Organisms as Agents: Beyond the Artifact Model of Teleology

by

Auguste Anatole Nahas

A thesis submitted in conformity with the requirements for the degree of
Doctor of Philosophy
Institute for the History and Philosophy of Science and Technology
University of Toronto

Organisms as Agents:
Beyond the Artifact Model of Teleology

Auguste Anatole Nahas

Doctor of Philosophy

Institute for the History and Philosophy of Science and Technology
University of Toronto

2025

Abstract

The debate regarding the purposiveness of living beings has long been framed in terms of an analogy and corresponding disanalogy between organisms and artifacts. While both are highly organized entities, artifacts derive their purpose or function from an extrinsic source while organisms appear intrinsically purposive. This dissertation traces the legacy of this idea in the 20^{th} century, from behavioristic philosophy and psychology to cybernetics and evolutionary biology. I argue that scholars in these disciplines all adopted "artifact models" of teleology which cast doubt on the distinction between organisms and artifacts and treated the analogy as sufficient to characterize purposiveness. By leveraging the critiques of cybernetics leveled by mid 20^{th} century continental philosophers, I argue that artifact models fail to make sense of organismal purposiveness. This is because such models cannot account for organismal agency as the locus of intrinsic purposiveness and a necessary condition for teleological explanations. I critically analyze recent attempts to conceptualize organisms as agents and characterize the relationship between teleological and causal explanations. Building on the recently developed Ecological Account of agency, I argue that agents are those systems that pursue a particular way of life. This leads to a more nuanced conception of the various explanatory roles of teleology. I then show how this account can provide its own account of artifact functions, which illustrates the conceptual and metaphysical priority of agency over artifacts.

Acknowledgements

We think more than we can say, we feel more than we can think,
we live more than we can feel—and there's much else besides.
—Eugene Gendlin (quoted in McGilchrist 2014)

I am eternally grateful to all those who played a role, small and large, in the writing of this dissertation. A full accounting would span countless pages and not be enough. Thank you, all of you, for everything.

If this journey had a beginning, then it was surely in the classrooms of Justin White, who instilled in me a love of philosophy. My dear great-uncle Ty Cashman, who is known for being in the right place at the right time, also played a pivotal role in those early years.

I owe a huge debt to Ty, Terrence Deacon, Jeremy Sherman, and the rest of the "Pirates" for lighting my mind on fire and pointing me in the direction that would lead me here. And I am grateful to all my professors at McGill University who helped me channel that energy. Special thanks to Maxwell Ramstead, whose guidance and mentorship proved invaluable.

My time at the HPS department at the University of Cambridge, while brief, left a huge imprint. I am so grateful to all the professors and dear friends from that time who shaped this journey and continue to do so, including Alex Rand, Una Jefferson, Zoe Nahas, and Avery Shoemaker.

Over the last few years, I've had the luck of being supported by the most wonderful colleagues and friends. My many discussions with Andrea Gambarotto and Carl Sachs generated countless insights which can now be found in the pages that follow.

Matteo Mossio and Denis Walsh deserve special credit for being the kindest and most brilliant mentors I could have hoped for. Thank you so much for your guidance, patience, and generosity. I'd also like to thank their many other students, and especially Greg Rupik, Emma Sigsworth, and David Rattray, for their friendship and our many great conversations.

Very special thanks to Denis, Matteo, Marga Vicedo, and Alejandro Fábregas-Tejeda, all of whom read through the entire dissertation and provided invaluable advice, feedback, and moral support.

I'd also like to thank the faculty, administration, and my fellow students of IHPST for all the support they have given me over the last four years. I'm so grateful.

Finally, none of this would have been possible without the love and support of all my wonderful friends and family. I dedicate this work to my parents, Anne Le Gall and Anthony Nahas, who have pushed me to pursue what I love from day one. This would not have been possible without them, to say the least. It also would not have been possible without Avery and Soma, who were there during the hardest days, and who taught me new ways to think, feel, and love. And a shoutout to all my dearest and most beloved friends, who kept me sane and laughing, including Rebecca Brosseau, Alex Smith, Nathaniel Lamont, Iris Rappoport, Keelan Horne, Zach Hollander, Hayley Birss, Miko Zeldes-Roth, Ben Mayer-Goodman, Katherine Wright, Mika Wrightson-Rivard, and so many others.

Contents

Introduction 1

I. THE ARTIFACT MODEL FROM BEHAVIORISM TO TELEONOMY 5

1.Purposive Behaviorism and Cybernetics 21

1.1 The Cybernetic Account of Teleology *23*

1.1.1 Behavior, Purpose, and Teleology 23

1.1.2 Mechanism and Behavior in Cybernetics 27

1.2 Cybernetic Teleology in Context *32*

1.2.1 Homeostasis and Teleology 32

1.2.2 Purposive Behaviorism 36

1.2.3 Intrinsic and Extrinsic purpose 43

1.3 Conclusion *51*

2. From Cybernetics to Teleonomy 52

2.1 Pittendrigh and the Origins of Teleonomy *54*

2.2 What Makes Pittendrigh's Teleonomy Cybernetic? *58*

2.2.1 The Purposive Behaviorist Connection: From Behavior to Description 58

2.2.2 Teleonomy and Mechanism 63

2.2.3 Teleonomy, Adaptation, and the Intrinsic/Extrinsic Distinction 64

2.3 Mayr's Teleonomy and the Genetic Program *65*

2.4 What Makes Mayr's Teleonomy Cybernetic? *71*

2.4.1 Proximate and Ultimate Causation 71

2.4.2 Mechanism and the Genetic Program 74

2.4.3 Intrinsic and Extrinsic Purpose: What Lies Beneath Mayr's Critiques? 76

2.5 Conclusion *81*

3. The Continental Critique of Cybernetics 83

3.1 The Continental Critics of Cybernetics *86*

3.2 The Missing Agent: Why Cybernetics Fails to Naturalize Teleology *91*
3.2.1 Individuating Agents 91
3.2.2 Normativity 97
3.2.3 Artifacts 99
3.2.4 Teleonomy and the Continental Critique 102

3.3 Autopoiesis and Teleology *103*
3.3.1 Second-order Cybernetics 103
3.3.2 From Second-Order Cybernetics to Autopoiesis 105
3.3.3 Autopoiesis and the Dispensability of Teleonomy 108

3.4 Conclusion *111*

II. BEYOND THE ARTIFACT ANALOGY 112

4. Two Approaches to Teleological Explanation 122

4.1 The Autonomy Theory: Teleology as a Causal-Consequence Etiology *125*
4.1.1 An Organizational Account of Intrinsic Purpose 125
4.1.2 Explaining Existence with Causes 130

4.2 The Ecological Approach: Teleology as a Non-Causal Explanation *134*
4.2.1 The Ecological Approach to Agency 134
4.2.2 Explaining with Goals 137

4.3 Consequence-Based Explanations: Two Interpretations of Wright's Etiological Account *139*

4.4 Tradeoffs of the Autonomy Approach *144*
4.4.1 The Problem of Action 144
4.4.2 The Problem of Artifact Function 147
4.4.3 The Problem of Normativity 151

4.5 Tradeoffs of the EA *153*
4.5.1 Assuming VS Explaining 153

4.5.2 Are There Really Non-Causal Explanations? 154
4.4.5 Teleological and Equilibrium Explanations 157

4.6 Conclusion *160*

5. Ways of Life 162

5.1 Goal-directed Behavior and Ways of Life *165*

5.2 Self Constitution, Unity, and Way of Life *166*
5.2.1 Aristotle on Bios and Self-Constitution 166
5.2.2 Bios and Contemporary Evolutionary Biology 169
5.2.3 Goals, Ends, and the Autonomy Theory 172

5.3 How and What Ends and Goals Explain *174*
5.3.1 Ends and Coordination 174
5.3.2 What Ends Explain 176
5.3.3 Are End-Explanations Really Teleological? 179

5.4 Way of Life and Essentialism *181*
5.4.1 Typological and Teleological Essentialism 182
5.4.2 Evo-Devo and Teleological Essentialism 184

5.6 Conclusion *189*

6. Functions and Teleological Explanations 191

6.1 The Goal-Contribution Account of Function *194*

6.2 Functions, Accidents, and Teleological Explanations *198*

6.3 Function Ascription and the Importance of Context *203*

6.4 Biological Functions *209*

6.5 Conclusion *211*

Conclusion 213

Bibliography 216

Introduction

The teleology in modern biology is analogical. The organic world
seems as if it is designed; therefore we treat it as designed.
The artifact model is the key to biological teleology.
— Michael Ruse (1981, 93)

This dissertation is concerned with the purposiveness of living beings and the nature of teleological explanations. Consider, for instance, the countless examples of camouflage in the animal kingdom, from the leaf-tailed gecko, to stick insects and seahorses. These creatures are exquisitely organized to mimic features of their environment and do so to seamlessly blend into their surroundings. These structures are propagated through complex and robust reproductive and developmental processes that appear reliably directed at the production of offspring with the same structures and capacities. And these organisms, in turn, display complex adaptive behaviors that enable them to hide from predators and harvest resources necessary for their survival. The appearance of such phenomena, and the ubiquity of purpose and function in biological discourse, has long posed great conceptual challenges insofar as they require explanations that appeal to goals, norms, functions, organizational principles, vital forces, and the like, all of which seem incompatible with proper scientific explanations that only cite causal-mechanical facts. The artifact model of teleology has been thought to present an elegant framework for resolving this problem: to make sense purposive phenomena, and the teleological mode of explanation they call for, we need only view organisms as if they were designed artifacts.

The debate regarding the purposiveness of living beings has long been framed in terms of an analogy and corresponding disanalogy between organisms and artifacts. While both are highly organized entities, artifacts derive their purpose or function from an extrinsic source while organisms are intrinsically purposive. In the 20th century, several scholars adopted artifact models of teleology which cast doubt on the distinction between organisms and artifacts and treated the analogy as sufficient to characterize purposiveness. In what follows, I argue that such attempts have failed. By attempting to collapse the distinction between organisms and machines, these approaches have radically distorted our understanding of purposiveness and teleological explanations. This is most evident in their inability to make sense of the distinction between intrinsic and extrinsic purpose, but it is also clear in their inability to make sense teleology as a distinct, scientific mode of explanation—rather than a mere heuristic or description—irreducible to a causal mode of explanation. In turn, I defend the view that the ultimate source of purposiveness is to be found in the agency of organisms. In doing so, I will show how a proper understanding of artifact purposes or functions are subsidiary to an account of agency. In other words, instead of deriving an account of the intrinsically purpose agency of organisms from artifacts, we must do the opposite.

My argument comprises two parts. The first is concerned with the legacy of the cybernetic account of teleology and those that it inspired, most notably "teleonomy". I approach the issue by attempting to understand why supporters of teleonomy, most notably Colin Pittendrigh and Ernst Mayr, claimed that their approach was deeply influenced by the cybernetic movement, despite clear differences in their views. My focus will be primarily on the philosophical issues this question raises, though it raises many important historical ones too. In so far as this work may be of value to a historian, it will be in virtue of the particular historical questions that an interdisciplinary perspective such as mine raises, including why teleonomy was deemed to be an appropriate resolution to the problem at that point in time.

Chapter 1 begins with a close reading of Rosenblueth, Wiener, and Bigelow's 1943 paper *Behavior, Purpose, and Teleology*, which is thought to have formed the cornerstone of the then-emerging cybernetic movement. Their account is a paradigmatic case of the artifact model of teleology, based on an analogy between organisms and machines regulated by negative feedback, or servomechanisms. The article is now famous for proposing that "teleological behavior" can be explained by such negative feedback mechanisms. The second half of the chapter traces the conceptual lineage of these ideas to earlier work in behavioristic philosophy and psychology. It reveals that notable ambiguities regarding both the distinction between intrinsic and extrinsic purpose and the place of teleology as a distinctive mode of explanation were inherited by cybernetics from purposive behaviorism.

In Chapter 2, I critically analyze Pittendrigh and Mayr's claim that teleonomy is largely inspired by cybernetics. First, I show that their respective senses of teleonomy are different, and I argue despite claims to the contrary, neither carries over the most distinctive features of the cybernetic account. However, I also show that the differences between these senses of teleonomy stem from distinct attempts at resolving the same set of problems which the cybernetic account of purpose aimed to solve.

In Chapter 3, I revisit several critiques that were leveled at the cybernetic account and show that they also apply to teleonomy. At the heart of these critiques is the idea that the behavioristic, cybernetic, and teleonomic conceptions of teleology all failed to recognize that goals must have their origin in an agent to whom one can ascribe that goal. The very meaning of teleology requires a notion of agency to ground it.

Part 2 of this dissertation brings my analysis to the present day and defends a particular conception of purposiveness that builds on the recently proposed Ecological Account (EA) of agency.

Chapter 4 compares the ecological account to another recently proposed account dubbed the Autonomy Theory (AT). I use the argument of Part 1 to motivate a distinction between the way one conceptualizes the phenomenon of intrinsic purposiveness or agency, which acts as a necessary *condition* for teleological explanation, from the way one conceptualizes the *content* of teleological explanations themselves. This analysis reveals that the AT is primarily focused on giving an account of intrinsic purpose in terms of its physical realization, at the expense of a more restricted conception of teleological content. In contrast, the EA is more concerned with making sense of the rich content of teleological explanations, at the risk risks of underspecifying the conditions under which it applies. The two chapters that follow leverage the distinction between the content and conditions of teleological explanations.

In Chapter 5, I argue that while the EA treats the phenomenon of intrinsically purposive agency as an empirically observable fact, there is more work to be done in conceptualizing it. I argue that agency is characterized by the pursuit of a way of life, which is realized through the coherence of myriad goal-directed activities across the life cycle of the organism. This leads to a more nuanced understanding of the explanatory roles of teleology.

Chapter 6 closes the dissertation by returning to the topic of artifacts and their functions. I defend an account of artifact and biological function in the context of teleological explanation. Function statements, in my view, explain why a process or entity takes part in a goal-directed activity, and in so doing contributes to an agent's goal. This is a complete inversion of the supposition employed by artifact models of teleology discussed in Part 1. It shows that one cannot analyze agential teleology in terms of artifact teleology, because the latter presupposes the former.

Teleology is one of the perennial problems of philosophy and has more recently earned its place as a major topic of debate in philosophy of biology as a modern discipline. To paraphrase Wilfrid Sellars, it inexorably brings into play the major concerns of philosophy of science, including the nature of causation, models, metaphors, and analogies in scientific explanation and understanding (Sellars 1962, 527). The question, then, is not whether making sense of purpose and teleology is worth our while; rather, it is about how the ways in which the problem is dealt with either enables us or misleads us in our understanding living beings.

I

THE ARTIFACT MODEL
FROM BEHAVIORISM TO TELEONOMY

In 1974, evolutionary biologist Ernst Mayr proclaimed the problem of teleology entirely resolved thanks to a series of conceptual and empirical breakthroughs that married the language of computers and information theory with molecular biology (Mayr 1974, 98). This verdict carried some weight. Mayr was one of the most prominent biologists of his day, and one of the founders of philosophy of biology as a modern discipline (Smocovitis 1994; D. L. Hull 1994; Nicholson and Gawne 2015; Peterson 2016; Rupik 2024). He recognized teleology as both a canonical problem of philosophy and one of the core conceptual issues that the new biology, which he and others dubbed "The Modern Evolutionary Synthesis", needed to address.

As Mayr recognized, it is a testament to the enduring importance of Aristotle's thought that mention of him remains unavoidable in discussions of teleology. One of the most significant questions which he posed is whether living things can be understood like artifacts. Since then, the organism-artifact (dis)analogy has periodically generated new interest and controversy with the emergence of new and more complex machines, from the clocks and automata of the Renaissance to the engines of the 19th century, to the computers of the 20th (Tresch 2012; Riskin

2016; Jones-Imhotep 2020). These new kinds of machines, along with changes in the philosophical, theological, and socio-political outlook of scholars over time, reconfigured the way the analogy was conceptualized.

In the modern era, one of the most careful treatments of the organism-artifact (O-A) analogy was given by Immanuel Kant in his *Critique of the Power of Judgment* ([1790] 2000). He saw organisms and artifacts as analogous because they exhibit intricate order and structure that we associate with the products of rational thought and skilled craft, whether human or divine (Ginsborg 2004).[1] Such order is otherwise highly contingent, and therefore extremely unlikely to have emerged spontaneously from the powers of matter alone. In this sense, this order is *inexplicable* in terms of such powers; and this is equally true for organisms and artifacts. But there is also a notable disanalogy between organisms and artifacts, namely the fact that organisms do things—grow, self-repair, and reproduce—that no artifice has ever been able to replicate (ibid). For Kant, this analogy and corresponding disanalogy leads to two 'senses' of purposiveness: one associated with the appearance of design which is common to artifacts and organisms, which is usually dubbed 'extrinsic' purposiveness, and one with the appearance of agency uniquely found in the living, which is dubbed 'intrinsic' purposiveness.[2] For both Aristotle and Kant, intrinsic purpose is a defining feature of organisms just as extrinsic purpose defines artifacts.[3] One speaks of an artifact like a knife having extrinsic purpose because that purpose or function is bestowed upon it by an external source, i.e., its creator or user. Organisms, in contrast, appear purposively directed toward their own growth, reproduction, and survival in virtue of their own endogenous activity.

---

[1] According to Hannah Ginsborg (2004), this is also true of Aristotle.

[2] It should be noted that this usage of "intrinsic" and "extrinsic" is different from the notion of intrinsic and extrinsic *properties*, where the former are those that are causally independent from external forces (Ellis 2002).

[3] See Witt (2015) for a contrary view of Aristotle's teleology; see also Lennox (2024).

Such a distinction between intrinsic and extrinsic purpose implies that organisms are the kind of things that can be intrinsically purposive, and artifacts like knives are not. What is it about organisms that makes them such? The Aristotelian answer, famously maligned in the early modern period for being unintelligible, is that it is in the *nature* of living things—their natures being constituted by their having a soul—to be intrinsically purposive (c.f. Mercer 1993; Osler 2001). Kant is often thought to have offered a decisively modern alternative (c.f. Toepfer 2011; 2012; 2023): organisms are intrinsically purposive insofar as they are organized in a fundamentally different way, namely, *self-organized*, such that each of their parts is reciprocally means and ends of another. Kant used this idea to frame the distinction between extrinsic and intrinsic purpose in terms of two nested problems: first, given what we know about the physical laws that govern the universe, how is the emergence of complex, intricately ordered structures possible? Second, how is it that a subset of those intricate organizations—namely, the biological ones—display the distinctive goal-directed behaviors that they do?

Many have argued that Darwin resolved the problem of teleology in the 'extrinsic' sense by articulating how, through a process analogous to the artificial selection and breeding of organisms, evolution could give rise to the intricate structure of living beings.[4] If Darwin solved this first sense of the problem, Mayr believed that 20th century genetics had cracked the second. The intrinsic goal-directedness of organisms was to be explained by the fact that they are controlled and regulated by a *genetic program*. This program was thought to be genetically encoded information which provides the instructions which every cell follows in its development and behavior. The genetic program promised to be a powerful resolution to the problem of teleology in part because it brought together the two sides of the problem. The genetic program is 'written' and 'edited' by the process of natural selection in phylogeny, and 'read' or 'transcribed' in ontogeny. Mayr

---

[4] According to Ernst Haeckel, this made Darwin 'the Newton of the grass blade' whose very possibility Kant had categorically denied (Haeckel 1889, 95–95; Kant [1790] 2000; see also: Lenoir 1982; Cornell 1986; Zammito 2012; Moss and Newman 2015).

saw this as such a radically new conception of biological teleology that he, following Colin Pittendrigh (1958), suggested a new term to characterize it, 'teleonomy', defined as any "process or behavior ... which owes its goal-directedness to the operation of a program." (Mayr 1974, 98).

Part 1 of this dissertation seeks to answer the following questions: in what sense is teleonomy an adequate, let alone complete solution to the problem of teleology? Is it really a resolution to the problem, or a cryptic abandonment of it altogether? And if it is the latter, what would it take to address the problem adequately?

Teleonomy exemplifies what I will call the *artifact model of teleology* (after Ruse 1981, 93). As I see it, this is a particular conception of intrinsic purpose that scholars believed could be applied equally to organisms and artifacts. In this view, intrinsic purpose is no longer something that the analogy fails to capture, as both Aristotle and Kant saw it, but something that falls squarely within its purview. This amounts to a fundamental revision of the O-A analogy; one which, I contend, radically distorts the very meaning of purpose. In the two following chapters, I will argue that this distortion is at the heart of teleonomy but does not originate there. This transformation dates to the early years of the cybernetic movement, and further still to the thinking of philosophers and psychologists in the tradition of behaviorism and American Pragmatism. Placing cybernetics within this intellectual lineage will offer a clearer picture of its account of teleology, which in turn will illuminate the shortcomings of teleonomy as an artifact model of teleology. Specifically, I will focus on three closely interrelated issues that define the cybernetic account of teleology, all of which one can also find in the literature on purposive behaviorism and teleonomy. The first lies in the purported compatibility between mechanism and teleology, which is predicated on a particular way of understanding the analogy between organisms and machines. The second concerns whether the purposiveness of living beings is best characterized in terms of their behavior rather than their structural organization, which has implications for how one conceives of teleological explanations. The third, concerns the nature of intrinsic purposiveness and what distinguishes from extrinsic purpose. I will now discuss each of these in turn.

Analogies, Metaphors, and Machines

The analogy between organisms and artifacts, and more specifically between organisms and machines, is at the heart of the debate over teleology. As discussed above, the analogy generates two senses of teleology; but it also implicitly frames the way we think about the opposition or compatibility between teleology and mechanism, where the latter usually means a reductive explanation of a phenomenon in terms of its component parts. Much of the confusion about the meaning of teleology (intrinsic and extrinsic), the meaning of mechanism, and their relationship can be traced back to the analogy. For this reason, the O-A analogy has been central to the various failed attempts to illuminate the nature of teleology.

Both analogies and metaphors are indispensable tools for thinking and communicating in virtue of the way they illuminate the unknown via the known (Black [1962] 2019; Haraway 1976; Hesse 1988; Keller 2003). Whether or not there is a meaningful difference between analogies and metaphors is a matter of debate. I will treat metaphors as figures of speech which are intended to convey their meaning directly without being noticed. As many have remarked, language is saturated with metaphors, to the point where it seems that language is primarily metaphorical, and literal in exception only. This has important implications for science. Metaphors play an essential *cognitive role* in science, most of all in communication and understanding (B. Smith 2009; Nicholson 2019; Stuart and Wilkenfeld 2022). But for this reason, they can also be misleading and pernicious. In fact, by one account metaphors are useful precisely because their meaning is essentially open to interpretation and prone to elision (Keller 2003).

Analogies, on the other hand, refer to more precise and explicit attempts at uncovering invariant features between two entities where the comparison is made explicit, as in “X is to Y as A is to B”. For this reason, analogies—unlike metaphors—can be properly taken to be forms of reasoning and argument. In their book-length treatment of analogical reasoning in Charles Darwin’s argument for natural selection, Roger White et al. (2021) distinguish two main senses of analogy. The first is analogy as similitude, which is a form of ampliative reasoning that makes inferences about unknown similarities between entities based on a series of known similarities. For example, “Saturn, Mars and other planets are known to resemble the Earth in orbiting and being lit by the Sun, in having their own moons and so on. Inferably, then, they probably resemble the Earth in being inhabited.” (ibid, 6). Like any ampliative inference, such conclusions are never certain and always strengthened by the discovery of further similarities.

