. Transac. R. Soc. B. 376, 20190757. doi: 10.1098/rstb.2019.0757

Brancazio, N. (2023). Interactive agential dynamics. *Synthese* 201 (6):1-20.

Brugues, J., and Needleman, D. (2014). Physical basis of spindel self-organization. Proc. Nat. Acad. Sci. 111, 18496–18500. doi: 10.1073/pnas.1409404111

Brunet, T. D. P., and Halina, M. (2020). Minds, machines, and molecules. Philos. Topic. 48, 221–241. doi: 10.5840/philtopics202048111

Butler, J. 1993. *Bodies that Matter: On the Discursive Limits of Sex*. London: Routledge.

Campbell, R J. & Bickhard, M. H. (2011). Physicalism, Emergence and Downward Causation. *Axiomathes* 21 (1):33-56.

Carabotti M, Scirocco A, Maselli MA, Severi C. 2015. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. Ann Gastroenterol. Vol. 28(2):203-209. PMID: 25830558; PMCID: PMC4367209.

Cates, E. and M. R. Evans. (Bristol; Philadelphia: Edinburgh, Scotland: Institute of Physics Pub.; copublished by Scottish Universities Summer School in Physics). doi: 10.1201/9781420033519.ch1

Cejkova, J., Novak, M., Stepanek, F., and Hanczyc, M. (2014). Dynamics of chemotactic droplets in salt concentration gradients. Langmuir 30, 11937–11944. doi: 10.1021/la502624f

Chalmers, D. 1995. Absent Qualia, Fading Qualia, Dancing Qualia. In Metzinger, T. *Conscious Experience*. Imprint Academic.

Checa, S., Rausch, M. K., Petersen, A., Kuhl, E., & Duda, G. N. (2015). The emergence of extracellular matrix mechanics and cell traction forces as important regulators of cellular self-organization. *Biomechanics and modeling in mechanobiology*, *14*(1), 1–13. https://doi.org/10.1007/s10237-014-0581-9

Chirimuuta, M. (2018). Marr, Mayr, and MR: What functionalism should now be about. Philos. Psychol. 31, 403–418. doi: 10.1080/09515089.2017.1381679

Ciaunica, A. & Levin, M. (2022). The Brain is not Mental! Coupling Neuronal and Immune Cellular Procesing in Human Organisms. *PsyArXiv Preptints*. Doi: 10.31234/osf.io/fgcy5

Ciaunica, A., Safron, A., and Delafield-Butt, J. (2021). Back to square one: the bodily roots of conscious experiences in early life. Neurosci. Conscious. 2021, niab037. doi: 10.31234/osf.io/zspm2

Cichos, F., Gustavsson, K., Mehlig, B., and Volpe, G. (2020). Machine learning for active matter. Nat. Mach. Intell. 2, 94–103. doi: 10.1038/s42256-020-0146-9

Coole, D. & Frost, S. (2010). *New Materialisms: Ontology, Agency, Politics.* Duke: Duke University Press.

Cusimano, S., & Sterner, B. (2020). The Objectivity of Organizational Functions. *Acta biotheoretica*, *68*(2), 253–269. https://doi.org/10.1007/s10441-019-09365-9

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

Craver, C. (2013). Functions and Mechanisms: A Perspectivalist Account. In P. Huneman (ed.), *Functions: Selection and Mechanisms*, Dordrecht: Springer, pp. 133–158.

Crick, F. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul.* Scribner Reprint Edition.

Cryan JF, Dinan TG. 2012. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. Nat Rev Neurosci. Oct;13(10):701-12. doi: 10.1038/nrn3346. Epub 2012 Sep 12. PMID: 22968153.

Dalous, J., Burghardt, E., Müller-Taubenberger, A., Bruckert, F., Gerisch, G., and Bretschneider, T. (2008). Reversal of cell polarity and actin-myosin cytoskeleton reorganization under mechanical and chemical stimulation. Biophys. J. 94, 1063–1074. doi: 10.1529/biophysj.107.114702

Darwin, C. (1859/2009). *On the Origin of Species.* Oxford: Oxford World's Classics.

Davies, P. (2008). The Physics of Downward Causation. In Philip Clayton, and Paul Davies (eds), *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion.* Oxford: Oxford University Press.

Deacon, T. (2012). *Incomplete Nature: How Mind Emerged from Matter.* New York, New York: W.W. Norton & Co.

Dennett, D. (1996). *Darwin's Dangerous Idea: Evolution and the Meaning of Life.* London: Penguin Press.