The second is analogy by proportion, which involves the specific analogical comparison between two relations. In such an argument, one takes a particular relation *R1* between two terms *A* and *B* in one domain, to infer a similar relation, *R2*, between two terms *C* and *D* in another domain. A simple mathematical example might be “1 is to 2 as 5 is to x”, where one infers the value of x. An empirical counterpart might be as follows: “Socks are related to feet as gloves are related to hands. Since socks warm feet, gloves, which cover hands as closely as socks cover feet, are correctly inferred to warm hands.” (White et al. 2021, 2). In the mathematical case, one infers a mathematical relation; in the empirical case, one infers a causal relation of warming. According to White and colleagues, analogies of proportion began in the mathematics in Ancient Greece and were first taken up by Aristotle to be used in an empirical context. This use of proportional analogy appears across his corpus, including in his articulation of the contrasting O-A analogy and disanalogy introduced above, which can now be explained with more precision.

In the *Generation of Animals* II.1, Aristotle uses his fourfold distinction between types of causes to argue that the matter from which an artifact is composed is insufficient to explain how it comes to be. In addition to this material cause, a statue requires the efficient causes of an artisan’s hands, along with the goals which guide its movements. In this sense, artifacts are inexplicable by the powers of matter alone. The same is true of organisms and their component matter: just as there is a causal relation between the movement of artisans and the artifacts they create, there is also an analogical relation between the movement of the male parent and their offspring (Barnes 1984, vol. 1, ll. 734b27-735a3). In Kant, philosophical (rather than mathematical) analogies play a similar role, and this is unsurprising given how much his logic was taken from Aristotle.[5]

Unlike analogies of similitude, analogies of proportion do not seek to infer unknown similarities between organisms and artifacts, nor do they attempt to establish the fact that the analogies are more numerous and important than the disanalogies. On the contrary, analogies of proportion involve acknowledging the differences between the domains in question and seek to identify "partial isomorphism" rather than similarity in the sense of "generic identity" (M. Esposito and Baravalle 2023). Arguments about whether organisms 'really are' just like machines are implicitly based on analogies of similitude (ibid; see e.g., Nicholson 2019; Bongard and Levin 2021). After enumerating several similarities or dissimilarities, one induces the fact of essential sameness difference between them. Counter arguments then involve finding different kinds of machines or organisms that might invalidate the inference. But such debates are futile, and inexorably lead to confusion since they ultimately rest on the definition of 'machine', which is always historically contextual, fleeting, and ambiguous. There is no essential, generic sense of machine upon which one can infer an essential identity between them and organisms. In other words, to say in what sense organisms are 'machine-like' requires specifying what kinds of machines we have in mind. It is from such a perspective that anti-mechanistic thinking has often been criticized: "There is no question but that the animal organism is different from any

[5] It appears that Kant used philosophical analogies both in the similitude and proportional senses (Callanan 2008, 751–53; van den Berg 2018, 68–72) though matters are complicated by the role analogies play in the broader context of his project to systematically rehabilitate metaphysics (see also: Breitenbach 2014; Nassar 2016; van den Berg 2024). Van den Berg (2018) makes the distinction in terms of simple analogies and analogical arguments. As he sees it, the former are a direct comparison between two entities whereas the latter resemble the kind of ampliative inference that White et al. call an analogy as similitude.

type of machine so far devised by man. But it may yet prove to be a machine or mechanism of some type peculiar to itself… *There are machines and other machines; and our knowledge of mechanical possibilities is not yet complete*." (Warren 1916b, 37; emphasis added).

In light of this, Esposito and Baravalle (2023) suggest that we restrict ourselves to analogies of proportion, where they are genuinely productive. In this way, analogies of proportion resemble a particular kind of modelling relation, whereby machines are constructed as physical models for particular biological processes which we wish to comprehend (Black [1962] 2019). The mistake is to forget the limited domain of application of these models and unwittingly reify them, or to treat the analogy as similitude. As I see it, this pernicious reification tends to happen precisely when analogies become metaphors. Such a metaphorical shorthand collapses the carefully circumscribed analogical distance between the relata in question and becomes a way of speaking that may suggest more (or less) than what was originally intended. In this regard, the essential ambiguity of metaphors makes them rhetorically useful and dangerous in equal measure. As Norbert Wiener is claimed to have said, "the price of metaphor is eternal vigilance" (Lewontin 2000, xv).

In what follows, I will suggest that the distinction between these two types of analogy is crucial for understanding the way 20th century scholars re-shaped our understanding of the O-A analogy around particular kinds of machines, and what it meant to provide a "mechanistic" account of teleology.

From Behaviorism to Cybernetics and Molecular Biology

According to Ernst Mayr, teleonomy marked a veritable "breakthrough in thinking about teleology" which was made possible by "the introduction of new concepts from the fields of cybernetics and new terminologies from the language of information theory" (Mayr 1974, 92). The cybernetic movement was an ambitious if short-lived attempt at creating a universal discipline which would apply insights from mathematics and engineering to the life and social sciences (Heims 1991; Galison 1994; Cordeschi 2002; Kline 2015). More precisely, the field was founded on an analogy between the behavior of particular *kinds* of machines, namely, information-feedback machines, and organisms.

Both Pittendrigh and Mayr referred to a 1943 paper by Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow entitled *Behavior, Purpose, and Teleology* which is widely considered the founding document of cybernetics. Less well known is the broader conceptual role that teleology played for the movement in its infancy. In 1944, Wiener brought together a group of experts in engineering, mathematics, and neurophysiology for a meeting on a range of topics of common interest. Wiener, John Von Neumann and Howard Aiken suggested they name it *The Teleological Society*, and considered starting a periodical they named *Teleology* or *Teleologia* (Aiken, Von Neumann, and Wiener 1944). Around this time, they began to search for a name for their newly established "science", which they described as being "devoted on the one hand to the study of how purpose is realized in human and animal conduct and on the other hand how purpose can be imitated by mechanical and electrical means." (ibid). Rosenblueth et al's paper formed the cornerstone of the emerging field, and the emphasis on teleology carried over to the post-war Macy conferences, sponsored by the Josiah Macy Jr. Foundation.[6] The first, held in March 1946 was entitled 'Teleological Mechanisms and Circular Causal Systems' and was followed by three more meetings in the fall of the same year, all of which were devoted to teleological mechanisms (Von Foerster, Mead, and Teuber 1953, xix). For Macy Group-member Gregory Bateson, the movement represented "the greatest single shift in human thinking since the days of Plato and Aristotle because it ... finally resolves the problems of teleology and the body-mind dichotomy..." (Ruesch and Bateson 1951, 177). In 1948, Wiener christened the movement in his best-selling *Cybernetics*: *Or Control and Communication in the Animal and the Machine* (Wiener [1948] 2019). In the introduction he described the

[6] There were ten such conferences between the years 1946-1953. Transcripts of the last five were published, and were recently reprinted (Pias 2016).

book as a direct outcome of his collaborations with Rosenblueth and Bigelow which, while separate at first, converged in their joint publication of the 1943 paper. The paper is now remembered for defining teleology in terms of negative feedback.

Though the connection between teleonomy and cybernetics is well known, the precise nature of this relationship has not received much attention. It is natural to view teleonomy as part of a broader uptake of concepts from cybernetics and information theory by molecular biology, including notions like the *genetic program* or *code*, along with its *transcription*, *translation*, and so on (Kay 2000). But when it comes to the precise connection between teleonomy and the cybernetic account of teleology, things are less clear. One problem is that the cybernetic focus on teleology was very short lived and was already over when the movement was christened in 1948. While the key concepts associated with teleology, such as adaptivity and homeostasis, were further developed by Ross Ashby and others, almost everything of substance that was said on the topic of teleology is contained in the nine-page long 1943 paper which, by the authors' own admission, left many questions unanswered. One possible reason for this is that the proto-cybernetic focus on teleology was somewhat rhetorical: that its ambitions as a unifying science required showing the compatibility between the mechanistic explanations of the physical sciences and the teleological explanations of the life sciences (Abraham 2012; see also: Bowker 1993; Galison 1998; Bates 2014). Viewed in this light, to say that machines are teleological and organisms are machines is to license the unrestricted movement of concepts from the physical sciences to the life sciences. Teleology, long a concept that had stood as a bulwark against such reduction, now played the opposite role. Teleology seemed to have given the burgeoning discipline a sense of identity in its early years (see: Brigandt 2020, 7).

Why then did this rhetoric not persist into the late 40s and beyond? My tentative suggestion is that its role was taken over in 1948 by a newer term, *information*, which came to prominence around the same time that references to purpose and teleology died in the cybernetic literature (Morange [1998] 2020). This new term could fulfill the same rhetorical functions as teleology, particularly in its unificatory role (Wilkinson 1961). "Information" signaled the continuity between the mechanical and the living; it licensed the unchecked use of machine analogies, which quickly turned into machine metaphors; and it did all that without the baggage of two thousand years of debate. And yet, perhaps precisely due to its brevity and ambiguities, the short 1943 paper on teleology had a profound and outsized impact on the debate in the latter half of the 20th century. Its reverberations can be felt most acutely in Pittendrigh and Mayr's teleonomy, along with the work of analytic philosophers of science, such as Ernest Nagel. By my best estimations, it now stands as the most cited modern piece of work on teleology, and the third most cited paper in *Philosophy of Science*.[7] For these reasons, it is important that it be properly situated in the intellectual currents of the 20th century.

Ostensibly, the profound "breakthrough" which Mayr saw in the cybernetic account of teleology referred to the way in which it was rendered compatible with mechanism. In the cybernetic view, there is no fundamental difference between the goal-directed behavior of a torpedo and one of a cat chasing a mouse. Intrinsic purpose, qua goal-directed behavior, is now illuminated in virtue of an *analogy* between organisms and artifacts, rather than a disanalogy. I take this to be a defining feature of an artifact model of teleology. Yet it is crucial to notice that the cyberneticians' argument fit the description of an analogy of *proportion*, rather than an analogy as similitude. As a physiologist by training, Arturo Rosenblueth was well aware of how different machines are from organisms in their material components and structure. He and his co-authors did not downplay these differences but recognized that they were irrelevant for their argument (Rosenblueth, Wiener, and Bigelow 1943, 23). What they believed they had uncovered was a specific analogical *relation*, in the form of behavior, between certain machines and their surrounding environments on the one hand, and organisms and their environments on the other. So, while it is true that cybernetics

[7] My thanks to Alejandro Fábregas-Tejeda for bringing this to my attention. See: https://www.cambridge.org/core/journals/philosophy-of-science/most-cited

gave a "mechanistic" conception of teleology, such claims must always be understood as being based on, and restricted to, a particular analogy with a particular kind of machine.

As several historians have emphasized, cyberneticians were far from the first to theorize feedback mechanisms Mindell (2002) has critiqued the way Wiener crafted a mythology around the emergence of cybernetics which highlighted the legacy of great minds such as Leibniz and Gibbs as ancestors of the movement, while obscuring the work of engineers and mathematicians during the decades prior to cybernetics. What Mindell and others have not emphasized is that the same is true of the philosophers that Wiener knew and learned from, and who developed behavioristic accounts of teleology of their own that cybernetics drew from. These philosophers and psychologists were concerned with making mechanism and teleology compatible by adopting a "middle way" between extreme forms of mechanism and vitalism. In this way, they reflected the broader trends of the interwar period which saw a veritable flourishing of such thought, which might be dubbed as broadly organism-centric or holistic, which stretched from biology, to medicine, and psychology (Lawrence and Weisz 1998; M. Esposito 2011; Nicholson and Gawne 2015; M. Esposito 2017; Peterson 2016; Baedke 2019; Brandt 2022). These included the organicist movement in the USA and England, various forms of emergentism, theoretical biology in Germany, and certain strands of psychology, such as Gestalt psychology in Europe and behaviorists in North America who took seriously the reality of purposiveness. As I shall show in the next chapter, there is good reason historically and philosophically to think that cybernetics was at least partly in continuity with such movements, particularly—though not exclusively—with purposive behaviorism. In doing so, I hope to contribute to the ongoing conversation regarding the need for reevaluating the legacy of American Pragmatism in 20th century analytic philosophy (Misak 2013; 2016), psychology (Heft 2001), and biology (Pearce 2020). For this reason, the legacy of this thought is felt especially in recent work that re-combines these once-intertwined stands that frayed over the course of the 20th century (Walsh 2015).

The conceptual continuity between purposive behaviorism and cybernetics highlights the fact that the analogy between the purposive behavior of servomechanism and organisms already assumed a definition of intrinsic purpose as a kind of *behavior*. This fact is crucial, yet rarely acknowledged, because few now remember the attempts of American psychologists and philosophers to conceptualize teleology within a behavioristic conception of life and mind. This school of thought, which I will dub *purposive behaviorism*, was supposed to be an alternative to—or in some cases, merely an improvement on—the thinking of overly reductionistic behaviorists like John Watson. Defenders of purposive behaviorism, such as Edward Tolman, insisted on the distinction between 'molar' and 'molecular' behaviorism, corresponding to holistic—or what some called 'properly psychological' behaviorism—and reductive conceptions that seemed to approach physiology (Calkins 1921; Perry 1921, 85). Tolman and others such as his Harvard professors Edwin Holt and Ralph Barton Perry, as well as Clark Hull, drew analogies between purposive behavior of organisms and machines like thermostats. As Hull saw it, the robot approach was a way to keep oneself from straying toward vitalism, i.e., a way of ensuring that one's theorizing kept within the bounds of naturalism (C. L. Hull 1943, 27–28; Craik 1943, 10; see also: Cordeschi 2002, 20). This fear reflected a more general anxiety about the properly scientific status of psychology—and the life and social sciences more broadly—which cybernetics promised to resolve.

For purposive behaviorists, as well as cyberneticians and defenders of teleonomy, one of the keys to making teleology scientifically respectable was to legitimate the use of teleological *explanations.* As has long been recognized, teleological explanations imply backward causation, by citing a future event—a goal—causing their means in the present. Various strategies were attempted to deal with this problem. Purposive behaviorists such as Edwin Holt argued that one could formulate teleological laws of behavior that mirrored the explanatory uses of laws in physics. Cyberneticians distinguished explanations operating at the "behavioral mode of study", which treat the internal workings of a mechanism as a black box, from causal explanations operating at the "functional mode of study". Colin Pittendrigh distinguished between a description in terms of goals and an explanation in terms of goals. Finally, Mayr, thought teleology could be understood as a species of causal explanations, albeit one whose originating cause is to be found in the genetic program.

In sum, one cannot understand teleonomy without seeing how it emerged from attempts at finding a properly naturalistic place for teleology by locating it within the analogy—rather than disanalogy—between organisms and artifacts.

The Missing Agent

The foregoing analysis, which will be covered in depth in the first two chapters, is necessary for seeing the fundamental shortcomings of these views; considerations which will also prove invaluable for making clear the desiderata for a proper account of teleology. My critique will reinforce the connection which I have elucidated between cybernetics and teleonomy, by showing how many of the critiques that were levelled at cybernetics in the 1950s and 60s also apply to teleonomy (Churchman and Ackoff 1950; Taylor 1950a; 1950b; Von Bertalanffy 1950; Wisdom 1951; Scheffler 1959; Wimsatt 1971). My focus will be on the most radical of these, levelled by continental philosophers, most notably Hans Jonas, Georges Canguilhem, and Gilbert Simondon (among others, e.g.: Ruyer 1961; Merleau-Ponty 1964). Though their critiques ranged in scope, content, and motivation, I argue that a clear convergence can be found in their arguments that cybernetics erased the organismal individual, self, or subject—what I will call "the agent"—which they saw as an integral component of any theory of biological purposiveness. This key overlap, in turn, provides the tools for understanding that the fundamental difference between teleonomy and many other "teleological" accounts was not, as Mayr would have it, that the latter appealed to vital forces or backward causation (Mayr 1974, 92–93). Rather, we find in teleonomy an explicit attempt at naturalizing purpose without any recourse to the notion of an individual as both the intrinsic locus from which purposiveness originates, and the beneficiary of these goals as they are accomplished.

What the continental critiques clearly highlight is the philosophical move that cybernetics willingly or unwillingly legitimated, and how it radically alters the meaning of teleology. What interests me here is the very idea—never explicitly stated, but strongly implied—that extrinsic purpose is all there is. If this is the case, then teleology must be something very different than what we usually think it to mean. Extrinsic purpose is, after all, a relational property. It is parasitic on something else being intrinsically purposive--so it can't be all there is, by definition! To claim otherwise is to perniciously reify something that doesn't truly exist, such as an intrinsically purposive "blind watchmaker" which designs organisms over the course of evolution (Winther 2020; Mazviiita Chirimuuta 2020; Walsh and Rupik 2023). And yet this is precisely what I believe defenders of cybernetics and teleonomy did when they argued that intrinsic purpose could be illuminated by the O-A analogy (rather than the disanalogy). This move presupposes that extrinsic purpose is separable or independent from intrinsic purpose in the traditional sense, and it is pervasive in both the cybernetic account and teleonomy, and even the various accounts of biological function that emerged in the 1970s. The fundamental problems with these claims were systematically obscured by defenders of the cybernetic and teleonomic views who reinscribed the notion of intrinsic purpose within an artifact model, which simultaneously presupposes and brackets what it is that grounds the purposiveness of artifacts in the first place: the agent who creates and uses the artifact. For teleology to be anything like what we mean by "purpose", it must be intrinsic purpose, meaning that it must include a grounding locus, what has been called an 'agent', 'self', 'subject' or 'beneficiary'. To say that teleonomy is an adequate—let alone complete—solution to the problem of teleology means giving an account of teleology without agency, and therefore failing the task of naturalizing teleology. And in this regard, teleonomy stands as an exemplar of the failings of the artifact model of teleology.

My standards for success and failure are pragmatic: we humans are intimately familiar with teleological explanations, because we use them daily to explain the behavior of our fellow human and non-human organisms. To offer a naturalistic account of a concept like purposiveness is to make sense of what we are doing in these instances, given the ever-evolving, patchwork picture of the world offered by our best sciences. This process may well require some distortion; it may be that, under certain reasonable constraints of naturalism, the content of a teleological explanation isn't quite what we thought it to be in the first place. But for any concept to be naturalized rather than reduced or eliminated means not overly distorting the meaning of the concept. And if there is significant distortion the onus is on the naturalizer to show why the distortion is warranted and why it preserves the essential "core" of the concept. As I see it, the intrinsic/extrinsic distinction, and the role that agency plays in that distinction, is precisely what lies at the 'core' of teleology, and to give an account of teleology without agency necessarily means that one has either eliminated the concept or failed at naturalizing it.

CHAPTER ONE

Purposive Behaviorism and Cybernetics

The 1943 article *Behavior, Purpose and Teleology* by Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow is one of the most influential pieces of scholarship on teleology in the 20th century. Building on decades of research on negative feedback mechanisms, the authors famously argue that such mechanisms explain the goal-directed behavior of organisms (Mindell 2002). In this regard, the paper stands as one of countless attempts to tear down any putative barrier between studies of the inanimate and the living; to show, in other words, that organisms can be understood in the same manner as the rest of nature. But while many of these debates had a distinctly metaphysical flavor, this project had other motivations: it was an attempt to lay the methodological groundwork for a universal discipline, later called *cybernetics*, that could unify the study of complex systems across the physical, life, and social sciences (Abraham 2012).[8]

Twenty five years after its publication, the paper was already considered a "classic" in philosophical circles where the topics of explanation, reduction, and teleology were in vogue (Wright 1968). But despite its enduring influence, the central thesis of *Behavior, Purpose and Teleology* remains, in my view, largely misunderstood. To remember it as one among many attempts to mechanize teleology risks ignoring both what is most distinctive and strange about this attempt at doing so, namely, within a *behavioristic* framework. Attending to this

---

[8] I will follow others in using the label "cybernetic" to describe the account of teleology put forward in Rosenblueth et al.'s 1943 paper, despite it being anachronistic.

context is crucial, for two reasons. First, it provides a more detailed and accurate picture of the way in which the authors defined teleology, and the sense in which it could be reconciled with mechanism. Second, it allows us to properly situate the cybernetic account of teleology within a broader, distinctively American school of thought about teleology which I will dub *purposive behaviorism.* This chapter aims to do both of these things, in order to arrive at a threefold characterization which captures what is distinctive about the cybernetic account of purposiveness. This will provide the groundwork for the two chapters that follow, which will critically evaluate the relationship between it and teleonomy, along with a raft of critiques which were leveled at cybernetics in the 1950s and 60s.

In the next section (1.1), I provide a close analysis of the 1943 paper. I begin by emphasizing the importance of the behavioristic framework within which the authors define purpose and teleology (1.1.1). I use this analysis to explain the sense in which the cybernetic account of teleology was mechanistic (1.1.2). In section 1.2, I will locate the cybernetic account of teleology among the many others which were developed by non-reductive thinkers in behavioristic psychology and philosophy in the decades prior to cybernetics (1.2.1). I will introduce purposive behaviorism as a school of thought which included the likes of Edwin Holt, Ralph B. Perry and Edward Tolman, whom Wiener knew from his time as a graduate student at Harvard University (1.2.2). I will not only argue that Wiener and his colleagues understood the concept of teleology in a way that was in continuity with these philosophers and psychologists, but also that it carried within it problematic ambiguities about the distinction between intrinsic and extrinsic purposiveness that permeated their thinking (1.2.2). To conclude, I will suggest a three-fold characterization of the cybernetic account of teleology as one that characterizes purpose: first, as a property of a system's gross dynamics or behavior; second, as one apt to be modeled with artifacts, and particularly machines; and third as one that was unable to make sense of the distinction between intrinsic and extrinsic purpose. This will prove important in the next chapter, when I examine the continuities and discontinuities between cybernetics and teleonomy.

1.1 The Cybernetic Account of Teleology

1.1.1 Behavior, Purpose, and Teleology

Behavior, Purpose, and Teleology is first and foremost a proposal for classifying different kinds of behaviors in a hierarchy, to then identify purposive and teleological behaviors as ones of particular interest (1948, 18). Notably, the classification is initially laid out in the abstract, without any specific reference to the kinds of behaving entities in question. This speaks to the intentionally broad scope of the paper achieved by a sweeping methodological distinction, made at the outset, between two ways in which one can study a system. The first corresponds to the behavioristic method of study which "omits the specific structure and the intrinsic organization of the object", and studies systems in relation to their environment in terms of "the output of the object and of the relations of this output to the input." (ibid). This proposed method operates in parallel with the second, "functional" method of study, which focuses on the internal workings of a system and largely brackets considerations pertaining to the relation between it and its environment.

Each method is driven by particular aims, most notably the extent to which the environment is deemed an important factor in the analysis; and each thereby reveals one aspect of a phenomenon while obscuring another. While a functional mode of analysis may reveal deep differences between two types of systems, the behavioral mode of study may highlight the ways in which their gross behaviors are similar. And it is precisely by bracketing considerations pertaining to what the system is made of, or how it is organized, that the behavioristic mode of study gains significant interdisciplinary scope. Moreover, the behavioral mode of study appears to have a certain independence from the functional mode insofar as the same behavior can be physically realized in different ways—a thesis which Roberto Cordeschi calls "machine functionalism", and which had its roots in functionalist and behaviorist psychology from decades prior (2002, 247; see also: Tamburrini and Datteri 2005; Dupuy [1994] 2009; Ashby 1962, 260–61). As

Rosenblueth and Wiener claimed in a paper written two years later, it "is possible to obtain the same output for the same input with different physical structures" (1945, 318). At the same time, they acknowledge that the functional organization of a system constrains the possible complexity of its behavior, and that in this regard insights from one of these perspectives could be enhanced from the other (Rosenblueth, Wiener, and Bigelow, 1943, 21).