Di Paolo, E., Buhrmann, T., and Barandiaran, X. E. (2017). Sensorimotor Life: An Enactive Proposal. Oxford: Oxford University Press doi: 10.1093/acprof:oso/9780198786849.001.0001

Dinet, C., Michelot, A., Herrou, J., and Mignot, T. (2021). Linking single-cell decisions to collective behaviours in social bacteria. Philos. Transac. R. Soc. B 376, 20190755. doi: 10.1098/rstb.2019.0755

Drossel, B. (2021). Strong Emergence in Condensed Matter Physics. In Jan Voosholz & Markus Gabriel (eds.), Top-Down Causation and Emergence. Springer Verlag.

Dupre, J. (2014). *Processes of Life: Essays in the Philosophy of Biology*. Oxford, Oxford University Press.

Ebrahimkhani, M. R., and Levin, M. (2021). Synthetic living machines: a new window on life. IScience 24, 102505. doi: 10.1016/j.isci.2021.102505

Egbert, M. (2022). Towards Adaptive Sensorimotor Autonomy: Developing A System That can Adapt to Its Own Emergent and Dynamic Needs. doi: 10.1162/isal_a_00503

Egbert, M. D. (2021). Self-preserving Mechanisms in Motile Oil Droplets: A computational model of abiological self-preservation. R. Soc. Open Sci. 8, 210534. doi: 10.1098/rsos.210534

Ellis G. F. (2012). Top-down causation and emergence: some comments on mechanisms. *Interface focus*, *2*(1), 126–140. https://doi.org/10.1098/rsfs.2011.0062

Emmanuel, P., Yeomans, J., and Doostmohammadi, A. (2021) Activity pulses induce spontaneous flow reversals in viscoelastic environments. *Journal of the Royal Society Interface.* Vol. 18: 20210100.20210100

Emmeche, C., Køppe, S. and Stjernfelt, F. (2000). Levels, Emergence, and Three Versions of Downward Causation. In Andersen, P, Emmeche, C., Finnemann N., and Christiansen, P. eds. *Downward Causation. Minds, Bodies and Matter.* Århus: Aarhus University Press.

Fields RD, Araque A, Johansen-Berg H, Lim SS, Lynch G, Nave KA, Nedergaard M, Perez R, Sejnowski T, Wake H. 2014. Glial biology in learning and cognition. Neuroscientist. Vol.20(5): 426-31. doi: 10.1177/1073858413504465.

Fields, C., Bischof, J., and Levin, M. (2020). Morphological coordination: a common ancestral function unifying neural and nonneural signalling. Physiology 35, 16–30. doi: 10.1152/physiol.00027.2019

Fields, C., Glazebrook, J. F., and Levin, M. (2021). Minimal physicalism as a scale free substrate for cognition and consciousness. Neurosci. Consc. 2021, niab013. doi: 10.1093/nc/niab013

Fields, C., and Levin, M. (2020). Scale-free biology: integrating evolutionary and developmental thinking. BioEssays 42, 1900228. doi: 10.1002/bies.201900228

Figdor, C. (2017). On the Proper Domain of Psychological Predicates. Synthese 194 (11):4289-4310.

Fodor, E., and Marchetti, C. (2018). The statistical physics of active matter: From self-catalytic colloids to living cells. Phys. A Stat. Mech. Appl. 504, 106–120. doi: 10.1016/j.physa.2017.12.137

Friston, K. J. (2019). A free energy principle for a particular physics. ArXiv E-Prints, arXiv:1906.10184.

Froese, T. (2016). “Life is precious because it is precarious: individuality, mortality, and the problem of meaning,” in Representation of Reality: Humans, Other Living Organisms and Intelligent Machines, eds. D. Dodig-Crnkovic, and R. Giovagnoli (Springer: Berlin). doi: 10.1007/978-3-319-43784-2_3

Gagliano M. (2017). The mind of plants: Thinking the unthinkable. *Communicative & Integrative Biology*, *10*(2), e1288333. https://doi.org/10.1080/19420889.2017.1288333

Gallagher, S. (2018) Rethinking Nature: Phenomenology and a Non-reductionist Cognitive Science, Australasian Philosophical Review, 2:2, 125-137.

Gambarotto, A. 2017. *Vital Forces, Teleology, and Organisation: Philosophy of Nature and the Rise of Biology in Germany*. Springer, Berlin: Springer Nature Group.

Gershman, S. J., Balbi, P. E., Gallistel, C. R., and Gunawardena, J. (2021). Reconsidering the evidence for learning in single cells. ELife. 10, e61907. doi: 10.7554/eLife.61907

Gillett, C. (2007). Understanding The New Reductionism. *Journal of Philosophy* 104 (4):193-216.

Ginsburg, S. & Jablonka, E. 2019. *Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*

Ginsburg, S., & Jablonka, E. (2021). Evolutionary transitions in learning and cognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *376*(1821), 20190766. https://doi.org/10.1098/rstb.2019.0766

Godfrey-Smith, P. (1996). Complexity and the Function of Mind in Nature. Cambridge, UK: Cambridge University Press.