It is within this broad methodological context that the authors turn to their goal of motivating the usefulness of the concepts of purposiveness and teleology. These concepts prove useful *within the behavioristic mode of study*, from which they classify particular types of behaviors. The authors begin by distinguishing active from passive behavior, which depend on whether the object is the source of the output energy (ibid, 18). Active behavior is further subdivided into purposive behavior, defined as that which is directed toward a goal, and non-purposive behavior.[9] They define the goals of purposive behavior as "*a final condition* in which the *behaving object* reaches a definite correlation in time or in space with respect to another object or event." (ibid; emphasis added). Such purposive behavior is subdivided into *teleological (purposive) behavior*, where the behavior is continuously modulated by negative feedback (such as the tropistic behaviors of organisms or servomechanisms such as torpedoes and thermostats) and *non-teleological (purposive) behavior*, where the action is too rapid to be modulated once initiated (such as when a snake strikes its prey).[10] Finally, teleological behavior can be predictive or non-predictive. The former is particularly important to the authors, as "it suggests the possibility of systematizing increasingly more complex tests of the behavior of organisms" (ibid, 22).

[9] In a follow-up paper, the authors revise this classification to make room for passive purposeful behavior, such as the movement of a compass or weighted roulette (Rosenblueth & Wiener 1950, 319).

[10] This distinction between purpose and teleology appears to be undermined by the assertion that "All purposeful behavior may be considered to require negative feedback" (Rosenblueth et al. 1943, 19). However, the equivocation here appears to pertain to the notion of feedback. While purposive behaviors are merely initiated by a signal from a goal-object, teleological behaviors are only those where the behavior is continuously modulated by negative feedback (ibid, 20).

It is worth emphasizing that the behavioristic mode of analysis is maintained throughout this classification, meaning that negative feedback (and by extension, teleology) is defined purely in terms of externally observable inputs and outputs, i.e., without any appeal to the internal structure or organization of the system (see Wimsatt 1971 for an analysis). Moreover, the authors make clear that teleological behavior is the kind which they are particularly interested in, and which they believe is most fruitful for classifying "increasingly more complex types of behavior of organisms", particularly those which involve prediction or anticipation (Rosenblueth, Wiener, and Bigelow 1943, 22). In this view, human beings are organisms that display teleological behavior with many orders of prediction (ibid).

The paper concludes with some opaque claims regarding the relationship of teleology to the concepts of causality and determinism which, by their own admission, require further elaboration. They attempt to assuage fears that teleology necessarily implies a "final cause", and therefore backward causation. In contrast, the authors claim that "purposefulness, as defined here, is *quite independent of causality, initial or final*" and therefore that "a teleological study is useful if it *avoids problems of causality* and concerns itself merely with an *investigation* of purpose." (ibid, 23; emphasis added). This is very odd. While their account of purpose does avoid backward causation, given that a purpose is defined as "a final state" and is therefore an effect rather than a cause, to say that purposiveness has nothing to do with causality *whatsoever* seems unnecessarily radical. As I see it, there are at least three ways to interpret these claims.

First, it might be interpreted as a way of contrasting linear causal thinking from the kind of circular causality which they see to be at play in teleological behavior. On one interpretation, circular causality is in fact the key to avoiding backward causation (Heylighen 2022, 9). But this interpretation, while plausible, would need to contend with the fact that their claim is explicitly about "*purposiveness, as defined here*" (ibid). Recall that teleological behavior, which involves negative feedback, is as a *special case* of purposive behavior more broadly, which does not imply any kind of causal circularity.

Second, it can be understood in the way that was by Colin Pittendrigh and other biologists decades later, namely as a claim that their view of teleology is merely descriptive, or a useful heuristic, as opposed to being truly explanatory in a causal sense. This possibility is more compelling, particularly given their claim that "since we consider purposefulness a concept necessary for the understanding of certain modes of behavior we suggest that a teleological study is *useful* if it *avoids* problems of causality and concerns itself merely with an investigation of purpose." (ibid; emphasis added). And yet, much of this hangs on what exactly they mean by "causality".

Finally, that purpose is "independent of causality" might be a suggestion that the very idea of backward causation is a kind of category mistake that conflates functional and behavioral modes of study. I deem this possibility most likely considering the concluding sentence of the paper: "causality implies a one-way, relatively irreversible *functional relationship*, whereas *teleology is concerned with behavior, not with functional relationships*." (ibid 24; emphasis added). This admittedly odd claim strongly implies what we would now call a kind of perspectivalism, namely, the idea that purpose and teleology are phenomena that properly belong to a behavioristic perspective, just as "causality" properly belongs to a functional perspective.[11] As a consequence, if the kind of causality which we appeal to in the backward causation only applies in the functional perspective, then the problem dissolves given that purpose strictly belongs to the behavioristic perspective (ibid 23). Again, this does not preclude that we can also study systems that behaving purposively from a "functional"—viz., structural and organizational—perspective, and uncover certain organizational preconditions for this kind of behavior, such as the structures (such as sensors and effectors) that

[11] The authors also claim that teleology is not opposed to determinism, because "both teleological and non-teleological systems are deterministic when the behavior considered belongs to the realm where determinism applies" (1943, 24). The meaning of this largely hangs on the notion of a "realm", which as previously suggested, may be understood as a kind of perspectivalism.

make negative feedback possible in the first place (see also: Rosenblueth and Wiener 1950, 320).

The distinction between the behavioral and functional mode of study therefore proves to be crucial for Rosenblueth et al.'s account of teleology. As I shall argue, the distinction is also essential for understanding the way in which it was supposedly compatible with mechanism, which suggests a particular interpretation of the concept.

1.1.2 Mechanism and Behavior in Cybernetics

For Rosenblueth et al., it was by adopting a behavioristic mode of study that one could most fruitfully compare and analogize machines to organisms. Indeed, while they admitted that there are many possible ways to classify behavior, their proposed typology was motivated in part by the fact that "it reveals that *a uniform behavioristic analysis is applicable to both machines and living organisms*, regardless of the complexity of the behavior." (22; emphasis added). Indeed, from this perspective the "broad classes of behavior are the *same* in machines and in living organisms" (ibid). No such uniformity could be found from the contrasting functional mode of analysis, which reveals "deep differences" between organisms and machines (ibid). The relevant organism-machine analogy was therefore to be situated at the behavioral level. But as discussed in the introduction which preceded this chapter, analogies come in at least two varieties: analogies of proportion, which pick out specific invariant features between systems; and analogies as similitude, which aim to infer unknown similarities from known similarities. The behavioral organism-machine analogies proposed by Rosenblueth et al., I would suggest, are best interpreted as analogies of the former kind, where the invariant relations have to do with the relation between the system and its environment. It is from this perspective that teleology, as behavior controlled by negative feedback, "gains precision by a sufficiently restricted connotation" (ibid, 24).

The analogy between organism and machines proposed by the cyberneticians lay at the foundation of their modeling efforts. In both publications and private correspondences, Wiener and other cyberneticians emphasized that their thesis was about the way machines are usefully interpreted as *imitations* or *models* of biological purposiveness (Aiken, Von Neumann, and Wiener 1944). In the 1943 paper, the authors spoke of teleology in terms of its *usefulness* in the context of behavioristic investigation. This attitude was in line with the broader methodological and operational spirit of cybernetics, which had its roots in the wartime origins of the major theses of the paper, namely, Wiener and Bigelow's research on anti-aircraft artillery and the behavior of airplanes (Galison 1994). Taking after Wiener's description of the struggle against wartime enemy intelligence, Peter Galison dubs cybernetics, along with game theory and operations research, one of the "Manichean sciences" (ibid, 232). These sciences put a premium on successful control and prediction. Questions of metaphysics, while not unimportant, were secondary (Kay 2000, 83–84); though that did not stop the propensity of perniciously reifying these models (Dupuy [1994] 2009, 138; Abraham 2012; Mazviiita Chirimuuta 2020; Mazviita Chirimuuta 2021).[12]

Also notable in this context was Norbert Wiener and Arturo's Rosenblueth's common interest in methodological questions in the philosophy of science, and modeling in particular (1945; Wiener 1964, 171 cited in Montagnini 2017, 83).[13] Their black-boxed approach to modeling had important methodological benefits in the way that it permitted bracketing the complexities of internal mechanisms, especially when many of these are unknown, and allowed for modeling (physically and mathematically) the behavior of organisms and machines in the

[12] It is difficult not to see Wiener doing exactly this when he states, for example, that "the modern automaton exists in the same sort of Bergsonian time as the living organism" (Wiener [1948] 2019, 62).

[13] Indeed, Rosenblueth and Wiener are perhaps the first to outline what we now call the "semantic view" of scientific theories, which sees them as collections of models: "scientific knowledge consists of a sequence of abstract models, preferably formal, occasionally material in nature." (1945, 320).

same way.[14] Such modeling practices established a particular kind of 'scientific style' with ambitions to not only unify scientific disciplines, but to provide tools for answering questions at high levels of abstraction—such as the nature of purposive behavior across scales—otherwise difficult to study empirically (Rosenblueth and Wiener 1945; Abraham 2012, 651).

It is in this light that cyberneticians suggested novel scientific hypotheses that could be tackled by their theoretical and mathematical biology. Wiener and Bigelow had noticed that machines operating by negative feedback can at times overshoot their target, which sends them overcorrecting in the opposite direction, causing increasingly large deviations from their target. The possible fruitfulness of applying the logic of machine mechanism to living beings was vindicated when Wiener approached Rosenblueth with the question: "Is there any pathological condition in which the patient, in trying to perform some voluntary act like picking up a pencil, overshoots the mark, and goes into an uncontrollable oscillation? Dr. Rosenblueth immediately answered us that there is such a well-known condition, that it is called purpose tremor, and that it is often associated with injury to the cerebellum" (Wiener [1943] 2019, 13). This was one of the key upshots of their first collaboration: "the analogy with the behavior of a machine with undamped feed-back is so vivid that we venture to suggest that the main function of the cerebellum is the control of the feed-back nervous mechanisms involved in purposeful motor activity." (Rosenblueth, Wiener, and Bigelow, 1943, 20). In other words, the analogy of proportion between organisms and servomechanisms had led to the formation of a specific hypothesis about the function of the cerebellum, that could be investigated further by neurophysiologists (ibid). According to Wiener, this key point was at the heart of a paper that was "a statement of a program for a large body of experimental

[14] In this way, the functional/behavioral distinction also resembled Craik's distinction between analytic and synthetic methods, where the former is concerned with physiological structure and the latter the construction of models to simulate organic functions (Cordeschi 2002, 144; Craik 1966).

work", which they hoped could be carried out at a new interdisciplinary institute (Wiener [1948] 2019 p.13; see also Abraham 2012).

The idea that the logic of negative feedback underlies complex behavior was a crucial insight for Wiener and his colleagues. From a "functional" perspective, the central nervous system appeared as a highly complex, intricately organized system. From the "behavioral" perspective, organisms appeared as black-box systems constantly coming to grips with and modulating their relationship to the environment through the feedback between sensation and action. As Wiener had it, this behavioristic conception of the nervous system marked "a new step in the study of that part of neurophysiology which concerns not solely the elementary processes of nerves and synapses but the performance of the nervous system *as an integrated whole*." (emphasis added, Wiener [1948] 2019, 13). Similarly, Ross Ashby, called cybernetics "a 'theory of machines', [which] treats, not things but ways of behaving. It does not ask 'what is this thing?' but 'what does it do?'" (Ashby 1957, 1; see also: Pickering 2010)

In my view, it is with the foregoing considerations in mind that the notion of 'teleological mechanisms' is to be understood. In so far as cybernetics offered a *mechanistic* account of teleology, one must understand this notion of "mechanism" in the context of the particular behavioral analogies of proportion between organisms and servomechanisms. Even though mechanism has almost always been associated with reductionism, the notion of 'mechanism' in 'teleological mechanism' was tied to the behavioristic mode of study (described in the previous section) that seeks to understand the action of a system as a unified whole in relation to its environment. That such a broadly 'holistic' perspective is deemed compatible with (machine) mechanism requires a specific conception of a machine—a servomechanism—which is so closely coupled with its environment that its behavior can only be understood with respect to it (see also: Allen 2005, 264).[15] Some, such as Ross Ashby, went even further in claiming that his

[15] I use the term 'holism' here in a broad sense, to capture anti-reductive views of life, and in this case of the cybernetic variety which emphasized the irreducibility of the system-environment coupling (see Baedke 2019, for an analysis of the challenges associated with this kind of view). This

servomechanical contraption should be understood as "the whole thing, organism *and* environment" (Ashby 1953, 94). It is this inextricability of organisms from their environment that servomechanisms captured. The same would not evidently be true of an analogy between, say, a clock and an organism. And since analogies are two-way streets, it is in this behavioral respect one can say not only that organisms are machine-like, but machines organism-like, and on this basis, the living and mechanical could legitimately be predicted and controlled using the same models.

To summarize the argument so far, I've shown the senses in which the cybernetic account of purpose was both distinctively *behavioristic* and *mechanistic*. It was behavioristic in the sense that it theorized purpose from a behavioral (as opposed to the functional) mode of study which brackets any considerations regarding the internal constitution of the system in question, and in so doing evaded the problem of backward causation (or so Rosenblueth et al. claimed). This, in turn, opened the possibility of modeling purposive behavior via the behavior with the kinds of machines that also instantiate such behavior. This made their account of teleology mechanistic in a very particular sense that did not imply a reductive, decompositional thesis that would lose sight of system-environment interactions. But what is crucial to notice here is that the possibility of such a rapprochement between organisms and machines *was contingent on defining teleology and purpose in behavioral terms in the first place*. That being said, Rosenblueth et al. were far from the first to conceive of these concepts in this way. Indeed, the influence of behaviorism on the early years of cybernetics continues to be underappreciated today. I take this to be largely a result of the fact that the historical context of these ideas has not been sufficiently appreciated. In the next section, I will aim to remedy this.

---

usage should be distinguished from more specific senses of the term used by adherents of schools such as German Holistic biology (Harrington 1996; see also: M. Esposito 2017).

1.2 Cybernetic Teleology in Context

Important clues to the thinking that inspired the cybernetic account of purpose and teleology can be found in a 1948 foreword to a special issue on teleological mechanisms published in the Annals of the New York Academy of Sciences by social scientist Lawrence K. Frank (1948). To explain the motivations behind the topic of the issue, he spells out how some members of this group saw their conception of teleology as fitting in within their intellectual milieu. According to Frank, scientists had long made progress in their rejection of teleology as a supernatural or religious concept, and instead favoring reductive mechanistic thinking (Frank 1948). Though this approach had enormous success, recent years had seen "growing perplexity in biology and psychology" due to the way that a "mechanistic and deterministic view of nature" had drawn researchers "further and further away from the organisms and the personalities they were trying to understand" (ibid, 190). This had in turn caused a search for new approaches, concepts, and methods "capable of dealing with the *larger wholes of organisms and personalities*" (ibid; emphasis added). Frank proceeds to locate the efforts of the proto-cybernetic group within this holistic line of thought: "the concept of teleological mechanisms" seeks to escape from older, inadequate mechanistic conception, and instead seeks "*a conception of 'natural teleology*'" (ibid; emphasis added). Given how heterogenous the cybernetic movement was, it should not be assumed that Frank was speaking on behalf of all its supporters. But his *Forward* poses important questions, particularly about the cybernetic movement's relationship to other anti-reductive movements of the interwar period concerned with naturalizing purpose.

1.2.1 Homeostasis and Teleology

Though cybernetics emerged out of a melting pot of research and debate in many fields of study, many scholars have emphasized the parallels and possible lines of influence between Walter Cannon's research on homeostasis and cybernetics

(Riskin 2016; Cueto 2016; Hagen 2021; Burbano and Reyes 2022). As Perlman (2000) emphasizes, Cannon saw regulation as the central problem of physiology and a fundamental principle underlying the structure of living systems. In his 1945 book, *The Way of an Investigator*, Cannon also explored the rationale for legitimizing teleological reasoning in physiological research (Cannon 1945, 108–13). The 1943 paper suggests that this perspective seemed shared by Rosenblueth, who worked with Cannon at the department of Physiology at Harvard from 1930 to 1942-43, and later by Wiener (ibidem; Wiener [1948] 2019, chap. 4; Ashby 1957, 196; Montagnini 2017, 83). This connection was also remarked by other cyberneticians, such as Warren McCulloch: "[Servomechanisms] are, as we say, error-operated. The state toward which they return the system is the goal, or aim, or end in and of the operation. This is what is meant by function. *On these circuits Cannon founded his theory of homeostasis, and Rosenblueth and Wiener their theory of teleological mechanisms*." (McCulloch 1950, 197; emphasis added). Ross Ashby also made explicit this connection when he created a servomechanism intended to model the capacity for adaptivity and learning, which he dubbed the "homeostat", after Cannon (Ashby 1960). Furthermore, like Cannon and many other physiologists of the interwar period, cyberneticians and complexity theorists extended their thinking about complex systems to the social sphere, where notions of control, breakdown, and pathology could be applied (Bates 2014; Geroulanos and Meyers 2018).

While there is no doubt that work on homeostasis influenced cybernetics, it is far less clear that it also inspired the discussions of purpose and teleology core to the 1943 paper. Equivocations notwithstanding, the cybernetic definition of teleology as negative feedback was only deemed a special case of purposive behavior more generally. Furthermore, there was an obvious sense in which the servomechanisms built by cyberneticians to model teleology completely failed to capture the essence of homeostasis, namely, the fact that it is in service of keeping the organism alive (Maturana and Varela [1972] 1980). Perhaps for this reason, there is a clear *lack* of emphasis in the cybernetic literature on self-maintenance and homeostasis as a defining feature of teleology. Indeed, such claims would have directly challenged the organism-machine analogies central to the cybernetic project (Lillie 1942). Even Ashby, who was the one who took this idea most seriously, also systematically undermined the point. At the Ninth meeting of the Macy Conferences on Cybernetics in 1952, he declared: “I shall consider the organism, then, as a mechanism which faces a hostile and difficult world and has as its fundamental task keeping itself alive”, all the while analogizing an organism’s homeostatic processes to a ship’s dials which must remain within a specific range (Ashby 1953, 73; see also: 1962; Froese and Stewart 2010). Evidently, the latter has nothing to do with the ship’s continued existence. At the same time, it is notable that Rosenblueth—who directly bridged the thought of Cannon and cybernetics—downplayed the importance of purpose and teleology, instead showing more interest in empirical problems pertaining to, in his words, “theoretical biology” (Rosenblueth 1947; cited in Abraham 2012) And in a book chapter published some years later on homeostasis, he made no reference to purpose and teleology whatsoever (Rosenblueth 1963).

How then should we locate the cybernetic account of purpose and teleology among those that preceded it? While not immediately obvious, *Behavior, Purpose, and Teleology* reflects the many significant and interrelated debates about teleology in the late 19th and early 20th centuries, and for which the journal *Philosophy of Science* was an important venue from its founding in 1934 until at least the 1950s (e.g.: Lillie 1943; Krikorian 1943; see Malaterre, Chartier, and Pulizzotto 2019). These centered on the distinction between automatic and plastic behavior, and particularly over the concepts of instinct, reflex and cognition; the possibility of purposive behavior without cognition; the continuities and discontinuities between humans and other organisms, or the extent of non-human cognition; and finally, the emergence of self-regulating machine as models of purposive behavior and learning (Richards 1987; Cordeschi 2002; Kremer 2009; Klein 2018). In other words, it is critical to notice the ties which cybernetics had to the many fields other than physiology which contributed to the study of animal behavior, including zoology, evolutionary biology, psychology, and philosophy.

In the next two sections, I will focus on specific parallels between the cybernetic account of purpose and a lesser-known strand of anti-reductive thinking in behaviorist philosophy and psychology, which I will dub *purposive behaviorism*. These scholars, which included the neo-realist philosophers Ralph Barton Perry and Edwin Holt, along with the psychologist Edward Tolman, pitted themselves against what they deemed to be reductive 'Watsonian' behaviorism, and instead sought a richer conception of behavior that would accommodate concepts like purpose.[16] The similarity of their accounts to the cybernetic account was noticed almost immediately by critics and supporters of cybernetics alike (Taylor 1950a; Churchman and Ackoff 1950). Furthermore, there is some evidence—albeit limited and circumstantial—that key cyberneticians were influenced by them. Consider, for instance, that Norbert Wiener was as much a philosopher as a mathematician by training, and came out of his graduate training at Harvard well versed in debates about mechanism and teleology (Wiener 1924a; 1924b; 1924c). These were topics that interested Pragmatist and neo-realist philosophers he knew personally and whose seminars he attended: Josiah Royce, George Santayana, Ralph Barton Perry, and Edwin Holt (Montagnini 2017, 15; Peters and Peters 2016). Wiener corresponded with Perry during his time at the University of Cambridge, and spoke highly of Royce's seminar on scientific method, reportedly giving him "some of the most valuable training I have ever had" (Wiener 2018, 130). This seminar was attended by the biologist Lawrence J. Henderson, who had recently defended an account of teleology in his book *The Fitness of the Environment*, which Royce (1914), Perry (1917b), and Holt (1920) engaged with and critique. Closer to the dawn of cybernetics, Wiener corresponded with the Harvard psychologist Edwin Boring on the replication of brain functions in electrical systems (Boring 1944); Boring had previously corresponded with Clark Hull on issues of behaviorism, teleology, and the

[16] Neo-realism was a short-lived Harvard-based movement which included Holt and Perry, and which challenged the idealism of Josiah Royce and others (Holt et al. 1910; Holt, Marvin, and Montague 1912; Kuklick 2001, chap. 11; Charles 2011). A behaviorist view of mind and purpose was a central component of their program.

mechanical simulation of purpose (Cordeschi 2002, chap. 3). Wiener expressed his grand ambitions in a letter to J. B. S. Haldane, while simultaneously distancing himself from his mentors: "Behaviorism as we all know is an established method of biological and psychological study, but I have nowhere seen an adequate attempt to analyze the intrinsic possibilities of types of behavior. […] At present, however, we are quite clear on the fact that *no behaviorist has ever really understood the possibilities of behavior*" (Wiener 1942). It is these connections which cybernetics had to conceptions of teleology that were found primarily in philosophy and psychology which provide crucial context to the cybernetic account of purpose.

1.2.2 Purposive Behaviorism

Historians have long challenged the received view of behaviorism as a monolithic movement inaugurated by John B. Watson (1913). Many of its supposedly distinctive tenets—such as its opposition to "unscientific" introspectionism in contrast to the "objective" study of mind qua behavior—were already defended by psychologists and philosophers in decades prior (Klein 2020; see also Richards 1987). Furthermore, though behaviorism is seen as an approach that sought to eliminate all mentalistic concepts from psychology, it was in fact a heterogenous movement that comprised many different perspectives on the place of such concepts and their relationship to behavior (Samelson 1981; Hatfield 2003; Costall 2006; Malone 2014). Several scholars focused particularly on finding a place for the concepts of purpose and teleology in behaviorism.

Support for a purposive conception of organismal behavior was shared by philosophers, such as Edgar Singer, Arthur Ernest Davies, Ralph Barton Perry; psychologists such as Howard Warren, Leonard Hobhouse, Edward Tolman, William McDougall; and scholars who straddled both worlds, such as Edwin Holt (among others). Similarly, articles on the topic ranged from philosophical publications (*The Monist*; *Philosophical Review*; *The Journal of Philosophy*), to psychological ones (*Psychological Review*; *American Journal of Psychology*) and hybrid journals (*Psychology and Scientific Methods*; *Philosophy of Science*). Though perhaps not a school in its own right, there is certainly a shared set of philosophical commitments among these authors, along with institutional ties and common inspirations. For example, many of the most prominent defenders of purposive behaviorism were trained at Harvard or held positions there, and almost if not all of them were heavily influenced by American Pragmatism (Kuklick 2001; Carroll 2017; Misak 2013; James [1890] 1983, xx). Indeed, the basic tenets of purposive behaviorism are core to Pragmatist philosophy, and the latter's major proponents frequently disparaged attempts at eliminating teleology (James 1878; Short 1981). Along the lines of Peirce's pragmatist maxim which urges us to consider the 'practical effects' of our beliefs, purposive behaviorism saw the mind and its associated concepts—such as intention, belief, and purpose—as being intimately tied to behavior in the world, rather than something that resides "in the head" (Perry 1909; Singer 1911; Holt, Marvin, and Montague 1912; Tolman 1928). In this view, rejecting introspectionism did not mean eliminating mental and agential concepts, but was rather a call for properly resituating them. In the words of the pragmatist Stephen Pepper: "…there is no reason why behaviorism should stand or fall with "materialism," or why all behaviorists should be committed to the metaphysical ventures of J. B. Watson. And there is very good reason why psychology should continue its efforts to the utmost to give a behavioristic account of mind." (Pepper 1923, 244). In other words, defenders of purposive behaviorism sought to expand the concept of behavior such that it could be understood to comprise mindedness rather than eliminate it (Kuklick 2001, 426; Carroll 2017, 33).[17] This view went hand in hand with a gradualist understanding of life and mind, which saw no fundamental difference between the purposive behavior of a simple organism and the intelligent behavior of humans (Cordeschi 2002, 124). Of the founding fathers of Pragmatism, John Dewey was perhaps the one who took this view the furthest, arguing that "that there is no breach of

[17] And even outside of purposive behaviorist circles, philosophers such as Albert Hofstadter would promote an approach to identifying teleological behavior "independently of internal structure of the agent in the process" (1941, 32).

continuity between operations of inquiry and biological operations and physical operations" and second, "that rational operations grow out of organic activities, without being identical with that from which they emerge" (Dewey 1939, 19).