Godfrey-Smith, P. (2016a). Mind, matter, and metabolism. J. Philos. 113, 481–506. doi: 10.5840/jphil20161131034

Godfrey-Smith, P. (2016b). Individuality, subjectivity, and minimal cognition. Biol. Philos. 3, 275–286. doi: 10.1007/s10539-016-9543-1

Godfrey-Smith, P. (2017). The subject as cause and effect of evolution. Interface Focus 7, 22. doi: 10.1098/rsfs.2017.0022

Goertzel, B., and Pennachin, C. (2007). Artificial General Intelligence. Cognitive Technologies. Berlin: Springer. doi: 10.1007/978-3-540-68677-4

Gold, I., & Stoljar, D. (1999). A neuron doctrine in the philosophy of neuroscience. *Behavioral and Brain Sciences*, *22*(5), 809–869. https://doi.org/10.1017/S0140525X99002198

Gould, S. (2002). *The Structure of Evolutionary Theory*. Boston MA: Harvard University Press.

Green, S. (2022). Cancer Beyond Genetics: On the Practical Implications of Downward Causation. In Brooks, D.S., DiFrisco, J., and Wimsatt, W.C. *Levels of Organization in the Biological Sciences*. Boston, MA: MIT Press.

Green, S., & Batterman, R. (2017). Biology meets physics: Reductionism and multi-scale modeling of morphogenesis. *Studies in history and philosophy of biological and biomedical sciences*, *61*, 20–34. https://doi.org/10.1016/j.shpsc.2016.12.003

Halley, J. D., & Winkler, D. A. (2008). Consistent concepts of self-organization and self-assembly. *Complexity*, *14*(2), 10-17. https://doi.org/10.1002/cplx.20235

Hanczyc, M. (2011). Metabolism and motility in prebiotic structures. Philos. Trans. R. Soc. B 366, 2885–2893. doi: 10.1098/rstb.2011.0141

Hanczyc, M., and Ikegami, T. (2010). Chemical basis for minimal cognition. Artif. Life 16, 233–243. doi: 10.1162/artl_a_00002

Hanson, A. (2021). Spontaneous electrical low-frequency oscillations: a possible role in hydra and all living systems. Philos. Transac. R. Soc. B. 376, 20190763. doi: 10.1098/rstb.2019.0763

Haugeland, J. (1998). Understanding Natural Language. Harvard University Press.

Hazelbauer, G. L., Falke, J. J., & Parkinson, J. S. (2008). Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends in biochemical sciences*, *33*(1), 9–19. https://doi.org/10.1016/j.tibs.2007.09.014

Heck D. H., McAfee S. S., Liu Y., Babajani-Feremi A., Rezaie R., Freeman W. J., et al. (2017). Breathing as a fundamental rhythm of brain function. *Front. Neural Circ.* 10:115. 10.3389/fncir.2016.00115

Hodges, B., and Baron, R. (1992). Values as constraints on affordances: perceiving and acting properly. J. Theory Soc. Behav. 22, 263–294. doi: 10.1111/j.1468-5914.1992.tb00220.x

Hodges, B., and Raczaszek-Leonardi, J. (2021). Ecological values theory: beyond conformity, goal-seeking, and rule-following in action and interaction. *Rev. General Psychol.* 26, 86–103. doi: 10.1177/10892680211048174

Hoyningen-Huene, P. (1994). Niels Bohr's argument for the irreducibility of biology to physics. In Jan Faye & Henry J. Folse (eds.), N*iels Bohr and Contemporary Philosophy*. Kluwer Academic Publishers. pp. 231--255.

Hull, D. L. (1981) Metaphysics and common usage. *Behav. Brain Sci.* 4:290-291

Jablonka, E. & Lamb, M. (2011). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioural, and Symbolic Variation in the History of Life*. Boston, MA: MIT Press.

Jekely, G. (2021). The chemical brain hypothesis for the origin of nervous systems. Philos. Transac. R. Soc. B 376, 761. doi: 10.1098/rstb.2019.0761

Jékely, G., Keijzer, F., and Godfrey-Smith, P. (2015). An option space for early neural evolution. Philos. Transac. R. Soc. B. 370, 20150181.doi: 10.1098/rstb.2015.0181

Jonas, H. 1966. *The Phenomenon of Life: Towards a Philosophical Biology.*. Minneapolis, MN: Northwestern University Press.

Jones, J. (2015). Applications of multi-agent slime mould computing. Int. J.Parallel, Emerg. Distr. Syst. 31, 420–449.

Juarrero, A. (2015). What does the closure of context-sensitive constraints mean for determinism, autonomy, self-determination, and agency? Progress in Biophys.Molec. Biol. 119, 510–521. doi: 10.1016/j.pbiomolbio.2015.08.007

Jülicher, F., Grill, S. W., and Salbreux, G. (2018). Hydrodynamic theory of active matter. Reports Progr. Phys. 81, 076601. doi: 10.1088/1361-6633/aab6bb

Kanao, K., Harada, S., Yamamoto, Y., Honda, W., Aerie, T., Akita, S., et al. (2015).Highly selective flexible tactile strain and temperature sensors against substrate bending for an artificial skin. RSC Adv. 5, 30170. doi: 10.1039/C5RA03110A

Kaspar, C., Ravoo, B. J., van der Wiel, W. G., Wegner, S. V., and Pernice, W. H. P. (2021). The rise of intelligent matter. Nature 594, 345–355.doi: 10.1038/s41586-021-03453-y

Kauffman, S. A. (1993). The Origins of Order: Self-Organization and Selection in Evolution. New York: Oxford University Press. doi: 10.1007/978-94-015-8054-0_8

Keijzer, F., Van Duijn, M., and Lyon, P. (2013). What nervous systems do: early evolution, input–output, and the skin brain thesis. Adapt. Behav. 21, 67–85.doi: 10.1177/1059712312465330

Kelso, S. (2016). On the self-organising origins of agency. Trends Cogn. Sci. 20,490–499. doi: 10.1016/j.tics.2016.04.004

Kempf, F., Mueller, R., Frey, E., Yeomans, J. M., and Doostmohammadi,A. (2019). Active matter invasion. Soft Matter 15, 7538–7546.doi: 10.1039/C9SM01210A

Kim J. (1993) Supervenience and mind. Cambridge University Press.

Kim J. (1998) Philosophy of mind. Westview Press.

Kim J. (1999) Making sense of emergence. *Philos Stud* 95:3–36

Kriegman, S., Blackiston, D., Levin, M., and Bongard, J. (2020). A scalable pipeline for designing reconfigurable organisms. Proc. Nat. Acad. Sci. 117,1853–1859. doi: 10.1073/pnas.1910837117

Kriegman, S., Blackiston, D., Levin, M., and Bongard, J. (2021). Kinematic self-replication in reconfigurable organisms. Proc. Nat. Acad. Sci. 118, e2112672118.doi: 10.1073/pnas.2112672118

De Landa, M. 1997. *A Thousand Years of Nonlinear History.* Princeton, NJ: Princeton University Press.

Levin, M. (2019). The computational boundary of a 'Self ': Developmental bioelectricity drives multicellularity and scale-free cognition. Front. Psychol. 10,2688. doi: 10.3389/fpsyg.2019.02688

Levin, M. (2020). The biophysics of regenerative repair suggests new perspectives on biological causation. BioEssays. 42, 1900146. doi: 10.1002/bies.201900146

Levin, M. (2021). Life, death, and self: Fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. Biochem. Biophys. Res. Commun. 564, 114–133. doi: 10.1016/j.bbrc.2020.10.077

Levin, M., and Dennett, D. C. (2020). Cognition All The Way Down. Aeon. Available online at: https://aeon.co/essays/how-to-understand-cells-tissues-and-organisms-as-agents-with-agendas (accessed November 8, 2022).

Lewens, T. 2004. *Organisms and Artifacts: Design in Nature and Elsewhere.* Cambridge, MA: MIT Press.

Lewes, G. H. (1875). *Problems of Life and Mind. First Series: The Foundations of a Creed.* Vol. 2. Boston: Osgood.

Lewis, D. (1986). *Philosophical Papers, volume II.* New York: Oxford University Press.

Loewer, B. (2009). Why Is There Anything Except Physics? *Synthese*, *170*(2), 217–233. http://www.jstor.org/stable/40271331

Luppi, A. I., Mediano, P. A. M., Rosas, F. E., Harrison, D. J., Carhart-Harris, R. L., Bor, D., & Stamatakis, E. A. (2021). What it is like to be a bit: an integrated information decomposition account of emergent mental phenomena. *Neuroscience of consciousness*, *2021*(2), niab027. https://doi.org/10.1093/nc/niab027

Lyon, P. (2006). The biogenic approach to cognition. Cogn. Process. 7, 11–29.doi: 10.1007/s10339-005-0016-8

Lyon, P. (2015). The cognitive cell: Bacterial behavior reconsidered. Front. Microbiol. 6, 264. doi: 10.3389/fmicb.2015.00264

Lyon, P. (2020). Of what is 'minimal cognition' the half-baked version? *Adaptive Behaviour*, vol. 28: 6. Doi: 10.1177/105971231987136.