To varying degrees, purposive behaviorists pitted themselves against alternative, reductive views of behaviorism. A prominent target was the self-styled father of behaviorism, James B. Watson, who was heavily influenced by his teacher, the physiologist Jacques Loeb (Pauly 1987). In this view, behavior ought to be understood solely through the lens of its atomistic components. The British psychologist William McDougall—who took up the William James chair in psychology at Harvard in 1920—not only emphasized the shortcomings of Watson's 'mechanistic' behaviorism, but went so but argued that mechanism and teleology were incompatible (though what exactly he meant by "mechanism" is unclear; see Boden 1972).[18] On the Watsonian, reductive mechanistic view, the brain "appear[s] as nothing but a vast automatic telephone exchange" and humans as a vastly complex "penny-in-the-slot machine […] Hence the intelligent purposive operator, whom Descartes imagined as presiding over the switch board, has been pronounced an unnecessary hypothesis" (McDougall 1923, 278). According to McDougall, such an overly simplistic view of organisms forces the mechanistic psychologist to have recourse to other concepts, such as "'motor sets,' 'trends,' 'drives,' and 'determining tendencies,' all of which are terms of the Purposive Psychology, thinly disguised" (ibid, p. 288; c.f. C. L. Hull 1943, 25-26). Ralph B. Perry was more conciliatory: Watson's view, while limited, was compatible with the purposiveness of behavior as they defined it (Perry 1918, 3).

The distinction between reductive and holistic forms of behaviorism also appeared under the monikers "molar" and "molecular" behaviorism, which

[18] McDougall's position is somewhat difficult to characterize, and some may think that he does not rightly belong among those seeking a 'middle way' between vitalism and mechanism. This is partly due to the fact that he labeled himself a dualist and animist, though he qualified these as not implying "metaphysical Dualism, or indeed any metaphysical or ontological doctrine" (McDougall 1911, x; see: Boden 1972; Lawrence and Weisz 1998, 12). I take this to perhaps point to the limits of framing the work of these authors as such a "middle way". For McDougall's relationships to his American colleagues, see Rose (2016).

originally came from the British emergentist Charlie D. Broad (Broad 1925). The former referred to behavior in the sense of gross "overt actions" of the whole organism, while the latter referred to behavior 'inside' the organism, such as change in blood pressure (ibid, 616). Broad used this distinction to argue against behaviorism as a whole, arguing that "a man starts as a Molar Behaviourist and is then pushed back by criticism into Molecular Behaviourism, at which stage his theory has lost most of its interest." (ibid, 617). In effect, Broad contended that the molar/molecular distinction was a false one in practice. But purposive behaviorists thought that there was a genuine distinction to be had here. Tolman, for instance, critiqued Watsonian "molecular" behaviorism as being a theory of "muscle twitchism" while insisting that molar behavior has certain "emergent" properties, including purposiveness. Harvard philosopher Ralph B. Perry said that purpose requires "an epistemological construction that lies beyond the scope of a strictly physiological (viz., molecular) behaviorism", without necessarily contradicting the basic theses of behaviorism (Perry 1918, 7).

Similar arguments had been made about a decade earlier by Edwin B. Holt who, alongside Perry, taught courses at Harvard which were followed by both Tolman and Wiener between 1911-1913 (L. D. Smith 1982; Montagnini 2017, 15; Carroll 2017, 21; 184). As Holt had it: "nerves and muscles, sense-organs, reflex arcs, stimulation, and muscular response...these things do not reach the core of the matter, and that they never can [...]" the biologist has "overlooked the *form of organization of these reflex arcs*" (Holt 1915b, 49–50; 76; emphasis added). Out of this larger "form of organization" emerges a novel capacity for *behavior* that makes psychology irreducible to physiology. Perry presented the methodological divergence between these fields in slogan form: "in so far as psychology divides the organism it approaches physiology, and in so far as physiology integrates the organism it approaches psychology." (Perry 1921, 85). Tolman emphasized the fact that reactions must be understood in terms of the broader purposes and functions they serve (Carroll 2017, 76). Holt, and much like Tolman after him, is careful to specify that the organized whole is not necessarily 'more' than the sum of the parts "thus organized", but can *do* something which the parts on their own cannot do. As such, "new phenomena, new laws and functions have been developed." (Holt 1915a, 365).

Holt illustrates the way behavior 'emerges' through the coordinated and organized integration of multiple reflexes with a thought experiment. We are asked to imagine a phototropic creature, much like the servomechanical robots that cyberneticians built decades later (Riskin 2016, chap. 9). The creature has a light receptor on one side of its body that stimulates the action of a fin on the opposite side. When light strikes the eyespot, it flaps its fin to rotate toward the light in a circle. With this single reflex arc, the creature is capable of motion but not *bona fide* behavior. With a second light receptor and fin on the opposite side, however, it can now swim *toward* the light. As light hits one eye, it will rotate toward it, and light will hit its opposite eye, causing its opposite fin to flap and redress the animal so that it now swims toward the light, giving it the capacity to respond to a specific feature of its environment. With the addition of a third reflex that senses heat, it would now be able to move toward a bright and hot source without burning itself. So whereas a single reflex acts like an ignited rocket that happens to fly in the direction that it was pointing, and therefore "does not refer to anything beyond itself" (Holt 1915b, 54), full-blown *behavior* has reference to an object of the environment.[19]

Purposive behaviorists such as Holt and Tolman saw purposiveness as an emergent property of the behaving organism as a whole in an environment, which cannot be explained in terms of individual reflex actions caused by stimuli impinging on sense organs. Such an explanation would fail to account for the behavior emerging from the organized integration of these reflexes (Holt 1915a, 366).[20] Put differently, while we can fully understand the rocket based on its

[19] This transition also involved the 'recession of the stimulus': a process whereby the behavior is not merely a direct product of stimulation, but of the luminous object itself (Holt 1915b, 76); or even cases in which sophisticated behavior of complex organisms is not orientated toward a clear physical object in the environment, but rather a more complex "situation" (79; c.f. Rosenblueth and Wiener 1950, 321).

[20] Though whether such an integrated view of reflexes, famously defended by Charles Sherrington, was sufficiently "holistic" was itself a matter of debate (Goldstein [1939] 1995).

internal structure and initial conditions, we cannot understand purposive behavior by looking inside the organism, because the environmental object toward which it purposively behaves is part of the description of behavior. Purpose, in this view, is always a feature of behavior directed toward objects in an environment; an idea that would later strongly influence one of Holt's students, James J. Gibson (Holt 1915b, 54–55; Shaw 2011).

In this way, purpose became a central topic of interest for scholars interested in a middle way between vitalistic and reductive-mechanistic accounts of animal behavior. In an article published a year after Holt's, Princeton psychology professor Edward Warren claimed that "Notwithstanding the vigorous controversy that has recently waged between vitalists and mechanists, and the forceful arguments put forward on behalf of teleology and against it, neither side has so far given an adequate account of the nature of purpose itself. *A thorough examination of its fundamental character might reveal a middle ground on which the two opposing parties could unite*" (Warren 1916a, 6; emphasis added). The very possibility of finding a suitable middle position was, to some extent, contingent on the kind of 'machine mechanism' which I discussed in the previous section—though this is at times hard to see due to ambiguities in the meaning of causality and mechanism.[21] Those who subscribed to Clark Hull's "robot approach" to simulating purposive behavior did so in part as a way of validating their naturalist (i.e., non-vitalistic) credentials (Cordeschi 2002, 144). In a similar vein, Warren—who defended a conception of teleology as anticipatory behavior—emphasized the analogy this has with a governor regulating the speed of a wheel; an indication that "mechanistic processes may fall into the purposive type" and that this did not require attributing "a resident entelechy to the regulator".[22] This can be seen as an effort on the part of these scholars to simultaneously insist on the existence of purposive behavior, while taking care not to overemphasize its distinctiveness,

[21] For one attempt at articulating the relation between these concepts, see: Warren (1916c).

[22] From the cybernetic point of view, Warren was conflating negative feedback with anticipatory negative feedback—or what Perry called compensatory vs preparatory adjustment.

lest it be taken to be radically discontinuous with other natural (non-vital) phenomena. Perry simultaneously recognized that "purpose is supposed to be a kind of happening or chain of events" and therefore not completely unlike a mechanism; and yet it is also somehow "differing in its determination from that of mechanism." (Perry 1917a, 352). That precise difference proved difficult to spell out clearly (see also: Hempel and Oppenheim 1948).

Holt went further and insisted on the emergent novelty of behavior while also arguing that it falls within a broader view of scientific laws which applies equally to solar systems and behaving organisms (Holt, 1915a). This can be seen in his distinction between two kinds of views of causation. According to the outdated 'bead theory', the previous state of a system "causes" the next and implies that there is an entelechy or occult power residing *within* the system in question. According to the modern 'function theory' one ought to describe the behavior of the system as a constant function of it in relation to something else (not to be confused with Rosenblueth et al.'s quasi-*opposite* use of the term 'functional').[23] Just as the falling of an object is a function of its mass and the mass of the attracting object, purposive behavior is a function of the organism and an external feature of its environment, namely, objects toward which it behaves purposefully (Holt 1915a, 367; c.f. Rosenblueth and Wiener 1950). At the same time, this view allows us to describe behavior in such a way as to extract laws. As Holt sees it, every physical law is "in the last analysis the statement of a constant function between one process or thing and some other process or thing" (ibid, 367-368). And here too one finds hints of the precarious balance that had to be struck by middle way approaches: for while Holt emphasized such continuities between physics and psychology, he simultaneously used the parallel to admonish us to keep the

[23] Holt's distinction between the function and bead theory bears some resemblance to Rosenblueth et al's distinction between the behavioral and functional modes of study respectively, though the meaning of function here is opposed. Holt's function theory is explicitly behavioristic, and in this way resembles the behavioristic mode of study. In turn, what Holt calls the "bead theory" might be seen to loosely parallel what Rosenblueth et al. call the functional, i.e., organizational perspective. The important difference is that Holt dismisses the bead theory entirely, whereas Rosenblueth et al. deem their functional mode of study important. Furthermore, Holt's contrast is more explicitly ontological, whereas Rosenblueth et al.'s is methodological.

organism whole "if it is behavior that we are studying" (ibid, 369; c.f. Loeb 1918, 17–18).

In sum, one finds in the work of Holt and other purposive behaviorists a need to simultaneously highlight the continuities and discontinuities between psychology and the 'hard sciences' of physics and engineering. It is precisely in this context that one finds what may be the earliest expression of the cybernetic view that a servomechanism like a thermostat may be teleological (Warren 1916c, 61; Perry 1917b, 488). In fact, one notices that this was already a straightforward consequence of Holt's definition of purpose, which is remarkably similar to the cybernetic account: in this view, any entity capable of behaving with respect to a specific feature of the environment is purposive. But as mentioned above, there were others such as Warren who argued that this was too simple, and that teleology required anticipation as an additional ingredient. Perry took this further, arguing that true purposiveness required the capacity for learning. In what follows, I shall argue that underlying these disagreements was a more fundamental ambiguity about the very meaning of purpose and the limits of what a behavioristic account could accommodate. I shall conclude by showing how these same difficulties are to be found in the cybernetic account of teleology.

1.2.3 Intrinsic and Extrinsic Purpose

In Perry's 1917 article, *Purpose as Tendency and Adaption*, he argued that various ways of conceiving teleology behavioristically—including Holt's, Warren's, and even the organicist Ralph S. Lillie's—were essentially the same. In all cases, purposiveness is constituted by a specific kind of adjustment which a system makes to a change in conditions. These come in three types: 1) "Compensatory adjustment": the tendency to maintain equilibrium, as typified by a thermostat; 2) "Progressive adjustment": the creation and maintenance of new equilibria, such as in growth; 3) "Preparatory adjustment": the anticipatory maintenance of an equilibrium (Warren's view; or what Rosenblueth et al. called 'predictive behavior'). Perry deemed all three cases of adjustment to be the same in so far as "the independent variable is the environmental action, the dependent variable is the organic response, while the outcome or resultant is constant. Hence we may say that the organism behaves in a variety of situations in such a way as to produce a constant result (1917b, 493).

For his part, Perry expressed some agnosticism over whether such cases of equilibration, even in their most complex form, truly amounted to purposive behavior. On the one hand, he deemed the three kinds of adjustment to be worth distinguishing as a distinctive class of processes and recognized that one may find it 'useful'—as Rosenblueth et al. later claimed—to call these purposive. However, he also argued that there were good reasons to think that bona fide purposiveness only applies to behaviors more complex than any of these, because none of the three kinds of adjustments require that we speak of the behavior in terms of 'means' and 'ends'. In all these cases, there is "a give and take between environment and organism, in which the environment makes the first move, and in which the exchange of actions obeys a constant law…The response *is* complementary; but it means nothing to say that it occurs *in order to* be complementary, or *for the sake of* the complementary outcome." (emphasis in original). For us to meaningfully speak of behavior that occurs for the sake of something, or as a means to an end, Perry argued that we must look to more complex behaviors born out of a process of *learning*. Making sense of the way an animal selects and learns to deploy a particular action in a given context requires speaking of two 'propensities': a 'higher' or 'selective' propensity, and a 'lower' propensity (Perry 1918; 1921). The former 'higher' propensity refers to that which motivates the animal's efforts, and which determines what counts as success. "In the maxim 'if at first you don't succeed, try, try again,' the higher propensity both accounts for the repeated trials and defines what shall constitute success." (Perry 1918, 3). The lower propensity the specific actions which are adopted as a means toward reaching that end. These can rightly be called 'means' precisely because they have been *selected* in virtue of their suitability to reach the goal.

On the surface, such disagreements are over how basic and omnipresent teleology is in organismal behavior, i.e., whether teleology rightly applies only to the level of learning (Perry) or to more basic tropistic and equilibration processes (Holt). Tolman and McDougall defended positions in between these extremes (Cordeschi 2002, 136). But a deeper ambiguity underlaid these disagreements, namely, whether purpose can be defined without having recourse to claims about the nature of, or features intrinsic to, the behaving entity (c.f. Taylor 1950b, 327-329). As discussed in the previous section, for Holt the very scientific status of purpose—and the possibility of extracting laws of behavior—hung on it *not* appealing to causal powers internal to the system in question, as such claims would imply the outdated 'bead theory' of causation. This parallels Rosenblueth and Wiener's insistence that purpose "be recognizable from the nature of the act, not from the study of or from any speculation on the structure and nature of the acting object." (1950, 323). According to these views, goal-directedness is a process of behavior, understood in terms of the "functional relationship" (in the Holtian sense) between the behaving object and a feature of the environment. These goals were often characterized as physical objects in the environment, such that one can speak of 'signals' emanating "from the goal" (Rosenblueth, Wiener, and Bigelow 1943, 19). Indeed, on the more conservative extremes of purposive behaviorism defended by Holt and the cyberneticians, a goal can only ever be objectively individuated in terms of a feature of the environment external to the behaving system.

Whether purposive behavior must be accounted for in terms of facts intrinsic to the organism also emerged in the context of a disagreement over whether purpose was truly being *identified* in terms of the behavior, or if it was merely *inferred* from the behavior. Highly influenced by Holt and Perry, Tolman began in the former camp, and he critiqued McDougall for explicitly holding the latter (Tolman 1925, 37; see also: Cordeschi 2002, 135–36). But identifying purpose with behavior was easier said than done: Tolman was soon accused by his student Zing-Yang Kuo for implicitly inferring purpose from behavior (Kuo 1921; 1922;

1924; 1928; quoted in Carroll 2017, 77–78).[24] Tolman quickly conceded, while turning around to defend the inferentialist position on the grounds that it was still objective: "our concept of purpose, or drive, is not, even though inferred, therefore mentalistic or introspectionistic. We do not describe or define these purposes in terms of their introspected 'feels'" (Tolman 1928, 524). Tolman, in effect, aimed for a somewhat precarious compromise: one that inferred the source of purposiveness within the organism, while maintaining that it is no more than an objective feature of behavior.

The need for such a middle ground was in part driven by the need for accounting for goals that were to be clearly located in the environment—an issue was raised decades later by Richard Taylor in his critique of cybernetics (1950). Holt had attempted to circumvent this by stating that more complex behaviors occurred with respect to "a larger and more comprehensive situation". Perry, in turn, later made the case that the complementarity between behavior and object could be hypothetical rather than actual. In "the case of hunting for a pin, the organism is not, strictly speaking, responding to an object or fact of its environment. The organism is not hunting for any particular pin; and is quite capable of carrying on the hunt, even though there be as a matter of fact no pin in its environment (see also Scheffler 1959). The finding of any particular pin is the hypothetical complement to its present response. It is related to it as a hypothetical key is related to some lock which it would fit if it did exist." (Perry 1918, p. 7). Here too, we are forced to admit that the goal—since it is no longer an objective feature of the environment—is inferred from the behavior to the internal 'propensity'; a view which was later excoriated by Clark Hull (1943, 26). The problem is that it was never quite clear what this 'propensity' really is. It all too often appears as a stand-in for many more loaded concepts—many of which Perry used to describe the propensity—such as trying, agency, impulse, and even at times seems to resemble what Tolman later called a 'cognitive map' which is

[24] A similar critique was later made of Perry by Albert Hofstadter (1941).

widely considered to have been a prequel to the contemporary notion of a 'mental representation'.

Such debates about the nature of goals—whether they are internal to the purposive system in question; merely to the behavior itself; or to a 'goal-object' in the environment—closely paralleled another set of concerns over whether servomechanisms are intrinsically purposive or whether their goal-directedness is merely a projection of the observer. Since at least the time of Aristotle, philosophers have recognized a fundamental difference between the purposiveness *intrinsic* to a human mind, and the purpose—in the sense of the 'function'—of an artifact which is *extrinsically* bestowed onto it by its maker or user. The major problem here is making sense of what makes purpose *intrinsic* in the first place. It is no coincidence that the rise of servomechanical robots, combined with the redefinition of purpose in behavioristic terms that saw goals as empirically observable objects in the world, radically reconfigured the scope of the organism-artifact analogy to now comprise intrinsic purpose. Unsurprisingly, critics took this to be precisely a sign of the shortcomings of the view. As the philosopher Hans Jonas famously put it, such a conception of teleology confused 'serving a purpose' with 'carrying out' a purpose (Jonas, 2001, p. 303). Though a torpedo or thermostat's "goal" is bestowed onto it by its user or maker, and not intrinsic to it, it is not clear on what grounds a behaviorist and cybernetic account of teleology could make such distinctions—since purpose 'just is' any behavior that reaches "a definite correlation in time or in space with respect to another object or event" (Rosenblueth et al., 1943, 18).

Many supporters of purposive behaviorism recognized the significance of this issue. Warren (1916) emphasized that self-regulating systems are purposive only insofar as that purpose is bestowed onto them (see also: Hobhouse 1913). While Perry did not explicitly make the same claim, his concern that 'means-ends' talk does not apply to servomechanisms seems to drive towards the same idea. He quoted an article by William James on the matter, in which James argues that the mind "not only *serves* a final purpose, but *brings* a final purpose—posits, declares it" (emphasis in original). As Perry sees it "No one would now be disposed to dispute the essential soundness of this position. The human individual does not merely do things that are useful as judged by an external observer, but by its own activity adopts and seeks that result in relation to which its deeds are useful." Perry had addressed this topic in an earlier article, where he rejected a definition of teleology in terms of "systematic unity" (i.e., appearance of order and design) that we find in both artifacts and organisms, because teleology is about the existence of entities that "not only exhibit unity, but must be 'for the sake of' that unity". In other words, the mere fact of being a systematically unified whole, as artifacts are, is insufficient to define teleology because it does not tell us the purpose or function that such a unity serves. To elucidate that, one must account for the designer of the artifact whose intrinsic purposes are that for the sake of which the artifact exists or functions. The essential question for Perry is whether making sense of intrinsic purposiveness requires a dualistic sundering of mind and body. His answer was negative: "The better the organism is understood, the more does it assume just those characters which James insists upon as the prerogatives of mind." (Perry 1921, p.102).

But Perry also recognized the difficulty of accounting for intrinsic purpose (and mindedness more generally) in behavioristic terms, because it required going beyond a conception of behavior, defended by the likes of Holt, where one defines purpose purely in terms of external events and stimuli.[25] Perry's alternative was to account for the fact that purposive behavior is "spontaneous" and "internally conditioned". Though he hoped that such a view would "not … contradict the fundamental thesis of behaviorism" it did "forbid any hasty or contemptuous dismissal of the traditional association of purpose with nonphysical or 'ideal' entities." (Perry 1918, 7). In other words, while Holt defined purpose as a feature external to the animal, Perry emphasizes the need to refer to something *internal* to it, namely a propensity, or prior impulse of curiosity and exploration, which

[25] Defining purpose in terms of an environmental feature was particularly important for Holt as it allowed him to extract laws of behavior. He likely would have deemed appeals to features intrinsic to the behaving as a regression to the 'bead theory' (see 3.1).

allowed the organism to learn its behavior in the first place, and continues to govern this behavior (1918, 13). For Perry, distinguishing automatic action (in the sense of behavior as tendency and adaption) from intelligent action (in which one can speak of means and goals) requires "over and above the external action of the stimulus, [...] some *additional factor* [...] which decisively determines the direction which the discharge [viz., action] takes" (1917a 359; emphasis added). This "third and differential factor [...] constitutes their purposive aspect." (Perry 1917a, 360). It is this 'third factor' which Perry appeared to be describing when he talked of the 'general propensity' that drives an animal's behavior in scenarios of trial and error.

The importance that Perry saw in learning for demarcating intrinsic and extrinsic purpose did not carry over to cybernetics. For their part Rosenblueth, Wiener, and Bigelow—who defined purposiveness in a broadly Holtian way as "a final condition in which the behaving object reaches a definite correlation in time or in space with respect to another object or event"—attempted to re-inscribe the distinction between intrinsic and extrinsic purposiveness within such a definition, to strange effect. As Rosenblueth, Wiener, and Bigelow (1943) state: "the term servomechanisms has been coined precisely to designate machines with intrinsic purposeful behavior", as opposed to systems that perform in an orderly but non-purposeful way, such as a clock, since "there is no specific final condition toward which the movement of the clock strives" (19; emphasis added). But such conceptions of 'intrinsic' purposiveness seem to stray very far from their original meaning, given that it is no longer clear what the contrast with 'extrinsic' teleology would even amount to. Consider, for instance, a roulette machine which clearly plays a functional role (an extrinsic purpose) for the purposes of gambling. Such devices, on their view, are "designed precisely for purposelessness" (Rosenblueth, Wiener, and Bigelow 1943, 19). Taylor (1950a) was quick to show the odd consequences of this view: if one were to add a weight to a roulette, or if a clock happened to break down at midnight on the 1st of January, it would suddenly become intrinsically purposive. Furthermore, the critiques which Perry leveled at

Holt's account could also be extended to the cybernetic account, most notably the fact that there is no room for a distinction between 'means' and 'ends'.

In sum, the various attempts to naturalize purpose from a behavioristic standpoint in the decades prior to cybernetics struggled mightily to make sense of the distinction between intrinsic and extrinsic purpose. It seemed that more needed to be said about the nature of the organism itself to demarcate its *intrinsic* purposiveness from the *extrinsic* purposiveness of artifacts.

These shortcomings can be traced to Rosenblueth and Wiener's particular sensitivity to the need for purpose to be *useful* for science in practice. In this sense, their understanding of naturalization was closely tied to the scientific operationalization of teleology rather than any kind deeper metaphysical claim about its nature. Their view rendered moot the question of whether servomechanisms 'really are' intrinsically purposive, or whether this was a projection of the observer. Based on their distinction between functional and behavior modes of study, they saw it as essential that purpose "be recognizable from the nature of the act, not from the study of or from any speculation on the structure and nature of the acting object" (Rosenblueth & Wiener 1950, 323). As I argued in the first section of this chapter, this was also what was required both for purpose to be applicable to non-living machines, and to be rightly deemed by Frank to be in line with holism. But methodological holism of this variety was not enough, in so far as it remained omitted to keeping at arm's length the deeper metaphysical questions that inevitably reappeared at every corner. As Rosenblueth and Wiener saw it, one's operational commitments did not necessarily commit one to a metaphysical position (Rosenblueth and Wiener 1945, 316; Wiener 1924c, 528). But their constant attempt at seeking refuge in the heuristic "as if" stance put them at constant risk of perniciously reifying their own models and abstractions (Abraham 2012; Dupuy [1994] 2009, 138; Mazviiita Chirimuuta 2020). Despite being urged by their colleagues to "rework and expand" the major points of the 1943 article (Morison 1947; Wiener 1949), by the end of the 1940's the phrase 'teleological mechanisms' was replaced by a new term, ‘cybernetics’, and the movement swiftly moved on from the topic of teleology, leaving the purposive behaviorist project behind and incomplete.

1.3 Conclusion

This chapter attempted to reevaluate the cybernetic account of teleology and trace the many continuities between it and the purposive behaviorist accounts which preceded it. I put emphasis on three features in particular. First, both accounts take purposiveness to be a kind of behavior, constituted by a particular kind of relation between a system and its environment. As I shall emphasize in the forthcoming chapters, it is important to see that this perspective on teleology *qua* behavior had broad implications for issues in philosophy of science, but most of all how we conceive of the connection between causation and teleological explanation as it relates to the problem of backward causation. Second, these scholars also shared a commitment to the compatibility of teleology and mechanism, based on an analogy of proportion between the behaviors of servomechanisms and organisms—an analogy that, as I argued, lent itself to a form of holistic mechanism. Finally, purposive behaviorists and cyberneticians all struggled to make sense of the distinction between intrinsic and extrinsic purposiveness. This historical and philosophical context is crucial for understanding the way that the cybernetic account of teleology was incorporated into evolutionary and molecular biology during the second half of the 20th century. As shall be discussed in the following chapter, this transfer was marked by a new term, teleonomy, which its proponents used to highlight the differences they saw between their novel, cybernetically inspired conception of teleology that was compatible with modern biology, and the outdated, “Aristotelian” conception.

CHAPTER TWO

From Cybernetics to Teleonomy

The term 'teleonomy' was coined by behavioral biologist Colin Pittendrigh in a chapter of the edited volume *Behavior and Evolution*, which emerged as the outcome of two symposia organized by Anne Roe and George Gaylord Simpson (1958). The term denoted what Pittendrigh saw as a novel way of conceiving teleology, one which had emerged thanks to the "the engineering physicist" who "in building end-seeking automata has *sanctified* the use of teleological jargon" (Pittendrigh 1958, 394). Teleonomy was quickly taken up by other prominent biologists, most notably evolutionary biologist Ernst Mayr. In a letter to Mayr, Pittendrigh claimed that the 'engineer physicists' he was referring to were, as one might suspect, Norbert Wiener and Julian Bigelow, and that his coining of teleonomy was directly inspired by the 1943 paper by them and Rosenblueth (Mayr 1974, 115). Following Pittendrigh, Mayr took up the term and associated it with another new concept, the *genetic program*, and cemented it as the now better-known meaning of teleonomy. In doing so, he credited cybernetics even more explicitly for its genesis: "What finally produced a breakthrough in our thinking about teleology was the introduction of new concepts from the fields of cybernetics" (ibid, 92). For this reason, the term teleonomy later became synonymous with the cybernetic account of teleology, with biologists such as Francisco Ayala using the term "teleonomic systems" to refer to systems that self-regulate via negative feedback (Ayala 1970, 9).

The credit which was given to cybernetics for advancing the teleology debate is striking. As discussed in the previous chapter, nearly all of this came out of the short and somewhat strange 1943 paper by Rosenblueth and colleagues. Moreover, this praise was *uniquely* bestowed unto cybernetics despite the efforts of many other 20th century philosophers and biologists to provide a naturalistic basis for teleology, many of which attempted to carve out a middle ground between mechanism and vitalism (Nicholson and Gawne 2015; Peterson 2016; Donohue and Wolfe 2023; Baedke and Fábregas-Tejeda 2023). In some cases, this might be chalked up to the prominence that cybernetics had in popular and scientific discourse, and particularly the broader infiltration of cybernetic concepts throughout biology, particularly molecular biology.

Despite the homage which these scholars paid to cybernetics, it remains unclear what this intellectual heritage amounted to. By their own admission, they were excited by the prospects of a rigorously 'mechanistic'—and therefore 'scientific'—account of purposiveness which made no reference to vital forces and 'entelechies', and did not imply backward causation. But as discussed in the previous chapter, 'mechanism' is a polysemous term that has a complex relationship to the notion of 'proper science' (Allen 2005). Moreover, the very meaning of 'teleonomy' shifted over time and varied among its proponents. In response to these terminological ambiguities, Ernst Mayr, David Hull and others broadly distinguished two main senses of teleonomy: Pittendrigh's sense of teleonomy as a historical concept, referring to the appearance of design because of natural selection; and 2) Mayr's sense of teleonomy as a designation for processes initiated and controlled by a genetic program.

My aim in this chapter is to give an overview of these two senses of teleonomy, while also attempting to clarify their relationship to cybernetics and purposive behaviorism. To do so, I will use the three main characteristics of the cybernetic account of teleology discussed in the previous chapter. The first of these comprises two points: it is, on the one hand, a fact about the *behavioristic* perspective that the cyberneticians took on purpose and teleology; but it is also about the way that this seemed to resolve the problem of backward causation. The second feature lies in the claim that purposiveness can be modeled by machines. Finally, I argued that the cybernetic account was unable to account for the distinction between intrinsic and extrinsic purpose.

The next section will focus on Pittendrigh's original usage of the term. In section 2.1, I will first describe his view that teleonomy is synonymous with adaptation and does not imply that goals themselves play a causal-explanatory role. In this view, teleonomy is a descriptive thesis that does not imply an explanatory thesis that cites goals as causes. Section 2.2 will then seek to illuminate the connection with cybernetics based on the three distinctive characteristics mentioned above. Section 2.3 will turn to Mayr's definition of teleonomy, defined as those processes initiated and regulated by the genetic program. Section 2.4 highlights the ways in which Mayr's definition of teleonomy also has similar, albeit different continues and discontinuities with cybernetics. In this regard, I will argue that Mayr's conception of teleonomy can be seen as a legitimate, if failed attempt to overcome some of the major shortcomings with the cybernetic account of teleology.

2.1 Pittendrigh and the Origins of Teleonomy

The chapter in which Pittendrigh coins teleonomy is primarily on the topic of behavior and adaptation, which are concepts that he feels are excessively vague in their usage (1958, 390). After listing four possible meanings of adaptation, he argues that all of them "connote that aura of design, purpose or end-directedness which has, since Aristotle, seemed to characterize the living thing, to set it sharply aside from the nonliving. It is the connotation of adaptation—not the multiplicity of its denotations—that has been its greatest burden in the history of biology." (ibid, 392). Despite the progress made by Darwin, who explained adaptation without teleology "and thus conformable with the dominant scientific conceptual scheme that of physics", this association of teleology with design has lingered (ibid, 393; c.f. Lennox 1993). According to Pittendrigh, three developments since the Darwin's time conspired to make teleological description of adaptation safe for biology: 1) the recognition that evolution is not a perfecting process; 2) a better understanding of natural selection; and 3) the fact that "the engineering physicist in building end-seeking automata has *sanctified* the use of teleological jargon" (ibid, 394; emphasis added). Given the disrepute of the term 'teleology', Pittendrigh suggests the use of the term 'teleonomy' or 'teleonomic' "in order to emphasize that the recognition and description of end directedness does not carry a commitment to Aristotelian teleology as an efficient causal principle." (ibid; emphasis added).

According to Pittendrigh, the appearance of purpose or teleonomy in biological adaptation is due to the close relationship "if not an identity" between adaptation and organization. As he states, "to say that living things are *organized* is to say they are *adapted*", and vice versa; "the organism is not just a system some features of which may or may not be adaptive; the living system is all adaptation *insofar as it is organized*." (1958, 395; emphasis added). As such, Pittendrigh details what he takes to be two distinctive features of organization that are, by extension, also true of adaptation: 1) all organization exists relative to a goal and 2) all organization, in virtue of being a highly improbable state, requires an explanation in terms of "information" (ibid, 395). I will discuss these in turn.

First, the connection of organization to goal-directedness appears to be largely assumed rather than explained. In both the chapter and a letter to Ernst Mayr, Pittendrigh is committed to the idea that organization 'just is' goal-directed; an opinion which he claims to share with the polymathic cybernetician John von Neumann: "the objects of biological analysis are organizations [...] and, *as such*, are end-directed." (Mayr 1974, 115). Importantly, Pittendrigh says nothing about makes *biological* organization distinctively goal-directed; he claims that an army or a room is equally goal-directed, because neither of which can be organized with respect to nothing (ibid, 394). Indeed, the only thing which typifies biological organization appears to be the kinds of goals it is orientated towards, namely, survival and reproduction. The problem, however, is that he never makes explicit what it is about organization, broadly construed, which makes it fundamentally end-directed. Moreover, Pittendrigh insists that the inherent goal-directedness of adaptation *qua* organization is *teleonomic* rather than *teleological*. What distinguishes these two concepts? As he sees it, while teleology is an "Aristotelian" concept that commits one to the idea that goals are "materially efficient"—a rather unfortunate turn of phrase that betrays a lack of understanding of Aristotle on his part—while teleonomy is purely descriptive. But what Pittendrigh really means by 'description' remains highly ambiguous throughout the piece. While it is tempting to see it as a commitment to teleological language as a mere heuristic or projection of the observer, it seems that this is not what he had in mind. Rather, for teleonomy to be descriptive just means that it is not explanatory in a particular sense, namely, that goals are cited as *causes* of their means. The same is true of adaptation, which despite being a brute fact about living organization, does not come about *because* it will be adaptive for the organism.

The second distinctive feature of organization and adaptation is the explanation it requires in terms of "information", or more specifically, "the *origin* of that information which underlies and causes the organization" (Pittendrigh 1958, 396). This point is essentially a modern restatement of Immanuel Kant's observation, made a century and a half earlier, that it is nigh-impossible for biological organization to have emerged spontaneously. With the theoretical tools of thermodynamics, information theory, and the modern theory of evolution, Pittendrigh reformulates the question thus: "How has the information content of the genotype accumulated in face of the universal tendency to maximize entropy?" (ibid, 397) As he sees it, the general answer to this question can only be natural selection, which requires differential reproduction, and hence perpetuation of, "the information—the inherited message of the genotype—that specifies itself." (ibid). In turn, this sheds light on the fact that adaptation, as "the core of biological study", requires a historical perspective (ibid, 395; 396-397).

The connection between these two distinctive features of organization and adaptation—its goal-directedness and historicity—while never clear, provides some tentative suggestions for how we might interpret the notion of teleonomy as distinct from teleology. Pittendrigh, along with his colleague George C. Williams

([1966] 2018, 258), claims that while a historical perspective is required to *explain* adaptation, the mere recognition of adaptation as such does not require such a perspective: "The word 'adaptation' here denotes some feature of the organism—morphological, physiological, or behavioral—which serves a proximate end (food getting, escape, etc.) that the observer believes he can discern fully by direct observation and without reference to the history of the organism." (Pittendrigh 1958, p. 392). In other words, Pittendrigh is committed to the fact that one can empirically and directly *recognize* the goal-directedness of (organismal) organization without necessarily knowing its origin or cause. This in turn goes some way in explaining Pittendrigh's insistence that teleonomy is purely descriptive: one recognizes teleonomy directly when observing the behavior of organized systems; but one can then explain it by investigating the 'origin' of this organization.

Pittendrigh's chapter is full of ambiguities and unresolved questions, many of which foreshadow developments in evolutionary biology which were to follow (Dresow and Love 2023). The root cause of these ambiguities, I would suggest, lies in Pittendrigh's attempt to grapple with the enormously difficult question of the distinctiveness of biology as a science. Unlike physics, it deals with organized entities that are goal-directed and originate from a historical process of evolution. Indeed, these two features are precisely what makes biology, as a science that studies organized entities, irreducible to a "Newtonian" causal analysis (Pittendrigh 1958, 396). At the same time, Pittendrigh maintains that the explanations used by biologists *are* causal in the broader sense of 'naturalistic', given that they do not make use of vitals forces or goals as explanations (ibid). In this regard, the motivations which drive him to distinguish teleology from teleonomy are the very same which drove the purposive behaviorists in the context of psychology—a point which I will return to in the chapters that follow.

In the next section, I will focus on one particularly puzzling question which has received little attention: what is 'cybernetic' about Pittendrigh's teleonomy? From his correspondence with Mayr, we know that his "thinking about the whole thing [teleology] was very significantly affected by a paper which was published by [Rosenblueth], Wiener, and Bigelow" (Mayr 1974, 115). And while Pittendrigh used the notions of information and associated concepts, such as negentropy, it remains unclear how teleonomy relates to the cybernetic definition of purpose, defined as "a final condition in which the behaving object reaches a definite correlation in time or in space with respect to another object or event", or the narrower definition of teleology as negative feedback. In fact, Pittendrigh's conception is much broader than that of the cyberneticians, since he applies it to all organized entities irrespective of the ways in which they might behave. As such, it is not immediately clear what similarities there are between the cybernetic account and teleonomy. As I shall argue in the next section there are in fact some key continuities to be found here on closer inspection.

2.2 What Makes Pittendrigh's Teleonomy Cybernetic?

To evaluate in what sense Pittendrigh's teleonomy was cybernetic, I shall use the threefold characterization of cybernetic teleology which I introduced in the previous chapter. These are: 1) the commitment to purpose and teleology as *behavior* apprehended from a behavioristic rather than functional perspective; 2) the possibility of reproducing and making intelligible such behavior in a machine model; 3) significant ambiguities regarding the distinction between intrinsic and extrinsic purpose. I shall look at each of these in turn.

2.2.1 The Purposive Behaviorist Connection: From Behavior to Description

Pittendrigh never mentioned Rosenblueth et al.'s distinction between behavioral and functional modes of study; nor did he subscribe to the view that purposiveness is a behavioral phenomenon. However, the concerns which motivated the behavioral/functional distinction were the same as those which led Pittendrigh to distinguish descriptions and explanations in terms of goals. As I will argue in this subsection, the parallels between these two distinctions reveals a deeper convergence between Pittendrigh's work and that of purposive behaviorism.

In the previous chapter, I argued that the cybernetician's motivation for making the behavior/function distinction were two-fold. First, it allowed for a particular and specific analogy of proportion between the behaviors of organisms, i.e., one that compares the relations of organisms and servomechanisms to their environments. Second, for both the purposive behaviorists and the cyberneticians, thinking of teleology as a behavior provided a way to resolve the problem of backward causation. As I've argued, Rosenblueth et al.'s claims that "purposefulness, as defined here, is quite independent of causality, initial or final" (Rosenblueth, Wiener, and Bigelow, 1943, 23) should be read alongside the claim that "causality implies a one-way, relatively irreversible functional relationship, whereas teleology is concerned with behavior, not with functional relationships." (Rosenblueth, Wiener, and Bigelow 1943, 24). Pittendrigh's attempt to distinguish teleonomy as a descriptive thesis from teleology as an explanatory thesis was also intended to resolve the problem of backward causation. For this reason, it is not hard to see how Pittendrigh might have interpreted Rosenblueth et al.'s behavioral/functional as being equivalent to a distinction between description and explanation. The following passage is especially apt for such an interpretation: "Since we consider purposefulness a concept necessary for the understanding of certain modes of behavior we suggest that a teleological study is useful *if it avoids problems of causality* and concerns itself merely with an investigation of purpose." (ibid, 23; emphasis added). In other words, if one is committed to causal explanations appealing to the structural and organizational features revealed by a "functional" mode of study, then talk of purposiveness at the level of behavior could well be considered merely descriptive, or at best telling us about how to 'predict' and 'control' behavior without necessarily *explaining* it by appealing to the causal relations that show up in a "functional" mode of study. For Pittendrigh, such "functional" explanations would presumably come in the form of explanations that uncover "the *origin* of that information which underlies and causes the organization".

One can more fully appreciate this proximity of the behavioral / functional and the descriptive / explanatory distinctions by way of what is a rather astonishing historical coincidence. In the 1930s and early 1940s, the distinction between explanation and description was used in the context of teleology by purposive behaviorists such as philosopher Albert Hofstadter and social psychologist Floyd Allport (not to be confused with his younger brother Gordon, who was also a personality psychologist). Moreover, Floyd Allport coined teleonomy *precisely* for this purpose in 1937, twenty-one years before Pittendrigh. Though there is no evidence that Pittendrigh knew of Allport's account of teleonomy, it provides a clear conceptual—though not necessarily historical—link bridging Pittendrigh to the very same purposive behaviorist tradition which cybernetics inherited.

Like the other, older purposive behaviorists, and later Pittendrigh, Allport coined the term 'teleonomy' to talk of purposiveness without implying that goals have a causal role. To understand his thinking on the matter, one must situate his thinking in the context of his research on personality traits. His framing of the issue is very similar to those of the purposive behaviorists, especially Edwin's Holt's. Like Holt, he emphasizes how the psychological study of individuals is continuous with the goals of all branches of science: "the discovery of some regularity or consistency which gives us a *law* through which we can, in a measure, *predict* future events, in this case future behavior." (F. H. Allport 1937, 202). But for Allport and many later personality psychologists, traits are problematic insofar as they seem to imply a kind of efficient cause 'inside' the person in question. On this common-sense view, a person acts kindly because his trait of 'kindness' causes him to do so. But "[w]e really have no justification for projecting into an individual some organic disposition known as a trait. A trait, so far as we know, is only a convenient way of expressing a prediction about the individual's future instances of behavior." (ibid, 203). Instead of speaking of traits, it would be more accurate to simply speak of behavior trends. "We can describe the individual's trend of behavior *teleonomically*, that is, in terms of the purpose or purposes which he seems to be trying to carry out" (ibid, 204; emphasis added).

The influence of purposive behaviorism on Allport's thinking is clear. Allport and his brother Gordon received their bachelor's degrees and later their PhDs from Harvard during the same period as Wiener and Tolman, and were influenced by the very same professors: Hugo Münsterberg, Edwin Holt, and Ralph B. Perry (Parkovnick 2000). Both were critical of Watson's reductive behaviorism (G. W. Allport 1921; see also: High and Woodward 1980). While Floyd Allport saw behavior as purposive, he also took teleonomy to be an objective study of behavior that has no recourse to explanations in terms of conscious motives or purposes as forces "behind [the] activities" (F. H. Allport 1937, 205). Instead, Allport insisted that teleonomy is a purely "descriptive" thesis, though what he means by description is somewhat more complicated than it may seem.

Consider the parallels between Floyd Allport's argument and Holt's wholesale attack on the "bead theory" of causation, discussed in the previous chapter. As Holt sees it, to speak of intentions in the mind of a person causing their behavior reflects a broader metaphysical mistake regarding the meaning of causation, whereby past states cause future states. The correct alternative to this is the "function theory"—*not* to be confused with Rosenblueth et al.'s "functional mode of study"—which describes the behavior of an entity as a constant function of another. Just as the motion of two planets is a constant function of their respective masses and the distance between them, behavior may be described as a function of a behavior with respect to an "object, situation, or process" (369). From this perspective, the only pertinent, scientific question is "What is this organism doing? All agree that empirical study will elicit the answer to this question, and in the end the complete answer." (368). Such an analysis would "invariably reveal" a law or laws that describe such behavior.

Allport likely had such a "function theory" in mind when he describes teleonomy as a purely descriptive thesis. Indeed, he takes teleonomic descriptions to be the first step in coming up with a "hypothesis concerning a possible law of this individual's general behavior … which can be subjected to the test of verification through wider sampling and measurement." (207). Such discovery of laws "is precisely what we mean by the scientific study of personality" (204), and in this way obviates any need for explanations in terms of causes. But rather than making teleonomy non-explanatory, it simply made it explanatory in a different way. When Allport speaks of teleonomy as a descriptive thesis, he does not have in mind a contrast with a causal-mechanical account that would be veritably explanatory, and to which one could presumably reduce teleonomic descriptions. In contrast, teleonomic descriptions provide the essential data from which complete explanations of behavior may be derived in terms of *laws*, which would have the same explanatory status as gravitational laws. Laws of teleonomic behavior are therefore a species of functional theories of explanation more generally, which legitimates the new suffix 'nomic', meaning law. In sum, Allport's teleonomy is in many ways deeply aligned with the Holtian, purposive behaviorist conception of purposiveness.

Pittendrigh's teleonomy shares much with Allport's and the cybernetician's account of purpose. Broadly speaking, one sees a common recognition of purposiveness as a defining feature of the life science's objects of study, and a desire to simultaneously acknowledge the reality of this fact without compromising the naturalistic standards of a properly scientific discipline. The solution to this problem is also similar in these cases, which we see in Pittendrigh's insistence that teleonomy can be observed and recognized directly from the observation of organized systems (N. S. Thompson 1987).[26] And like Allport, he insists on a contrast between so-called Aristotelian teleology, which he understands to mean that goals are efficient causes, and teleonomy as a purely descriptive thesis. But there is also one key distinction between their views: there

---

[26] As Thompson (1987) argues, that idea was pursued even further by Albert Hofstadter (1941). For him, it was important that one distinguishes the "identification of teleological process independently of internal structure of the agent in the process. We take the agent as a unit and consider the macroscopic structure of its action. Under what conditions may that action be said to be teleological?" (ibid, 32).

is no reason to think that Pittendrigh's notion of 'description' was tied to a broader theory of scientific explanation in terms of laws, as it was for Holt and Allport, or of Rosenblueth et al.'s behavioral and functional perspectives. In this way, we find in Pittendrigh's description/explanation distinction what is in effect an impoverished version of both Allport's conceptualization of the distinction and the distinction between behavioral and functional modes of study. At the same time, there remains some continuity in the ways that all these scholars, Pittendrigh included, used these distinctions in large part (though not solely) to find a solution to the problem of backward causation.

2.2.2 Teleonomy and Mechanism

The second distinctive feature of the cybernetic account is its commitment to the analogy between the purposive behavior of organisms and that of servomechanisms. In this respect, Pittendrigh's teleonomy significantly parts ways with the cybernetic account. Elaborating the details of this divergence in turn reveals certain ways of thinking about teleology and mechanism that were distinctive of evolutionary biology.