Lyon, P., Keijzer, F., Arendt, D., and Levin, M. (2021). Reframing cognition: getting down to biological basics. Philos. Transac. R. Soc. B 376, 750.doi: 10.1098/rstb.2019.0750

Lyon, P., and Kuchling, F. (2021). Valuing what happens: A biogenic approach to valence and (potentially) affect. Philos. Trans. Royal Soc. B. Biol. Sci. 376, 20190752.doi: 10.1098/rstb.2019.0752

Machamer, P.K., Darden, L and Craver, C. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67:1–25.

Man, K., and Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. Nat. Mach. Intell. 1, 446–452. doi: 10.1038/s42256-019-0103-7

Marchetti, M. C., Joanny, J. F., Ramawamy, S., Liverpool, T. B., Prost, J., Madan,R., et al. (2013). Hydrodynamics of sft active matter. Rev. Modern Phys. 85,1143–1189. doi: 10.1103/RevModPhys.85.1143

Maturana, H. R., and Varela, F. J. (1980). Autopoiesis and Cognition: The Realization of the Living, Vol. 42. Springer Netherlands. doi: 10.1007/978-94-009-8947-4

Mayer EA. 2012. Gut feelings: the emerging biology of gut-brain communication. *Nat Rev Neurosci.* Vol.12:453–466. doi: 10.1038/nrn3071.

Mayer EA, Knight R, Mazmanian SK, Cryan JF, Tillisch K. 2014. Gut microbes and the brain: paradigm shift in neuroscience. J Neurosci. Vol. 12;34(46):15490-6. doi: 10.1523/JNEUROSCI.3299-14.2014. PMID: 25392516; PMCID: PMC4228144.

McCarthy, J., and Hayes, P. J. (1969). "Some philosophical problems from the standpoint of artificial intelligence," in Machine Intelligence, Vol. 4, eds B. Meltzerand D. Michie (Edinburgh: Edinburgh University Press), 463–502.

McGivern, P. (2020). Active materials: minimal models of cognition? Adapt. Behav. 28, 441–451. doi: 10.1177/1059712319891742

Meyer, R., and Brancazio, N. (2021). Putting down the revolt: Enactivism as a philosophy of nature. Front. Psychol. 13, 948733. doi: 10.3389/fpsyg.2022.948733

Miller, C. J., & Davidson, L. A. (2013). The interplay between cell signalling and mechanics in developmental processes. *Nature reviews. Genetics*, *14*(10), 733–744.

Mogilner, A., and Manhart, A. (2018). Intracellular fluid mechanics: Coupling cytoplasmic flow with active cytoskeletal gel. Ann. Rev. Fluid Mech. 50, 347–370. doi: 10.1146/annurev-fluid-010816-060238

Moreno, A., and Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial life*, *11*(1-2), 161–175. https://doi.org/10.1162/1064546053278919

Moreno, A., and Mossio, M. (2015). Biological Autonomy: A Philosophical and Theoretical Enquiry. London: Springer.

Mossad O., Nent E., Woltemate S., Folschweiller S., Buescher J. M., Schnepf D., et al.. (2021). Microbiota-dependent increase in δ-valerobetaine alters neuronal function and is responsible for age-related cognitive decline. *Nat. Aging* 1, 1127–1136. 10.1038/s43587-021-00141-4

Mossio, M., Bich, L. (2017). What makes biological organisation teleological?. *Synthese* **194**, 1089–1114. https://doi.org/10.1007/s11229-014-0594-z

Müller, V., and Hoffman, M. (2017). What is morphological computation? On how the body contributes to cognition and control. Artif. Life 23, 1–24. doi: 10.1162/ARTL_a_00219

Nakajima, K., Hauser, H., Li, T., and Pfeifer, R. (2015). Information processing via physical soft body. Sci. Rep. 5, 1–11. doi: 10.1038/srep10487

Needleman, D., and Dogic, Z. (2017). Active matter at the interface between materials science and cell biology. Nat. Rev. Mater. 2, 1–14. doi: 10.1038/natrevmats.2017.48

Newman, S A. and Linde-Medina, M. (2013). Physical Determinants in the Emergence and Inheritance of Multicellular Form. *Biological Theory* 8 (3):274-285.