Recall that the cybernetic analogy between organisms and servomechanisms specifically regards their behavior. And in both cases, this behavior is one that is not merely "organized with respect to" a goal in a general, but to a specific feature of the environment which counts as the systems 'goal state'. In this regard, the behavioral/functional distinction plays an important role not only in identifying a particular view of what purpose and teleology are (i.e., a behavior), but also in establishing a particular analogy of proportion between the behaviors of organisms and machines. As such, Pittendrigh's abandonment of the specifically behaviorist conception of purpose also undermines this machine-analogy, which is replaced by a generic analogy between organisms and artifacts founded on the fact of being organized. This was implicitly recognized by the biologist George C. Williams who, favorably commenting on Pittendrigh, claims that one can find "no better illustration of this process of recognizing functional design from careful observation than that provided by Paley" (Williams [1966] 2018, 258–59). But as is well known, Paley's association of organization in nature (or apparent design) with purpose is founded on a generic organism-artifact analogy which locates the purpose of an organization in the intention of an intelligent designer, or in the modern biological view, a 'blind' designer of natural selection.[27] It is this latter view that, I submit, is implicitly operating in the background of Pittendrigh's understanding of "end-directed mechanism".

Granted, one may also argue that Pittendrigh's usage of the term 'mechanism' is vague to the point of being almost meaningless; and indeed, such philosophical matters are not of particular interest to him. Nonetheless, it is worth noticing how the differences in the ways that the cyberneticians understood the meaning of 'mechanism' goes hand in hand with a particular understanding of teleology and the analogies it made possible; a conceptual linkage that carries over to Pittendrigh's more general conception of teleology as design. These differences show that insofar as Pittendrigh's teleonomy was mechanistic, it was in an altogether more general and vague sense.

2.2.3 Teleonomy, Adaptation, and the Intrinsic/Extrinsic Distinction

The third distinctive feature of cybernetic teleology which I identified is their confusion in the way its defenders framed the distinction between intrinsic and extrinsic purpose. And if I am correct in thinking that Pittendrigh is committed to a view of teleology grounded within a generic organism-artifact analogy, then this also poses serious problems regarding the possibility of making sense of the distinction in his view.

The heart of the problem lies in the fact that Pittendrigh never makes explicit what it is about an organization, whether animate or inanimate, that makes it goal-directed. As I've argued, one possibility is to be found in the 'two questions' raised

[27] More precisely, Paley's analogy is a *conclusion* of his argument rather than a premise. In this sense, it is an argument *to* analogy rather than *from* analogy (White et al. 2021, 58-62).

by organized entities: 1) the goals served by the organization and 2) the origin of the information which "underlies and causes" said organization. Plausibly, the goals served by an organized artifact such as a watch is ultimately explained by the originating causes of the artifact, namely, the intentions of its designer. Those intentions are intrinsically purposive and ground any claims regarding the corresponding *extrinsic* purpose (or function) of said artifact.

Extending the same logic to biological organization, one might wonder whether the information underlying it can plausibly explain that in virtue of which it is teleological. Pittendrigh makes explicit that in his view, the underlying cause of biological organization is the differential reproduction of organisms leading to natural selection (1958, 397). For this reason, biological organization is given a special name, "adaptation", and as such becomes "the core of biological study". But the only sense in which natural selection may be that which grounds the goal-directedness of biological organization is if it is an intrinsically purposive designer in the same way that a human designer is. Modern biology has no room for such a notion that isn't metaphorical (Lewens 2004); and if it is merely metaphorical, then this necessarily makes any claim regarding the teleological character of organisms metaphorical too. In this regard, Pittendrigh's teleonomy does nothing to clarify the distinction between intrinsic and extrinsic purpose. That being said, one does find in Pittendrigh's chapter a tentative suggestion at another way of thinking about the source of biological organization which, to Ernst Mayr, did promise to clarify this issue. This source was to be found within the genotype, as a repository of "information" containing "that inherited message, which specifies the constructional steps that comprise development" (Pittendrigh 1958, 395); or what would soon be dubbed "the genetic program".

2.3 Mayr's Teleonomy and the Genetic Program

Mayr first articulated his conception of teleonomy in his now famous article "Cause and Effect in Biology", where he introduced his distinction between "proximate" and "ultimate" causes. The distinction refers to two ways of explaining behavior. Proximate causes are typically the immediate, physiological and biochemical causes of a behavior, while ultimate causes are those which occurred in the evolutionary past (see Beatty 1994; Ariew 2003 for an analysis). Appeals to these different kinds of causes reflect two modes of explanation and two subfields within biology. Functional biology asks *how* questions and takes a more reductive and decompositional approach to its object of study.[28] In contrast, evolutionary biology asks *why* questions, and takes a more integrated, historical approach to organisms. former is appealed to in functional biology, which asks *how* questions, and the other in evolutionary biology, which asks *why* questions (Milam 2010). Counterintuitively, Mayr was adamant that teleonomy belongs to functional rather than evolutionary biology. In what follows, I will explain why he held this view.

A large part of Mayr's motivation for introducing the distinction between proximate and ultimate causes is to argue against the possibility of reducing biology to biochemistry. This concern for the eclipse of "organismic biology"—as a disciplinary configuration that included systematics, paleontology, and field ecology, among others— which he shared with his colleagues George G. Simpson and Theodosius Dobzhansky, was felt as coming both from within their discipline and without. Within their field, Mayr criticized both population geneticists such as Ronald Fisher, Sewall Wright, J.B.S. Haldane, and the increasingly dominant sub-field of molecular biology (Smocovitis 1996; Peterson 2016, 189–93; Morange [1998] 2020, chap. 23). Outside of biology, philosophers such as Ernest Nagel (1961) contrasted mechanistic biology with organismic biology, and argued that the latter would be eventually reducible to the former; and in fact, "Cause and Effect in Biology", which is Mayr's first piece on the philosophy of science, is aimed squarely at Nagel (Mayr 1961, 1501, fn. 1).

While Mayr recognized the success of the methods of biochemistry in functional biology, and granted that reduction was possible in this domain, he was

---

[28] This closely tracks Rosenblueth et al.'s usage of "functional".

staunchly against the possibility of reducing evolutionary modes of explanation to functional ones. In asking *why*, the evolutionary biologist seeks answers from the perspective of the whole organism and its contingent history. The essentially *historical* mode of explanation of evolutionary biology makes biology irreducible to other disciplines, because there is no history or contingency in physics and chemistry (Mayr 1961, 1502). In this regard, Mayr was following Pittendrigh's lead in claiming that biology deals with particular kinds of objects that require modes of explanation which, while distinct from those found in physics, remain naturalistic.

It is in this context that Mayr introduced the topic of teleology. As he saw it, the 'entelechies' or 'vital forces' posited by past thinkers had merely replaced one unknown with another (1974, 93). In contrast, he deemed the science of molecular genetics to be the first to provide a scientifically rigorous explanation for the appearance of goal-directed behavior in biology, in continuity with the fundamental insights of cybernetics. Acting purposefully, for Mayr, is made possible by the fact of being a *programmed* individual, which is true both of organisms and certain kinds of machines (Mayr 1961, 1503; 1974, 98). Computers are programmed to behave in specific purposeful ways, while organisms are controlled by the *genetic program*; a view which he shared with luminaries of the time such as François Jacob and Jacques Monod.

The genetic program emerged simultaneously in Mayr's 1961 article, as well as Monod and Jacob's "Genetic regulatory mechanisms in the synthesis of proteins", which was the result of many years of work on genetic regulation which earned them the Nobel Prize in 1965 (see: Kay 2000; Keller 2003 for a more detailed treatment of this work, and the many cybernetically-infused metaphors that came out of it, "such as messenger RNA").[29] While Monod and Jacob's understanding of the concept was initially slightly different to Mayr's, their conceptions converged over time. Particularly relevant for my purposes is the way they all

[29] Alexandre Peluffo (2015) has compellingly argued that it is likely that the notion of the genetic program had a common origin prior to the publication of these papers.

closely associated the genetic program with the notion of *teleonomy*.[30] Both Monod and Jacob expressed strong interest in the notion of teleology in their bestselling books published in 1971. In *The Logic of Life*, Jacob riffed on a saying—often misattributed to Haldane (Georgios 2017; Riskin 2020)—that the genetic program had given legal status to the biologist's illicit relationship with teleology. In doing so, he echoed Mayr's original claim that the genetic program made teleology, *qua* teleonomy, a "legitimate" concept (1961).

Mayr's concerns for keeping explanatory appeals to proximate and ultimate causes separate carried over to his conceptualization of teleology. Here, the distinction corresponds to two entirely different sets of phenomena and two different senses of teleology:

> On one hand is the production and perfecting throughout the history of the animal and plant kingdoms of ever-new programs and of ever improved DNA codes of information. On the other hand there is the testing of these programs and the decoding of these codes throughout the lifetime of each individual. There is a fundamental difference between, on the one hand, end-directed behavioral activities or developmental processes of an individual or system, which are controlled by a program, and, on the other hand, the steady improvement of genetic codes.
>
> (1961, 1504)

For the sake of simplicity, I will call these two senses of teleonomy 'proximate' and 'ultimate'. The latter applies to two distinct senses of teleology which Mayr is critical of: 1) unidirectional evolutionary sequences; and 2) teleology as a

[30] Jacob and Monod used both the notion of the genetic program and teleonomy in 1961, albeit in separate publications of that year. It was not until their later publications that they directly associated the concepts in writing. However, as Peluffo notes, the article which introduced the notion of program was originally intended to be part of a book written in 1959, entitled "Enzymatic Cybernetics" which was never completed, and which deals heavily with the issue of teleology (2015). It therefore seems that the association between program and purpose was already in Monod's mind in the late 50s. The pair may have first heard the term teleonomy from Davis' opening address in 1961.

designation for static 'systems' (as opposed to dynamic processes).[31] While I will discuss the latter in the next section, the former is worth briefly mentioning here. Applied to evolution, teleonomy implies that it has a direction or goal—an idea often associated with the theology of Teilhard de Chardin (Chardin [1955] 2008; see also: Dobzhansky 1968; Ayala 1972). Unsurprisingly, Mayr does not support such an interpretation. More interesting, however, is his critique of fellow biologists, such as Conrad Waddington and Francisco Ayala, who proposed non-theological interpretations of evolutionary teleology. For Ayala, evolution "can be said to be teleological in the sense of being directed toward the production of DNA codes of information which improve the reproductive fitness of a population in the environments where it lives." (Ayala 1970, 11; Mayr 1974, 96). The problem with such views is that they seriously distort the meaning of teleology, which is not merely the description of some trend but a process where future goals are set up: "Yet, natural selection is strictly an a posteriori process which rewards current success but *never sets up future goals*" (Mayr 1974, 96; see also: De Laguna 1962; Krieger 1998).

The correct, 'proximate' sense of teleology which Mayr wished to defend encapsulated all processes that have clear termination points. This includes what Mayr called "teleomatic" and "teleonomic" processes. The former refer to those inanimate processes that "are the simple consequences of natural laws", such as that of gravity of thermodynamics (ibid, 97-98). In this case, "They are 'end-directed' only in a passive, automatic way, regulated by external forces or conditions." (ibid, 98). Teleonomic processes, in contrast, are the result of a program, which not only guides the process but 'foresees' the "end point, goal, or terminus" on the basis of which behavior is regulated (ibid, 99).

Mayr was cautious to keep teleonomy restricted to the realm of proximate causes, as studied by the functional biologist. To muddle teleonomy with

[31] Mayr's argument on this point is somewhat confusingly laid out. While he makes this threefold distinction in section 2 of the paper, in which he proceeds to discuss the first two senses, the bulk of the discussion on 'teleological systems' occurs much later in section 4. What follows is therefore my best attempt at a charitable reinterpretation of his view.

adaptation, as Pittendrigh did, would breach this barrier, and bring teleological thinking into the realm of ultimate causation: "A system is capable of performing teleonomic processes because it was programmed to function in this manner. The origin of the program that is responsible for the adaptiveness of the system *is an entirely independent matter.* It obscures definitions to combine current functioning and history of origin in a single explanation." (ibid, 108; emphasis added). While he claims that "Natural selection does its best to favor the production of codes guaranteeing behavior that increases fitness", he is also adamant that the evolutionary process itself (and historical processes more generally) is not itself teleological (1961, 1503). These tasks correspond to the work of evolutionary biologists (who ask "why questions") and functional biologists (who ask "how questions") respectively. Such a division had the odd consequence of making teleonomic explanations answers to "how" questions rather than the "why" questions of the evolutionary biologists (Love and Dresow 2023, 109), since the ultimate causes of behavior cannot include reference to a historical purpose, or an *evolutionary* program (see: Waddington 1968 for a critique).

The notion of the genetic program, in other words, proved indispensable for Mayr not only to define teleonomy but to reinforce his distinction between ultimate and proximate causation, along with the division of labor among biologists which it implies (Peluffo 2015, 686).

At the same time, this account aimed to naturalize (rather than eliminate) teleology because it provided an answer to the "why" question distinctive of biology. An eliminative account would replace teleological statements like "the turtle went to shore *to* lay her eggs" with non-teleological statements like "the turtle went ashore *and* laid her eggs", and thereby makes "why" questions redundant. The teleonomic account, on the other hand, opened biologically relevant and interesting inquiry which could lead to investigations of the programs underlying behavior. Such a radical dichotomy between ultimate and proximate causation, as it was imposed on evolution and development, led to "[e]volution [being] thus seen as adaptive but not teleological, development teleological but not adaptive." (Moss 2004, 195–96; Rheinberger and Müller-Wille 2017, 89).

With these preliminary considerations in mind, we can now ask in what sense Mayr's own view was cybernetic—and whether he was successful in his attempt to resolve the ambiguities between intrinsic and extrinsic purpose which he perceptively identified.

2.4 What Makes Mayr's Teleonomy Cybernetic?

2.4.1 Proximate and Ultimate Causation

As I argued in the previous chapter, the behavioral/functional distinction played two interrelated roles for the cyberneticians, namely, the characterization of purposiveness as a behavior that could be modeled analogically with the behavior of machines, and a way of avoiding the problem of backward causation. While Mayr never explicitly took up the distinction, one sees how his own proximate/ultimate distinction in fact served those very same roles. I will focus on the latter of these roles in this subsection, before turning to the former in the next.

The behavioristic aspect of Mayr's view can be seen in his insistence that one should speak of teleonomy in processual terms, but not of static, teleonomic "systems". As he saw it, many scholars at the time had taken up Pittendrigh's notion of teleonomy as a referent for organized 'systems', because the structure or system in question was seen as "the product of evolution and of selective advantage". This 'system' view, for Mayr, was essentially a historical view of teleology as adaptation, and corresponded to the view which he saw being defended by the likes of Monod, George G. Simpson, R. Munson, and Davis (it was also held by George C. Williams, though Mayr did not mention him).[32] All

[32] Indeed, despite the similarities between Mayr and Monod's view, Mayr was critical of Monod's conception of teleonomy. Monod did not associate teleonomy quite as closely to the

these scholars, Mayr thought, misinterpreted both his own view and Pittendrigh's, who he wrongly believed to be in essential agreement with him, despite all evidence pointing to the contrary. As he argued, to extend the "the term teleological to cover also static systems leads to contradictions and illogicalities." (105). A torpedo, for instance, *can* behave teleonomically. But when it is sitting in storage, there is nothing that makes it *intrinsically purposive*.

This problem lies in a confusion between the potential *capacity* for an entity to behave purposively and teleology as an intrinsic property of an entity. Mayr considers this problematic because the notion of a teleological system in biology implies for Mayr the first sense of teleonomy as a "unidirectional evolutionary sequences" which he rejected: it "reflects just enough of the old idea of evolution leading to a steady progression in adaptation and perfection, to make me uneasy." (1974, 106). When calling a structure teleological we are naturally led to ask what goal the system is moving toward, or what goal the structure is "realizing" while it is inactive. Such questions are answerable only by attributing teleology to the process which brought the system about. And while this is a sound way of analyzing the goals of artifacts, it is not acceptable when it comes to evolution. In other words, Mayr sees the equation of teleology and adaptation as implicitly slipping to a teleological view of the evolutionary process itself: "To me, it is misleading to call adaptational statements teleological. 'Adapted' is an a posteriori statement and it is only the success (statistically speaking), of the owner of an adaptive trait, which proves whether the trait is truly adaptive (=contributes to survival) or is not." (Mayr 1974, 52). This critique of "system" talk is in effect an extension of his critique of historical or ultimate conceptions of teleonomy which clearly applies to Pittendrigh.

It is in this context that we can see how the genetic program, as a proximate cause in development, proves to be crucial to resolving the problem of backward causation for Mayr. In his 1974 article *Teleological and Teleonomic, a New Analysis*

---

genetic program, and when he did, he didn't have the proximate/ultimate distinction to keep teleonomy from just meaning adaptation (later reprinted with minor changes as 'The Many Meanings of Teleology', 1988), Mayr elaborated on what exactly he meant by 'program' and its relation to teleonomy: "a program is (1) something material, and (2) it exists prior to the initiation of the teleonomic process. Hence, it is consistent with a causal explanation." It is this "consistency" with causal explanation, also emphasized—but articulated differently by Pittendrigh, that is worth emphasizing. Consider his statement that the "genetic program has provided a mechanistic explanation of one class of teleological phenomena" (Mayr 1982, 48–49). What Mayr meant is that the program exists prior to the goals which it determines, while operating in concert with feedback mechanisms that ensure the end-state is reached in the face of perturbations. The source of goal-directedness (the genetic program) is entirely explicable in mechanistic terms, since the program is a material entity (Mayr 1974, 102). There is therefore "no conflict between causality and teleonomy" (Mayr 1961, 1506; 1974, 93–94). In other words, by arguing that teleonomic processes are initiated by the genetic program, Mayr can claim that the goal, being "foreseen" in the program, is a strictly causal process that requires no appeal to future states causing the present (see also: Carl G. Hempel and Oppenheim 1948, 142–43; Nagel 1961, 25; 1977, 264). In so doing, he was also able to give a privileged causal role to the genetic program in ontogeny, which played a critical role in legitimating the distinction between teleonomy in the 'proximate' and 'ultimate' sense.

The genetic program, in turn, allowed Mayr to resolve a major problem which he sees lying in the cybernetic account. Mayr is critical of the idea that negative feedback alone is sufficient for purposive behavior. Though negative feedback mechanisms do serve to increase precision and rectify errors—thereby allowing for a certain amount of plasticity and robustness in goal-directed behavior—Mayr argues that negative feedback never determines the goal itself. "Feedback devices are only *executive* mechanisms that operate during the *translation* of a program." (1974, 100; emphasis added). This is reflected in his claim that the "truly characteristic aspect of goal-seeking behavior is not that mechanisms exist which improve the precision with which a goal is reached, but rather that mechanisms exist which initiate, i.e. *'cause'* this goal-seeking behavior". Just as humans determine the target temperature of a thermostat and negative feedback mechanisms that execute it, the genetic program of an organism determines the goal-state of the system and the developmental and physiological mechanisms of the organism execute it (ibid).

In sum, while Mayr's teleonomy was not overtly behavioristic, the proximate/ultimate distinction played a role analogous to the behavioral/functional distinction, in so far as it allowed Mayr to argue for a particular dynamic and processual conception of teleonomy that does not impute goal-directedness to the mere fact of being organized in the right way (i.e., of being a particular kind of "system"). Secondly, it also allowed him to make clear the privileged causal role that the genetic program plays as the cause and regulator of these processes that simultaneously set the final state or goal which they will reach. By extension, a teleonomic explanation is fully consistent with a causal explanation. But in doing so, Mayr's view ends up going significantly counter to the cybernetic claim that "a teleological study is useful if it avoids problems of causality and concerns itself merely with an investigation of purpose…causality implies a one-way, relatively irreversible functional relationship, whereas teleology is concerned with behavior, not with functional relationships." (Rosenblueth, Wiener, and Bigelow 1943, 24). On the contrary, causality must play a role in the story for Mayr, because the goal—or rather, its representation in the genetic program—must act as a cause prior in time to its realization.

2.4.2 Mechanism and the Genetic Program

The second distinctive aspect of cybernetic teleology was its commitment to a mechanism derived from an analogy between organisms and servomechanisms. For the reasons stated above, this analogy fell short on Mayr's view: it included execution, but fell short of encapsulating the genesis of the goal-directed process. For this reason, Mayr's analogy (charitably interpreted as one of proportion) between genetic and computer programs had, in his own view, some distinctive advantages.

But it is worth emphasizing that this shift in mechanistic analogies had some important consequences for the relation between holism and reductionism. For while servomechanisms invited a somewhat holistic behavioral understanding of teleology, the genetic program locates the explanation of that behavior in a part of the behaving system.

It did not have to be this way. The genetic program not only built upon the idea of a genetic code and genetic information, which had given the genome privilege status among cellular organelles, but also doubled down on a long history of thinking about the genes as active agents (Keller 2003). But the idea that the genome contains information left open the possibility that this was a passive form of storage that required the action of a broader system, even an entire cell, to use it as a resource for cellular regulation. As David Nanney presented at a 1956 meeting on "The Chemical Basis of Heredity", this more holistic interpretation was countered by the interpretation of DNA as a "master molecule" that not only stored information but was also the active, self-controlling and self-regulating executor of that information (Nanney 1957). As Evelyn Fox Keller has argued, Monod and Jacob were decisive in shifting the cybernetic meaning of negative feedback and regulation as a holistic, decentralized property of a whole system to refer to specifically genetically controlled processes (2003, 170). In the words of Monod and Jacob, "The discovery of regulator and operator genes...reveals that the genome contains not only a series of blue-prints, but a coordinated program of protein synthesis and the means of controlling its execution." (Jacob and Monod 1961, 354). The notion of 'program' captured this dual capacity of information and execution, and in so doing enforced a particular *reductionist*-mechanist conception of teleology that was quite at odds with the cybernetic account. This fact was never addressed, perhaps because it was too subtle but also perhaps because Mayr wanted to keep the issue of reductionism and holism out of the discussion on teleology: "it only confuses the issue, when a discussion of teleology is mingled with that of such of extraneous problems, as vitalism, holism, or reductionism." (Mayr 1974, 95).

In sum, Mayr attempted to distinguish his conception of teleonomy from both those which he felt equated teleonomy with the notion of adaptation, and the cybernetic account that equated purposiveness with negative feedback. In the next section, I will show how both critiques can be seen as deriving from an attempt by Mayr to reestablish the distinction between intrinsic and extrinsic purposiveness.

2.4.3 Intrinsic and Extrinsic Purpose: What Lies Beneath Mayr's Critiques?

Even though Mayr never explicitly addressed the distinction between intrinsic and extrinsic purpose, he clearly had the distinction in mind when he defended his conception of teleonomy from the alternative 'ultimate' (i.e., historical) views.

Recall that the distinction between intrinsic and extrinsic purposiveness derives from the distinction between the purposiveness of agents and that of artifacts. A human agent may have the goal of chopping wood, and be said to be the originating locus of this purposeful activity. The axe which is used in the process, in contrast, can be said to have the function to chop wood, but only derivatively or extrinsically; i.e., on the basis of the agent's goals.

Consider, then, Mayr's critique of teleonomic systems or Pittendrigh's teleonomy as adaptation. His concern—namely, that a definition of teleonomy as adaptation renders the process of evolution teleological—is just another way to say that adaptation is only teleological in the extrinsic sense. Just like an artifact whose purpose is derivative of a history of creation *for* a purpose, an organism's adaptation would be said to be 'for a purpose' based on a history of natural selection. But for this to be true, evolution itself would need to be the intrinsic locus of purposiveness.[33] This is why one must distinguish purpose from function; a point which Mayr hinted at, albeit only briefly. In his discussion of stationary

[33] The alternative that Mayr did not consider, which emerged around this time and has since become mainstream in philosophy of biology, is that one can consider the products of design and design-like processes (such as natural selection) to be teleological without the process itself being teleological. As will become clearer in Chapters 3 and 4, this view is either incoherent or so radically distorts the meaning of teleology or function that it effectively eliminates the concepts.

systems like a torpedo in storage, he conceded that the system could be teleological in the same sense that a poison has the potential to be used to kill somebody "but this inherent property does not make it a goal-directed object [viz. intrinsically purposive]. Perhaps this difficulty can be resolved by making a terminological distinction between functional properties of systems and strict-goal directedness" (1974, 106). The distinction between "functions" and "goals" is essentially that of intrinsic vs extrinsic purposiveness. And Mayr is right in stating that an inactive torpedo can be deemed functional in the sense that it was designed to serve a particular purpose. He is also right to call biological organs "functional" in the same way that a hammer is functional, or even a rock if it happens to suit that purpose (ibid, 107). But none of these are "teleological systems" in the sense of *intrinsic purpose*. May's invocation of his proximate/ultimate distinction in this very context reveals the way that this distinction serves as a proxy for the intrinsic/extrinsic distinction (ibid).

All of this is also true, in my view, of Mayr's critique of the cybernetic view of teleology as feedback. Mayr here is clearly using the notion of being a *cause* or *initiator* of goal-directed behavior as a proxy for the notion of intrinsic purpose. On this basis he could claim that what truly initiates the goal-directed behavior of a thermostat is the human agent's setting of the desired temperature. But the biological analog to the human in this case had to be located within the genetic program (Moss 2004). Arguing against the likes of G. C. Williams who wished to replace Aristotelian final cause with "the material principle of natural selection" (Williams [1966] 2018, 258), Mayr argued that the genetic program in fact vindicated the Aristotelian notion of *eidos*—or the form giving principle that which "reflects the terminal telos of the full grown individual." (Mayr 1974, 111). Favorably quoting Max Delbrück, who stated that the notion of the "'unmoved mover' perfectly describes DNA: it acts, creates form and development, and is not changed in the process" (Delbrück 1971, 55), Mayr claimed that "the existence of teleonomic programs—unmoved movers—is one of the most profound differences between the living and the inanimate world" (Mayr 1974, 122).

To fully appreciate the philosophical move that Mayr was making here, it is worth emphasizing how pervasive and confused thinking about teleology in the context of adaptation and natural selection was at this time (Dresow and Love 2023). At the heart of this confusion is the analogy between natural and artificial selection, which goes back to Darwin himself. It is well known that Darwin explicitly used artificial selection as a model for thinking about natural selection, and may have used the analogy in his thinking about teleology (White, Hodge, and Radick 2021; see also: Lennox 1993; 2013). Less well known is the fact that in his own notes, Darwin used the notion of a 'selecting demon' as a stand-in for the human artificial selector (Canales 2020, 44–48). Much like Maxwell's demon, this creature operated as a stand-in for a not fully understood process, and thereby played a role as a cognitive aid or heuristic. Alfred Russell Wallace saw such a demon being implicitly invoked by the notion of "natural selection" and suggested to Darwin that they use Herbert Spencer's "survival of the fittest" instead (Alfred Russell Wallace 1866).[34] A century later Darwin's demon reappeared, alive and well, in George C. William's description of the sterile cast of bee workers whose existence is "entirely attributable to the unrelenting efforts of Darwin's demon" (Williams [1966] 2018, 34). The demon was also implicitly invoked by Jacob, who invoked evolution as the author of the authorless genetic plan (Jacob [1970] 2022, 2); a creature that would live as the "blind watchmaker" (Dawkins 1996). But if natural selection makes organisms teleological, *it can only be because the demon itself is the intrinsically purposive counterpart.*

If there is no such demon, then teleological ascriptions within this design framework become seriously confused. Ayala, for instance, wanted to reject the

[34] Interestingly, Wallace had previously analogized natural selection to the governor of a steam engine, which regulates the flow of fuel by negative feedback (Alfred Russel Wallace [1858] 2009). On a cybernetic interpretation, this would make the process of natural selection intrinsically purposive. Bateson later claimed that "If it had been Wallace instead of Darwin [who started the trend], we would have had a very different theory of evolution today. The whole cybernetic movement might have occurred one hundred years earlier as a result of Wallace's comparison between the steam engine with a governor and the process of natural selection" (Bateson 1979, 43; cited in: C. Smith 2004).

artifactual metaphor while also claiming that natural selection could account for intrinsic purpose:

> The end-directedness of living organisms and their features may be said to be "internal" teleology, while that of man-made tools and servo-mechanisms may be called "external" teleology… Internal teleological systems are accounted for by natural selection which is a strictly mechanistic process. External teleological mechanisms are products of the human mind… The overall process of evolution is not teleological in the external sense. Evolution can be explained without recourse to a Creator or planning agent external to the organisms themselves.
>
> (Ayala 1970, 11)

Similar confusions are to be found in a chapter by Dobzhansky on natural selection, in a book co-authored with Ayala, G. L. Stebbins, and J. W. Valentine (Dobzhansky 1977). While claiming that "The origin of organic adaptedness, *or internal teleology*, is a fundamental, if not the most fundamental problem of biology" the authors also claimed that internal teleology is "a product of evolution by natural selection" (ibid, 96, emphasis in original). The point was made most explicitly by Pittendrigh some years later, in an essay reflecting on his life's work:

> "The primary purpose of biological organization is self-perpetuation by self-copying, and as such is the handiwork of Darwin's Demon […] the commonplace nature of programmable machines at midcentury gave teleology (as teleonomy) complete respectability in the society of biological ideas. *The genome was the program of a Turing machine and Darwin's Demon was the programmer*."
>
> (Pittendrigh 1993, 20)

Mayr's sensitivity to the inherent problems of these views is commendable. But while his criticisms were on target, the genetic program was unable to resolve these problems by putting the intrinsic/extrinsic distinction on firmer ground (Keller 2003, 142). Indeed, the genetic program in fact only appears to be a resolution to these issues because of significant ambiguities which it traded on. Evelyn Fox Keller has identified four such ambiguities, but two are particularly relevant here.

First, there is an ambiguity regarding the status of genes as both subjects and objects of the genetic program, given that they are treated both as the controllers and the controlled, the regulators and the regulated, the switches and the entities to be turned on and off, the activators and the activated (Keller 2003, 136-137). This created an essential ambiguity in the locus of agency, which would have needed to be found in the genetic program for it to play the role of intrinsic purposiveness that Mayr needed it to. For this to be the case, genes would need to be "subjects capable of activating, regulating, and controlling" themselves (ibid 138). In other words, Mayr's teleonomy rested on a specific conception of the molecular genome, endowed with a range of extraordinary and contradictory powers ranging from the storage of a developmental blueprint to the active control and regulation of its execution (Moss 2004). The empirical evidence for this has not borne out (Walsh 2020; McKenna, Gawne, and Nijhout 2022).

A second, related though more serious source of ambiguity came from the invocation of the machine analogies and metaphors which threatened to undermine Mayr's attempts to overcome the shortcomings of the accounts of teleology which he criticized. Consider, for instance, his statement that it is the naval officer, rather than the torpedo itself, which determines its target (Mayr 1974, 100). Putting aside the obvious problem of individuating and locating "mental" programs that would ground intrinsic purposiveness in the naval officer rather than the torpedo itself, the very idea of a natural program undermines the barrier he wished to erect between ultimate and proximate senses of teleology. A program, as an artifact, is only teleological *in virtue of being an artifact*; of being *extrinsically* designed for a purpose, by definition. As such, there is no way to make this metaphor work as a ground for intrinsic purpose. This effectively suggests that Mayr's proximate-ultimate distinction collapsed when it came to distinguishing two types of teleology; that his "proximate" account of intrinsic purpose collapsed into the very evolutionary account he was criticizing. The goal of the genetic program is to direct the processes that it was evolutionarily "designed" to direct, and in this sense the program ends up being equivalent to a thermostat, and the evolutionary process to a demonic agent. In other words, the genetic program as a locus of intrinsic purpose not only required the genome to have some very special properties; it also required conception of teleology that could be fully extricated from the metaphors of artifice and intelligent design that it relied on.

2.5 Conclusion

This chapter closely analyzed Pittendrigh and Mayr's conceptions of teleonomy, along with their claim that cybernetics was a primary source of inspiration for them. Using the three distinguishing features suggested in the previous chapter, I have shown that both senses of teleonomy departed significantly from the cybernetic account of purpose.

First, neither Pittendrigh nor Mayr took seriously the distinction between functional and behavioral modes of study (though Mayr came close to it in his critique of teleonomy applied to static structures rather than processes). Despite this apparent difference, I emphasized that the behavioral/functional distinction was intended to play a role in reconciling a reductive causal explanation and a teleological explanation operating at a more coarse-grained level, thereby resolving the problem of backward causation. From this vantage, one notices how Pittendrigh and Mayr invoked their own distinctions—explanation/description and proximate/ultimate causation respectively—to deal with the same problem. While Pittendrigh argued that teleonomy is purely descriptive rather than explanatory, Mayr argued that one needed to see intrinsic purpose as a proximate cause that had its origin in a particular kind of entity: a genetic program.

Second, while both Pittenrigh and Mayr's teleonomy were mechanistic, neither subscribed to the particular analogy of proportion between organismal and machine behavior defended by the cyberneticians; nor did either seem to think carefully about the nature and scope of the analogies they were proposing. Pittendrigh leveraged a generic organism-artifact analogy, while Mayr put forward an analogy between genetic material and computer programs.

Finally, both struggled with the distinction between intrinsic and extrinsic purpose in similar ways. While Pittendrigh's teleonomy straightforwardly failed to make sense of the distinction, Mayr made a concerted if unsuccessful effort to cast the genetic program as a source of intrinsic purpose. As will become clearer over the next chapter, this is because both Pittendrigh and Mayr failed to escape the artifact model of teleology. It is, in fact, conceptually *impossible* to make sense of intrinsic purpose within the scope of an organism-artifact analogy, given that the only sense in which purpose enters into the analogy is as *extrinsic* purpose, which necessarily needs to be cashed out in an account of *intrinsic purpose*.

CHAPTER THREE

The Continental Critique of Cybernetics

The preceding chapters traced three core issues from purposive behaviorism, to cybernetics, and teleonomy. First among these is the question concerning the relationship between teleology and causation, and most notably the problem of backward causation. The struggle to articulate a realist conception of purpose that did not reduce to a causal explanation led scholars to draw various distinctions that contrasted different explanatory modes and perspectives. Purposiveness was variously given a place within laws of behavior, behavioral methods of study, or teleonomic descriptions. In this regard, Mayr proved to be a notable exception in the way that he attempted to assimilate teleonomy within a causal framework by citing the genetic program as a privileged kind of cause in development. The second recurring theme is the question of whether intrinsic purposiveness can be adequately modeled by machines and artifacts, to which nearly all the scholars so far mentioned gave a positive answer. This second issue proves closely related to the third, namely, the persistent struggle to make sense of the distinction between the intrinsic purposes of agent, and the extrinsic purpose that they bestow unto artifacts. In many cases, this issue was ignored or glossed over. In others, there was a concerted effort to show that intrinsic purpose could in fact fall *within* the scope of the organism-artifact (O-A) analogy. Such a claim radically alters the scope and meaning of the O-A analogy, given that intrinsic purpose had traditionally been one of the most important aspects of the *disanalogy* between them. As I have argued, there is good reason to think that this simply cannot work, and that it constitutes a very serious confusion regarding the scope of the analogy and the meaning of intrinsic purpose.

In this chapter, I will flesh out these critical claims and show more explicitly why the artifact model of teleology fails. Doing so will require revisiting the distinction between intrinsic/extrinsic purpose and showing that it lies at the core of the very concept of purposiveness. I will do so by drawing together a series of critiques which were made of the cybernetic account in the 50s and 60s. Although Richard Taylor's critique is often remembered for the response it received from Rosenblueth and Wiener (1950), their account of teleology was critiqued by many others during and after its heyday (e.g., Churchman and Ackoff 1950; Von Bertalanffy 1950; Wisdom 1951; Scheffler 1959; Wimsatt 1971). A recurring theme in many of these critiques was the behavioristic framework that the cyberneticians operated under. As discussed in the first chapter, these shortcomings were already evident decades prior to cybernetics, particularly when it came to identifying goals in terms of objects or features in the environment. For example, in 1950 Taylor argued that there are many cases in which a system's goal cannot be identified as a feature of the environment, because the goal simply does not exist (e.g., the Holy Grail) or the goal is an as-of-yet unrealized state (e.g., the adult form of an organism). Indeed, even torpedoes should be said to be aimed at the *destruction* of the ship, not the ship itself. Moreover, the dependence on a feature of the environment to be a goal-object seemed to render the behavior of the purposive system excessively passive, as merely reacting to these features rather than actively pursuing them 'from within' (Taylor 1950a; Von Bertalanffy 1951). In those cases, it seems that we need to appeal—at least in part—to something intrinsic to the system, such as a representation of the Grail in the nervous system of King Arthur's knights, or in Mayr's view a goal state encoded in the genetic program.

I contend that a similar conceptual thread runs through the critiques levelled by a group of continental philosophers, most notably Hans Jonas, Georges Canguilhem, Gilbert Simondon, Maurice Merleau-Ponty and Raymon Ruyer, who rarely if ever appear in anglophone histories of cybernetics (e.g., Heims 1991;

Johnston 2008; Pickering 2010; Kline 2015). Though their critiques ranged in scope, content, and motivation, one finds in their thinking a clear convergence on a set of ideas that, while broadly in line with Taylor's criticisms, go further in articulating a better account of the distinction between intrinsic and extrinsic purpose. Focusing on the work of Jonas, Canguilhem, and Simondon in particular, I endorse their claim that a grounding for the concept of teleology—along with making sense of the intrinsic/extrinsic distinction—requires thinking through the proper criteria by which something can be a locus of purposiveness, i.e., the kind of entity to whom we can ascribe such goals. I will use the term 'agent' to denote such an entity, even though the continental critics (henceforth "CCs") used alternative terms such as 'individual', 'self', or 'subject'. My use of the term 'agent' in this context is intended to highlight where these terms converged, namely on the idea that a proper account of teleology must be grounded on a particular kind of entity to whom we can ascribe those purposes intrinsically. I will argue that the CC significantly advances two of the three core issues so far discussed, namely that of the distinction between intrinsic and extrinsic purpose, and the analogy between organisms and artifacts. However, it does little to clarify the third issue, namely the relationship between causation and teleological explanations. This argument will set the stage for Part 2 of this dissertation, which will use this groundwork to tackle that third issue.

The structure of this chapter is as follows. In the next section (3.1), I briefly introduce the CC to both explain and contextualize the relevance of their thinking for understanding the limitations of the cybernetic account of teleology. In section 3.2 I will outline what I take to be the focus of their critiques. In section 3.3, I will trace the curious convergence of the CCs and cybernetics in the early 2000's, in the form of Hans Jonas' integration into the theory of autopoiesis. I conclude with reflections regarding the fundamental importance of the intrinsic/extrinsic distinction for any account of teleology.

3.1 The Continental Critics

Jonas, Canguilhem, and Simondon were among the most prominent continental biophilosophers born in the first quarter of the 20th century.[35] They shared a deep interest in several major philosophical questions, many of which circled around the problem of *life*, and the associated problem of teleology (or 'finality' as it appears in French). It is from this perspective that they each critically appraised the cybernetic movement and its claims.

While cyberneticians were keen to show that *intrinsic purpose* could, contra Aristotle and Kant, be rendered intelligible by the O-A analogy, the CCs attempted to re-establish what was the more traditional understanding of what such an analogy comprises. The CCs are particularly important for my purposes insofar as they help clarify the meaning of 'intrinsic' and 'extrinsic' teleology by venturing into questions which the cyberneticians assiduously avoided—namely, the metaphysics of organisms and artifacts which, in turn, determines the nature and scope of the O-A analogy.

They articulated those same ideas in different ways. Jonas framed his critique in terms of the perennial debate between mechanism and teleology which featured so prominently in the German tradition which he emerged from (Jonas 1953). In contrast, Canguilhem, Merleau-Ponty, and Simondon's responses reflected the distinctively 'French' uptake of cybernetics which occurred during the period (Mindell, Segal, and Gerovitch 2003; Johnson 2015; Le Roux 2018). This reception was wide reaching both in terms of its audience and its content; it therefore spanned far more than the issue of naturalizing teleology. As products and shapers of this culture, they certainly had much to say about cybernetics as a whole.

---

[35] In contrast to analytic philosophy of biology which engages in 'higher-order theorizing' on the work of biologists, biophilosophy takes insights from biology and associated disciplines and asks how these findings might impact philosophical debates, such as those between substantialist and processualist ontologies (D. L. Smith 2017; Wong and Wolfe 2014; Wolfe 2015). In other words, while the former takes philosophy to be a resource for biology, the latter sees biology as a resource for philosophy.

These scholars were also motivated by different concerns. Underlying Jonas' critique was a distinctive fear for the destructive powers of technology (Jonas [1966] 2001, xiv–xvii; see also: Aumann 2011 for the German reception of cybernetics). These concerns can be traced back to his early years as a Jewish doctoral candidate in the 1920s under the supervision of Martin Heidegger and Rudolf Bultmann (Wolin 2015, chap. 5). During this time, he completed his dissertation on freedom, spirit-matter dualism, and Gnosticism (Jonas 2001, xi–xii). This early work established a life-long concern for the question of nihilism and the effects of technology on our modern way of life, both of which influenced his reception of cybernetics. Jonas left Germany prior to the Second World War and eventually settled in North America in the early 1950s (ibid). During his time in Canada, he came into contact with another emigree, Ludwig Von Bertalanffy, thanks to whom he was likely first exposed to general systems theory (1951) and cybernetics (Jonas 1953; Tibaldeo 2008; 2012).

Canguilhem's philosophical work was also profoundly affected by his experiences during the Second World War, particularly his time in the French resistance (R. Esposito 2008, 189). While Jonas was centrally concerned with the destructive powers of technology and the threat of nihilism, in Canguilhem we see a special concern for issues in medicine, particularly the categories of normality and pathology, which were the topic of his first and most famous book, the 1943 *Essay on Some Problems Concerning the Normal and the Pathological*. Originally his doctoral thesis in medicine, it is a first attempt at developing what we might now call a holistic perspective on biomedicine. These were concerns he shared with many organicist and holist interwar physiologists, including Walter Cannon, who found their theories of the body unable to explain the profusion of new kinds of pathologies coming out of the first World War, most notably shell shock (Geroulanos and Meyers 2018), which contributed to the formation of the concept of homeostasis and influenced cybernetic thinking (ibid; Bates 2014). It was perhaps this overlap that drew Canguilhem's interest in cybernetics. Canguilhem's critique of cybernetics is particularly useful for my purposes in so far as he brings to the fore the relationship between the concepts of teleology, normativity, and function.

Simondon was one of Canguilhem's PhD students, and he paid close attention to—and most intimately understood—the scientific developments of the time, including quantum mechanics, thermodynamics, and cybernetics. In two of his major works, including his doctoral thesis (originally published in parts in 1964) entitled *Individuation in Light of Notions of Form and Information* (2005; henceforth *Individuation*), and *On the Mode of Existence of Technical Objects* (*On the Mode*; [1958] 2012) center on the notion of information and gaining a better understanding of the nature of technological objects. The term 'information' is a direct reference to information theory and cybernetics, which profoundly influenced his thought (Bardin 2015, 3). In his latter work, we find evidence that he read the transcripts from the Macy Conference which, as Dupuy remarks, was likely uncommon among his French contemporaries ([1994] 2009, 40; see also Mills 2014, 55). In 1964, he organized a conference entitled *The Concept of Information in Contemporary Science*, to which he invited Wiener as a guest speaker (Hart [1958] 1980; cited in: Mills 2014, 55).

Taken together, Jonas, Canguilhem, Merleau-Ponty, and Simondon's thought provides much needed clarity on the nature of machines and the limits of the O-A analogy. Chapter 1 discussed how central theoretical and mathematical modelling were to cybernetics and its transdisciplinary 'scientific style', and how these modelling efforts were founded on a proportional analogy between organisms and servomechanisms (Abraham 2012). The foundational importance of this analogy to cybernetic thinking, and the ways in which it became sedimented and accepted through metaphors that closed the analogical gap, were so deeply cemented that Gordon Pask, a second generation cybernetician, was reported to have called cybernetics "the science of defensible metaphors" (von Foerster 2003, 288). But insofar as the O-A analogy was a precise analogy of proportion at times and a metaphorical shorthand in others, one of the key questions that the CCs raised was the extent to which it was defensible. Rosenblueth and Wiener defended it on practical scientific grounds: analogies could generate useful models that might advance science by bridging and integrating previously separate fields of research, suggesting new hypotheses and experiments. On the one hand, one might think that this was a way of bracketing the issues which had troubled Perry and other purposive behaviorists, who by cybernetic lights seem overly concerned with metaphysical questions concerning the 'true nature' of intrinsic purpose. But this bracketing proved difficult to maintain against the tendency to perniciously reify their models and metaphors. The physiologist Ralph W. Gerard evoked this concern at the seventh meeting of the Macy Conferences on the cybernetics:

> It seems to me, in looking back over the history of this group, that we started our discussions and sessions in the ''as if'' spirit. Everyone was delighted to express any idea that came into his mind, whether it seemed silly or certain or merely a stimulating guess that would affect someone else. We explored possibilities for all sorts of ''ifs''. Then, rather sharply it seemed to me, we began to talk in an ''is'' idiom. We were saying much the same things, but now saying them as if they were so.
>
> (Gerard 1951, 11, cited in; Abraham 2012)

These were major concerns for the CCs. As David Mindell and colleagues state, in Europe "the 'as if' does not become an 'is' like in America" (Mindell, Segal, and Gerovitch 2003, 81; see also: Dupuy [1994] 2009, 127). The point was put more explicitly by Raymond Ruyer, a contemporary of Simondon: "Everyone knows feedback 'simulates' teleology; but we must know if this is an isomorphism or a caricature" (Ruyer 1961, 165). Merleau-Ponty was particularly concerned with the ways in which mechanistic thinking became self-referential, in this was leading to an "absolute artificialism" (Merleau-Ponty [1961] 2004, 292; quoted in Mazviiita Chirimuuta 2020, 499). As Chirimuuta explains, he worried that an analogy with machines inspired the creation of new devices which "are then projected back onto living organisms as models of their workings, and through the cumulative and recurrent effect of this process it becomes impossible to think of the organism—

including the human being—in any other terms than as a tool, a thing to be manipulated and an instrument at the service of interminable projects of intervention and control." (ibid, 450). Such a conceptually hermetic space was deemed dangerous because it loses sight of the real phenomenon which one sought to originally understand.

Simondon went further in arguing that the cybernetic analogy between organism and servomechanism ultimately does a disservice to both, by obscuring or even distorting the meaning of life and technology. Veritable analogical reasoning, for Simondon, does not use analogy in such a way as to collapse or assimilate the identity of the relata, but takes seriously the fact that analogy is always already based on the existence of difference (Simondon 2020, 107; Mills 2014, 62–63); a point which mirrors the distinction between analogies of proportion, which seek to uncover particular invariant features, and analogies as similitude, which seeks to uncover the essential 'sameness' of two things (see the Introduction to Part 1). And on the occasions that the cyberneticians were being careful, they did meet Simondon's standards of reasoning; the problem is that they frequently failed to do so. As I shall argue, this is a problem not just because it obscures the metaphysics of organisms, but also the metaphysics of artifacts. Perhaps less charitably, Jonas saw cybernetic terminology as "*intentionally* ambiguous and metaphorical" and thereby "facilitates the transfer back and forth between the artifact and its maker" (1953, 174). He not only saw this as misleading, but also took the more radical position that "the problem of mechanism or teleology, as also that of wholeness, are (to use the disreputable word) metaphysical and not logical issues. Mathematics cannot decide them, *indeed does not pronounce on them at all*" (Jonas 1951).