Newman, S. A. (2016). "Biogeneric' developmental processes: drivers of major transitions in animal evolution. Philos. Transac. R. Soc. B. 371, 20150443. doi: 10.1098/rstb.2015.0443

Newman, S. A. (2019). Inherency of form and function in animal development and evolution. Front. Physiol. 10, 702. doi: 10.3389/fphys.2019.00702

Newman, S. A. (2020). "The origins and evolution of animal identity," in Biological Identity: Perspectives from Metaphysics and the Philosophy of Biology, eds. A. S. Meincke and J. Dupre (London: Routledge). doi: 10.4324/978135106 6389-7

Newman, S. (2021). "Inherency," in Evolutionary Developmental Biology: A Reference Guide, Volume 1, eds L. N. de la Rosa and G. Muller (Cham: Springer), 121–132. doi: 10.1007/978-3-319-32979-6_78

Newman, S. A. (2022). Inherency and agency in the origin and evolution of biological function. Biol. J. Linnean Soc. 20, 1–16. doi: 10.1093/biolinnean/blac109

Nicholson, D. (2014). The machine conception of the organism in development and evolution: A critical analysis. Stud. History Philos. Biol. Biomed. Sci. 48, 162–174. doi: 10.1016/j.shpsc.2014.08.003

Nicholson D. J. (2019). Is the cell really a machine?. *Journal of theoretical biology*, *477*, 108–126. https://doi.org/10.1016/j.jtbi.2019.06.002

Noble D. (2012). A theory of biological relativity: no privileged level of causation. *Interface focus*, *2*(1), 55–64. https://doi.org/10.1098/rsfs.2011.0067

Nunes-Alves C. (2015). Bacterial physiology. Linking cell differentiation and growth. *Nature reviews. Microbiology*, *13*(7), 398–399. https://doi.org/10.1038/nrmicro3505

Oden, J. T. (2006). Simulation based engineering science: An NSF Blue Ribbon Report. www.nsf.gov/ pubs/reports/sbes_fnal_report.pdf.

Oppenheim, P. and Putnam, H. (1958). Unity of Science as a Working Hypothesis. *Minnesota Studies in the Philosophy of Science* vol.2: 3–36.

Pezzulo, G., Rigoli, F., and Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. Progr. Neurobiol. 134, 17–35. doi: 10.1016/j.pneurobio.2015.09.001

Pfeifer, R., Lungarella, M., and Iida, F. (2007). Self-organization, embodiment, and biologically inspired robots. Science 318, 1088–1093. doi: 10.1126/science.114580

Pfeifer, R., and Scheier, C. (1999). Understanding Intelligence. Cambridge, MA: MIT Press.

Piccinini, G. 2010. The Resilience of Computationalism. *Philosophy of Science*, vol.77(5): 852-861.

Pietak, A., & Levin, M. (2018). Bioelectrical control of positional information in development and regeneration: A review of conceptual and computational advances. *Progress in biophysics and molecular biology*, *137*, 52–68. https://doi.org/10.1016/j.pbiomolbio.2018.03.008

Pishvar, M., and Harne, R. L. (2020). Foundations for soft, smart matter by active mechanical metamaterials. Adv. Mater. 7, 2001384. doi: 10.1002/advs.202001384

Polger, T. W., and Shapiro, L. A. (2016). The Multiple Realization Book, 1st Edn. Oxford University Press.

Poon, W. C. K. (2000). "A day in the life of a hard-sphere suspension," in Soft and Fragile Matter: Nonequilibrium Dynamics, Metastability, and Flow. eds, M.

Popov, K., Komianos, J., & Papoian, G. 2016. Mechanochemical Simulations of Contractions and Polarity Alignment in Actomyosin Networks. PLOS Computational Biology, 12(4): e1004877. https://doi.org/10.1371/journal.pcbi.1004877

Popkin, G. (2016). The physics of life. Nat. News 529, 16–18. doi: 10.1038/529016a

Potochnik, A. (2017). Idealization and the Aims of Science. Chicago: The University of Chicago Press. doi: 10.7208/chicago/9780226507194.001.0001

Potochnik, A. (2021). Our World Isn't Organised Into Levels. In Daniel Brooks, James DiFrisco, and William Wimsatt Levels of Organization in the Biological Sciences. Cambridge, Mass: MIT Press.

Potter, H. D., & Mitchell, K. J. (2022). Naturalising Agent Causation. *Entropy (Basel, Switzerland)*, *24*(4), 472. https://doi.org/10.3390/e24040472

Prindle, A., Liu, J., Asally, M., Ly, S., Garcia-Ojalvo, J., and Süel, G. M. (2015). Ion channels enable electrical communication in bacterial communities. Nature 527, 59–63. doi: 10.1038/nature15709

Putnam, H. (1975). Mind, Language, and Reality: Philosophical Papers. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511625251

Putnam, H. (1988). Representation and Reality. Cambridge, MA: MIT Press.

Ramaswamy, S. (2017). Active Matter. *Journal of Statistical Mechanics: Theory and Experiment*, 054002.

Reynolds, A. (2007). The theory of the cell state and the question of cell autonomy in nineteenth and early twentieth-century biology. Sci. Context 20, 71–95. doi: 10.1017/S0269889706001141

Richards, R. 1992. *The Meaning of Evolution: Morphological Construction and Ideological Reconstruction of Darwin's Theory.* Chicago, Illinois: Chicago University Press.

Richards 2002. *The Romantic Conception of Life: Science and Philosophy in the Age of Goethe.* Chicago, Illinois: Chicago University Press.

Riskin, J. 2016. *The Restless Clock: A History of the Centuries Long Argument over What Makes Living Things Think.* Chicago, IL: Chicago University Press.

Rorot, W. (2022). Counting with cilia: The role of morphological computation in basal cognition research. Entropy, 24, 1581. doi: 10.3390/e24111581

Rosen, R. (1991). *Life Itself—A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life.* New York, New York: Columbia University Press.

Rosenberg, A & Kaplan, D. (2005). How to reconcile physicalism and antireductionism about biology. *Philosophy of Science* 72 (1):43-68.

Rosenberg, Alex. (2020). Scientism Versus the Theory of Mind. *Social Epistemology Review and Reply Collective* 9 (1): 48-57. https://wp.me/p1Bfg0-4Mf.

Russell, E.S. (1945). *The Directiveness of Organic Activities.* Cambridge: Cambridge University Press.

Sachs, Carl & Nahas, Auguste (2023). What's at stake in the debate over naturalizing teleology? An overlooked metatheoretical debate. *Synthese* 201 (4):1-22.

Sanchez, T., Chen, D. T., DeCamp, S. J., Heymann, M., and Dogic, Z. (2012). Spontaneous motion in hierarchically assembled active matter. Nature 491, 431–434. doi: 10.1038/nature11591

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., et al. (2020). Mastering atari, go, chess and shogi by planning with a learned model. Nature 588, 604–609. doi: 10.1038/s41586-020-03051-4

Seibt, J. 2018. Ontological Tools for the Process Turn in Biology: Some Basic Notions of *General Process Theory*. in J. Dupré and D. Nicholson (eds.), *Everything Flows: Towards a Processual Philosophy of Biology*, Oxford: Oxford University Press.

Senoussi, A., Galas, J. C., & Estevez-Torres, A. (2021). Programmed mechano-chemical coupling in reaction-diffusion active matter. *Science advances*, *7*(51), eabi9865. https://doi.org/10.1126/sciadv.abi9865

Seth, A. K., & Tsakiris, M. (2018). Being a Beast Machine: The Somatic Basis of Selfhood. *Trends in cognitive sciences*, *22*(11), 969–981. https://doi.org/10.1016/j.tics.2018.08.008

Shah, D. S., Powers, J. P., Tilton, L. G., Kriegman, S., Bongard, J., and KramerBottiglio, R. (2021). A soft robot that adapts to environments through shape change. Nat. Mach. Intell. 3, 51–59. doi: 10.1038/s42256-020-00263-1

Sims, M. (2020). *Strong Continuity of life and mind: The free energy framework, predictive processing and ecological psychology.* Unpublished Ph.D. Dissertation, University of Edinburgh. Doi: http://dx.doi.org/10.7488/era/1370

Sims, M. (2020). How to count biological minds: symbiosis, the free energy principle, and reciprocal multiscale integration. Synthese 199, 2157–2179. doi: 10.1007/s11229-020-02876-w

Sims, M. (2021). A continuum of intentionality: linking the biogenic and anthropogenic approaches to cognition. Biol. Philos. 38, 1–31. doi: 10.1007/s10539-021-09827-w

Smith, H. and Morowitz, H. (2016). *The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere.* Cambridge: Cambridge University Press.

Smith-Ferguson, J., and Beekman, M. (2020). Who needs a brain? Slime moulds, behavioural ecology and minimal cognition. Adapt. Behav. 28, 465–478. doi: 10.1177/1059712319826537

Soto, A. M., Sonnenschein, C., & Miquel, P. A. (2008). On physicalism and downward causation in developmental and cancer biology. *Acta biotheoretica*, *56*(4), 257–274. https://doi.org/10.1007/s10441-008-9052-y

Sonnenschein, C., and Soto, A. M. (2016). Carcinogenesis explained within the context of a theory of organisms. *Progress in biophysics and molecular biology*, *122*(1), 70–76.

Stotz, K & Allen, C (2012). From cell-surface receptors to higher learning: A whole world of experience. In Philosophy of Behavioral Biology, eds, Katie Plaisance and Thomas Reydon. Boston: Springer. pp. 85-123.

Sullivan KG, Levin M. 2016. Neurotransmitter signaling pathways required for normal development in Xenopus laevis embryos: a pharmacological survey screen. J Anat. 229(4):483-502. doi: 10.1111/joa.12467. Epub 2016 Apr 7. PMID: 27060969; PMCID: PMC5013058.

Thompson, E. (2010). Mind in Life: Biology, Phenomenology, and The Sciences Of Mind. Harvard University Press.

Tremblay M. E., Lowery R. L., Majewska A. K. (2010). Microglial interactions with synapses are modulated by visual experience. *PLoS Biol.* 8:e1000527. 10.1371/journal.pbio.1000527

Turkheimer, F. E., Hellyer, P., Kehagia, A. A., Expert, P., Lord, L. D., Vohryzek, J., De Faria Dafflon, J., Brammer, M., & Leech, R. (2019). Conflicting emergences. Weak vs. strong emergence for the modelling of brain function. *Neuroscience and biobehavioral reviews*, *99*, 3–10. https://doi.org/10.1016/j.neubiorev.2019.01.023

Vahabli, D., and Vicsek, T. (2023.) Emergence of synchronised rotations in dense active matter with disorder. *Commun Phys* **6**, 56. https://doi.org/10.1038/s42005-023-01173-5

Van Duijn, M., Keijzer, F., and Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. Adapt. Behav. 14, 157–170. doi: 10.1177/105971230601400207

Varela, F., Rosch, E., and Thompson, E. (1993). The Embodied Mind: Cognitive Science and Human Experience. Cambridge, Massachusetts: MIT Press.

Varela, F. (1977). On being autonomous: The lessons of natural history for systems theory, in: *Applied General Systems Research - Recent Developments and Trends* (Plenum, New York) pp. 77--84.

Varela, F. G. (1992). Autopoiesis and a biology of intentionality. In B. McMullin & N. Murphy (Eds.), Autopoiesis and perception: A workshop with ESPRIT BRA 3352 (pp. 4–14). Dublin.

Vaux D. L. (2011). In defense of the somatic mutation theory of cancer. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *33*(5), 341–343.

Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., I, & Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Physical review letters*, *75*(6), 1226–1229. https://doi.org/10.1103/PhysRevLett.75.1226

Walsh, D. (2014). *Organisms, Agency, and Evolution.* Cambridge: Cambridge University Press.

Walsh, D. (2018). Objectcy and Agency: Towards a Methodological Vitalism. In Daniel J. Nicholson, and John Dupré (eds), *Everything Flows: Towards a Processual Philosophy of Biology*. Oxford: Oxford University Press.

Wan, K. Y., and Jekely, G. (2021). Origins of eukaryotic excitability. Philos. Transac. R. Soc. B 376: 20190758.20190758. doi: 10.1098/rstb.2019.0758

Wang, S., Shen, T., and Wolynes, P. (2011). The interplay of nonlinearity and architecture in equilibrium cytoskeleton mechanics. *The Journal of Chemical Physics* 134, 014510 (2011); doi: 10.1063/1.3518450

Weber, A., and Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenol. Cogn. Sci.* 1, 97–125. doi: 10.1023/A:1020368120174

Winning, J. & Bechtel, W. (2018). Rethinking Causality in Biological and Neural Mechanisms: Constraints and Control. *Minds and Machines* 28 (2).

Winning, J. (2020). Mechanistic Causation and Constraints: Perspectival Parts and Powers, Non-perspectival Modal Patterns. *British Journal for the Philosophy of Science* 71 (4):1385-1409.

Wioland, H., Lushi, E., and Goldstein, R. E. (2016). Directed collective motion of bacteria under channel confinement. New J. Phys. 18, 075002. doi: 10.1088/1367-2630/18/7/075002

Wolfe, C. 2016. *Materialism: A Historico-Philosophical Introduction.* Berlin: Springer Briefs in Philosophy.

Wu, C., Dai, J., Li, X., Gao, L., Wang, J., Liu, J., et al. (2021). Ion-exchange enabled synthetic swarm. Nat. Nanotechnol. 16, 288–295. doi: 10.1038/s41565-020-00825-9

Yang K., Chen Q.-L., Chen M.-L., Li H.-Z., Liao H., Pu Q., et al. (2020). Temporal dynamics of antibiotic resistome in the plastisphere during microbial colonization. *Environ. Sci. Technol.* 54 11322–11332. 10.1021/acs.est.0c04292

Yuste, R. 2015. From the neuron doctrine to neural networks. *Nature Reviews Neuroscience*, vol. 16: 487-497.

Zammito, J. 2006. Teleology then and now: The question of Kant's relevance for contemporary controversies over function in biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, Vol. 37: 748-770.

Zammito, L. 2018. *The Gestation of German Biology: Philosophy and Physiology from Stahl to Schelling.* Chicago, IL: Chicago University Press.