The above points to an important idiosyncrasy of Jonas' thought that puts him at odds with Canguilhem and Simondon. For Jonas, the phenomenology of our own agency plays a critical role in his conception of biological purposiveness. As he saw it, our theorizing about organisms originates with the inference from one's 'first-person' experience of purposiveness to that of a living organism. As Jonas recognizes, such anthropomorphic claims amount to "scientific high treason"

([1966] 2001, 35). But since he sees no other, non-phenomenological grounding for teleology, he believed that we are forced to either take seriously the phenomenological perspective as evidence for the reality of teleology in nature, or to submit to the total domination of mechanistic thinking over all phenomena (including humans) (ibid, 35-37). I shall return to this at the end of this chapter.

3.2 The Missing Agent: Why Cybernetics Fails to Naturalize Teleology

3.2.1 Individuating Agents

At the heart of the CC is the claim that the behavioristic criteria used by cyberneticians to ascribe goals to systems fail to properly distinguish intrinsically and extrinsically purposive systems. To illustrate his point, Jonas appealed to our common-sense intuitions about the ways in which we ascribe purposes to humans when they are driving vehicles or ships. In such situations, the pilot is an agent controlling or driving the otherwise passive vehicle, even if the pilot is using mechanical receptors and actuators to guide the ship. In such cases, Jonas claims, there is no problem in localizing purpose in a particular part of a larger human-mechanical assemblage; and there is no sense in which the human and pilot have merged "into one purposive whole" (Jonas [1966] 2001, 118). And yet, this is the conclusion which cyberneticians would come to, just as they would treat the same ship if the pilot were replaced with a servomechanism. The locus of purposiveness would now be "the receptor plus the effector plus the coupling, and in the form of organization of the whole system" (ibid, 118), or in other words, the amorphous assemblage of the servo-mechanism and ship together. It seems that the cybernetic view is forced to take such a maximally 'holistic' or 'molar' behaviorist view, despite our strong intuitions that the locus of purposiveness indeed resides in the pilot.

Alternatively, if the cyberneticians could coherently endorse the view that purposiveness resides in the pilot, then the same would have to be true of the servomechanisms which replaces the pilot. Jonas rejects this: the pilot's purpose is not equivalent to a process of negative feedback between their senses and their actions. The pilot's actions to keep the plane stabilized is only a description of means to a greater end that is not accounted for. A description in terms of input and output *already assumes* a purpose "interposed between them", i.e., *intrinsic* to the pilot themselves (ibid, 185). This appeal to the pilot's intentions can be ignored only when they are following orders, i.e., intentions *extrinsic* to theirs. In such a case, the pilot is acting *as* a robot (which is derived etymologically from the Czech *robota*, meaning forced labor; ibid [1966] 2001, 123; c.f. Mindell, 2002, 8). Though a cybernetic description is more appropriate in this case, it still leaves the commander's purpose unexplained. The position therefore falls to infinite regress, which can only be stopped once the grounds of human purposiveness have been explained in the first place: the "ultimate concerns; the ends by which we live" ([1966] 2001, 120).

This critique introduces one of the key issues raised by the CCs, namely, the need to more adequately account for the phenomenon of agency found in organisms, and the fact that agents are the kinds of entities to whom we can ascribe goals intrinsically. For the CCs, being an agent means not simply being capable of *behaving* in a certain way if it remains agnostic—as the cybernetic account intentionally does—about the kind of thing that is behaving. In other words, for the CCs there can legitimately be something that *appears* to behave in an intrinsically purposive way without being so—a position which Wiener and colleagues had at first taken toward servomechanisms, but slowly abandoned as the 'as if' turned into an 'is'. As I shall argue, this critique can also be extended to the notion of teleonomy and the genetic program.

For the CCs, what it takes to be an agent is to be, in some way or another, an entity that strives to keep itself alive. According to Jonas, the 'ultimate concerns' of a pilot are grounded in a purposiveness common to all living beings: to actively resist disintegration, to exist in a way which promotes the continuation of that very existence (Jonas 1968, 233). As he famously puts it: "'To be' is the organism's intrinsic goal. Teleology comes in where the continuous identity of being is not assured by mere inertial persistence of a substance, but is continually executed by something done, and by something which must be done in order to stay on at all…" (ibid, 243). In other words, an organism's existence is dependent on the continuous and precarious need for metabolic self-maintenance, coupled with a constant flow of material-energetic exchange with a surrounding environment. It is this metabolic process of self-maintenance that constitutes a necessary condition being an agent or locus of intrinsic purpose, because in so doing the system actively constitutes itself as an individual and establishes its own survival as an intrinsic goal. Canguilhem hints at a similar view when he appeals to the distinctiveness of biological organization (Canguilhem [1989] 2012, 76). While he does not explicitly state what he means by this, he is certainly committed to the view that life is not merely a process of negative feedback, but one in which negative feedback is only part of a larger system which can create new set-points, new norms, in response to environmental challenges (Canguilhem [1966] 1991). It is only at this level of the whole organism, as an individual within its milieu, that we truly grasp the intrinsic purposiveness of life. And while the cyberneticians were right in emphasizing the entanglement between organism and environment, the CCs did not see this as necessitating an abandonment of any distinction between them (c.f. Ashby 1953, 94). For both Canguilhem and Jonas, the wellspring of intrinsic purpose and normativity is found in the organismal need to continuously and precariously negotiate their relationship with their milieu.

Simondon, for his part, puts emphasis on the ways in which agents are not just individuals but always in the process of individuating themselves. While this idea is shared with Jonas and Canguilhem, Simondon makes this point the most clearly. Simondon illustrates his conception of individuation with the example of crystal growth from a supersaturated solution. The crystallization process begins with a seed relative to which the supersaturated solution is a metastable milieu, where 'metastability' is a thermodynamic property used to describe a system that is in a precarious state of equilibrium that is not its lowest energy state.

Crystallization occurs as this metastable solution crosses a threshold and begins to find "resolution" in the form of a more stable crystalline state, the formation of which releases the energy stored in the solution in the form of heat (Simondon 2020, 5; 68-88). To conceptualize the crystal as an individual challenges any simplistic, synchronic notions of identity, because the crystal is in a continuous process of changing and growing, surpassing itself, and at each moment creating the potential to grow further if there remains enough energy in the solution. The crystal is in this sense in a relation with its pre-individual reality (the solution) which sustains its growth; a process which could theoretically continue indefinitely if the solution is continuously replenished (ibidem; 159).

The logic underlying a crystal's process of individuation is also at work in the individuation of living beings, but with some crucial differences, which begin to reveal where Simondon sees the cybernetic conception of organisms as servomechanisms to be incorrect. Much like a crystal, an organism exists in a milieu rich in potentials which it uses to individuate and differentiate itself. Unlike a crystal, however, in an organism we find a kind of recursivity whereby it "conserves within itself an ongoing activity of individuation; it is not merely a result of individuation, like the crystal or molecule, but a theatre of individuation" (ibid, 7). In other words, though crystals always have the potential to grow further, this potential is fully determined by the presence of supersaturated solution, which the individuation of the crystal will always tend to exhaust. Organisms, on the other hand, preserve this potential by keeping themselves in a metastable state, which in turn preserves that capacity to continue to individuate themselves. This capacity is akin to what Mark Bickhard has dubbed 'recursive self-maintenance' (Bickhard 2009; 2017); they maintain the very capacity for self-maintenance. It is precisely this capacity to amplify an initial individual "that is not done by the technical object to which cybernetic mechanism would want to functionally compare it" because the capacity is not captured by "an automaton that would maintain a certain number of equilibria or would seek compatibilities among several requirements based on a formula of a complex equilibrium composed of simpler equilibria" (Simondon 2020, 7). The key for Simondon is that organisms are not simply homeostatic systems driven by negative feedback, but networks of many homeostatic networks which act in concert—communicating across levels, which Simondon calls "internal resonance"—as part of a larger "theatre of individuation" (ibid, 7). In the living being the process of individuation does not merely occur at the edge of the system where it meets its milieu as is the case with crystallization, but rather is maintained 'from within' which gives the organism a "veritable interiority" (ibid).

The foregoing has important consequences for how we understand organisms that actively survive by constantly adapting to changing conditions, not merely by tracking a changing variable (as the servomechanism does) but by altering and reorganizing itself from within. As Simondon puts it, "the living being resolves problems, not just by adapting, i.e. by modifying its relation to the milieu (like a machine is capable of doing), but by modifying itself, by inventing new internal structures…" (ibid). This very same point is made in *On the Mode*: "the living being is grounded in homeostases so as to develop itself and to continue its coming-into-being, rather than remaining perpetually in the same state it surpasses them and makes use of them" (Simondon [1958] 2017, 162). The upshot here is that homeostasis—stability—*is a means to an end for the organism, not an end in itself*. This point was missed by the likes of Ross Ashby who claimed that the need to keep a ship's dials within a given range was a good analogy for the variables that an organism must keep within homeostatic balance (Ashby 1953, 73). In fact, for Simondon the organism must not only know when to remain stable, but also when to change its internal structures. Servomechanisms can remain in a stable equilibrium, but they do not maintain a state of *metastable* equilibrium. And since the "system of stable equilibrium is one that has attained the highest degree of homogeneity possible … it cannot explain action to any extent, for it is the system within which no transformation is possible, since all potentials have been exhausted: it is a dead system." (Simondon 2020, 235). Like Canguilhem, for Simondon technology is designed to be rigid and singular in its functions, whereas organisms have an open-ended potentiality: they not only pursue goals but create new ones.

A further disanalogy between automatons and living beings is articulated in *On the Mode*, where Simondon states that one of cybernetics' key shortcomings was its "improper identification of the technical object with the natural object and more specifically with the living being." (2017, 50). He then goes on to explicitly criticize the behaviorist grounds on which this identification is made, which he deems misleading because it does not describe how the system is internally exchanging energy and information with its milieu (ibid, 50-51). In doing so we lose sight of the fact that organisms are "entirely concrete", material beings that individuate themselves, whereas machines, as artifices, are part of a process of progressive individuation and "concretization" whereby technological entities move from abstract ideas and theoretical principles to the concrete (a point reminiscent of Canguilhem's analysis of machines as organs; [1952] 1992). The technical object begins its life "as the translation into matter of a set of notions and scientific principles that are deeply separate from one another, which are attached only through their consequences and converge for the purpose of the production of a desired effect." (ibid, 49). As such, the "primitive technical object is not a natural, physical system, it is the physical translation of an intellectual system." (ibid). This reflection leads Simondon to emphasize that causal-mechanical processes on their own could never have produced these processes of concretization in the absence of the intentional, purposive thinking of human beings (ibid; see also Ginsborg 2004), while also admitting that the origins of life requires that certain organized structures emerge from nature, and more specifically from a metastable state (ibid, 51). While servomechanisms might seem to be alive, Simondon insists that they are "still artificial" (ibid).

For Simondon, individuation as a problem is completely bracketed by cybernetics which is neither concerned with the initial genesis, nor with the ongoing process of individuation. The notion of the servomechanical automaton, and equally notions of information as a quantity received or exchanged, fall prey to the tendency to assume rather than explain individuated entities. And because the essential purposiveness of life is rooted in the unique, recursively self-maintaining individuating capacities of all organisms, the cybernetic account of purposiveness falls short in explaining teleology.

In these critiques one finds a concern for establishing a clear locus of purposiveness, in the form of the self-maintaining, self-asserting individual, who, as Jonas famously states, its 'being is its doing' (Jonas 1968). But here one begins to see some divergence in his views from those of Canguilhem and Simondon. Though Jonas recognizes the need for the organism to be 'open' to the environment, he also puts great emphasis on the idea of the interiority of the organismal subject. He deems the term 'self' "unavoidable in any description of the most elementary instance of life" and he also takes it to indicate "the emergence, with life as such, of internal identity-and hence, as one with that emergence, its self-isolation as well from all the rest of reality." (1968, 242) Or: organisms "are obviously organized for inwardness, for internal identity, for individuality" (Jonas [1966] 2001, 90). This is a picture of teleology that is radically opposed to any behaviorist stance which does not take seriously the idea of an organism's own perspective or internal point of view. But for Jonas in particular, this problem goes even deeper: for the inwardness of organisms is itself apprehensible in virtue of *our own* inwardness. As he argues, a disembodied observer would not be able to see the 'true meaning' of metabolism as a teleological process. It is only "on the evidence of each one's own organic awareness, the external manifestation of the inwardness of substance. To add the implications: there is no organism without teleology; there is no teleology without inwardness; and: life can be known only by life." (Ibid, 91).

3.2.2 Normativity

The establishment of agency as the necessary condition for the ascription of intrinsic purposes to a system has major implications for the concept of normativity, which is completely missing from the cybernetics account. Jonas makes this case by arguing that cybernetics cannot make sense of goal-failure (see also: Scheffler 1959). He references the example of the purpose tremor discussed by Rosenblueth et al., where a cerebellar patient will, in attempting to carry a glass to their mouth, "execute a series of oscillatory motions of increasing amplitude" in a way "strikingly similar" to undamped negative feedback (1943, 20). The problem with this analogy, for Jonas, is that one can clearly speak of a *failure* in the patient's capacity to drink water, since his motion is motivated by the goal of doing so. It is in reference to this goal that the tremor has normative significance for the patient, and indeed, the reason why we consider him a patient in the first place. This speaks to Ralph Barton Perry's concern, discussed in the first chapter, that veritably purposive behavior requires that we speak not just ends but of means: the patient reaches out *in order* to quench their thirst, and will try to compensate for the tremor as a means toward achieving that goal. Means introduce a normative element in the sense that they can either be conducive or ill-suited to bringing about their ends.

Canguilhem, who emphasized the importance of normativity as a crucial part of teleology, can help us further flesh out the problem. As highlighted by the case of the cerebellar patient, it is not obvious by what standard, on the cybernetic view, one can view their behavior as pathological. One option, which is Canguilhem's primary target in his *Essay*, is the prevailing understanding of normality in terms of a statistical average or an idealized state. In this view, the patient's behavior is pathological simply because it differs from typical behavior of other humans. Against this view, Canguilhem argues that we must understand health and pathology as a basic biological phenomenon, which exist "because there are men who feel sick, not because there are doctors to tell men of their illnesses" (Canguilhem [1966] 1991, 93). This raises the question: what kind of entity is a human being such that it can find itself in states of health and illness? For Canguilhem, the answer lies in the fact that human activity is always already saturated not just with social norms, but biological ones too. A patient's spontaneous reaction to illness "expresses the fundamental fact that life is not indifferent to the conditions in which it is possible...life is in fact a normative activity" (ibid, 126). The behavior of the ill patient is already an attempt to adapt to new circumstances, though with reduced tolerance for inconsistencies in the environment (ibid, 199).

Consider how this notion of ‘adapt’ diverges from the cybernetic notion. For the latter, the dysfunctional servomechanism’s behavior is a result of failed attempts to adapt to its previous overcorrection, which only causes further overcorrections. But for Canguilhem, to fully understand the cerebellar patient’s behavior as being goal-directed requires seeing it in a broader context, namely, of an organism learning to cope with a novel challenge across a variety of contexts, through which it learns new behaviors to compensate for its condition. Such a view of normativity reflects Canguilhem’s conception of all organisms as agents that act on their own behalf, and for whom the environment shows up in a normatively ‘polarized’ way: “living means preference and exclusion” (ibid, 136; 223). Even a creature as simple as an amoeba, for Canguilhem, “recognizes” categories of health (ibid, 198). Though the notion of recognition may be deemed overly cognitive, it points to the basic fact that organisms for Canguilhem (and Jonas) have *perspectives* on the world, relative to their current goals and capacities. The meaning of health, in this view, is the capacity for a flexible, adaptive, and normative relation with the environment (ibid, 200). It is not an objective phenomenon or state, but always a context-sensitive and future-oriented activity in which the organism is “capable of following new norms of life” (ibid, 196-197; 200). Illness is the instauration of a new norm, a new structure—what Simondon might call a shift in metastable equilibrium--when the previous one becomes disrupted (ibid, 194; Simondon 2020, 231-236). Canguilhem argues that the cybernetic conception of teleology cannot account for these rich notions of adaptation and normativity.

3.2.3 Artifacts

One of the important aspects of the CC is its implications for thinking through the nature of artifacts and extrinsic purpose based on an organism-centered account of agency. As Canguilhem reminds us, if analogies truly are two-way streets, then there is no reason why we would not use the organic to illuminate the technological, rather than the other way around. This is precisely the suggestion which he makes in his essay *Machine and Organism*, which he first presented as a lecture in 1947, and remains a key point of reference for understanding Canguilhem's perspective on cybernetics (Canguilhem [1952] 2005, 76; see also: Fiant 2018).[36] His reversal of the organism-artifact analogy begins with a historical claim about machines. Prior to the invention of the simplest automata during the early-modern period, all machines were directly operated by a human. It was self-evident that machines (such as pulleys) were no more than extensions or 'organs' of the human body, which aided people in realizing their goals. It was with the rise of automata—i.e., machines which had an internal store of energy which allowed them to operate independently of direct human input—that it became possible to forget the dependence of the mechanical on the living (ibid, 79-80). In this context it no longer seemed paradoxical to understand organisms via the machine, even though the nature of machines remained fundamentally unchanged. This resulted in a tautology whereby the machine, which is no more than a tool or organ of the human species, was used to explain the organ (ibid, 87).

Canguilhem's reversal gives us a different way to consider the way teleology and mechanism might be reconciled, one which is more radical than what purposive behaviorists attempted. As he argues, to understand the mechanical via the organic leads to a perspective from which it is "hard to see where the opposition between mechanism and finalism lies. No one doubts that a mechanism is needed to ensure the success of a given purpose, and inversely,

[36] Though cybernetics is clearly in the background of these discussions, in the 1965 version of the essay Canguilhem only explicitly mentions it twice. He first claims that self-regulating machines are no more than "machines superposed upon machines by man…The construction of servomechanisms or electronic automatons displaces the relationship of man to machine but does not alter its sense" ([1952] 2005, 88). I was unable to verify if these references are in the 1947 original or if they were introduced in the reprinted 1952 version. The second mention is found in a footnote at the end of the chapter, in which he lucidly describes cybernetics, and its 'bionics' offshoot, which he likely saw as the literal embodiment of his thesis regarding the nature of machines as organs (for a detailed historical account of bionics as an offshoot of cybernetics, see Kline 2015, 165–68).

every mechanism must have a sense, for a mechanism is not just an accidental series of interdependent movements." (ibid). In other words, by showing the continuity of the mechanical with the living, Canguilhem can also show the continuity of teleology (or "finalism" as he called it) with mechanism. Far from being opposed to finalism, machines are highly purposive given that they are engineered to fulfill specific goals and functions determined by their creator.

Canguilhem did not mean to say that organisms and machines are purposive in the same way, however. Canguilhem reiterates the fact that organisms are capable of "self-construction, self-conservation, self-regulation, and self-repair" (ibid, 88) in ways that machines as we know them (including servomechanisms) cannot. For machines to be conserved, there is the need for "periodic intervention" along with close attention and surveillance on behalf of the human. Moreover, Canguilhem insists somewhat enigmatically that in the machine, the "whole is strictly the sum of the parts", whereby effects are dependent on "the order of the causes" (ibid). This amounts to a kind of "functional rigidity", reinforced by practices of standardization, which allow for replaceability and interchangeability of parts (ibid). For Canguilhem, this leads to the somewhat paradoxical conclusion that a machine is more, rather than less purposive than an organism, in the sense that "the purpose of the machine is rigid, univocal, univalent." (ibid, 89). As a result, one kind of machine cannot replace another kind, because they are built for different purposes, and cannot adapt to new circumstances and adopt new functions. This is in stark contrast to organisms in which one witnesses "a vicariousness of functions, a polyvalence of organs" (ibid). For Canguilhem, the living in this sense has "less purpose and more potentialities… Life, by contrast, is experience, that is to say, improvisation, the utilization of occurrences; it is an attempt in all directions." (90). In other words, machines lack precisely the adaptive capacities which Canguilhem used to ground notions of normality and pathology. Consequently, "…there is no machine monster. There is no mechanical pathology …The distinction between the normal and the pathological holds for living beings alone." (ibid).

3.2.4 Teleonomy and the Continental Critique

While the continental critics were largely writing prior to the popularization of teleonomy, they would have seen it fail in the same ways that they believed cybernetics failed as an account of intrinsic purpose, precisely because of the ways in which teleonomy remains continuous with cybernetics in being an artifact model of teleology. As argued in the end of Chapter 2, Mayr's claim regarding the intrinsic purposiveness of the genetic program appear to trade on ambiguities around the notion of genes as active agents. One cannot simply "read off" goals from a sequence of nucleotides, and there is much in development and behavior that is not genetically coded, but rather arises from self-organizing processes in the cell. While Mayr did not know this, his silence when it came to explaining what it means for a goal to be inscribed in a genetic program is telling.

Moreover, it is completely unclear how normativity enters this picture. What would it mean for the genetic program to malfunction? One might think that the extra-genetic apparatus of the cell could fail to properly execute the goals inscribed in the program, though again one would need to specify in what sense those goals are specified. Furthermore, it seems that there is no way to think of a mutation in the program in functional terms. If a change in the code causes a change in the goal-instructions, then that is simply the new goal—even if it leads to disastrous outcomes for the developing organism. Cashing out any intuitions we might have about the normative status of such mutations requires falling back on the metaphorical implications of the program, namely, that it has failed to execute the goals that its author—natural selection—inscribed. And indeed, the idea that function and malfunction statements refer to past histories of natural selection is precisely an idea which became formalized and popularized around the time of Mayr's writing. But taking this route would go against Mayr's attempt to keep teleonomy a proximate causal process, and in the process, undermine the distinction between intrinsic and extrinsic purpose. The program, in this view, would have the function of executing goals set by an "agent" external to it, albeit one that is, necessarily, purely metaphorical. At this stage, teleonomy is really an outright elimination rather than a naturalization or rehabilitation of teleology.

This reveals that the notion of intrinsicality, for Mayr, has been replaced with the far thinner notion of being a cause "inside" the organism. Indeed, this is what distinguishes it from *teleomatic* processes, such as falling rocks, which "are 'end-directed' only in a passive, automatic way, regulated by external forces or conditions." (Mayr 1974, 98). This distinction betrays the fact that Mayr's understanding of intrinsic purposiveness is equivalent to the directing forces being merely being physically inside the goal-directed process in question. Such a thin notion of intrinsicality has no recourse to substantial facts about the kind of entity in question, let alone a richer notion of agency.

To conclude this section, let us review where things in relation to the three core questions that framed the discussion of the previous chapters. The notion of biological agency advanced by the CCs goes a long way in clarifying the third question, namely, how one ought to make sense of the distinction between intrinsic and extrinsic purpose. I have also argued that this properly reframes our understanding of the scope of the organism-artifact analogy. However, this still leaves the first question largely untouched: how are we to make sense of the relationship of causation and teleology? In what sense is the problem of backward causation resolved? Does a natural conception of biological agency necessarily also make teleological explanation a species of causal explanation? In the final section of this chapter, I will unpack these questions by way of an investigation of the historical legacy of the CC.

3.3 Autopoiesis and Teleology

3.3.1 Second-order Cybernetics

Despite the wide-ranging critiques levelled at the cybernetic account—of which the continental critiques constituted only a small part—the cybernetic vision of teleology never underwent any development or subsequent transformations. The reasons for this are likely manifold; beyond the fact that the critics had little contact with the cybernetics group, who seemed more interested in applying these ideas far and wide rather than questioning their basis, the dream of a unified cybernetic research program quickly dissolved after the last Macy Conference, and splintered into a multitude of research projects such as artificial intelligence and artificial life which took its place in the mainstream (Erickson et al. 2013, 2020).

As such, it becomes more difficult to trace the history of cybernetic teleology after the tenth and last Macy conference of 1953. Some members of the Macy Conference group pursued the cybernetic project under the aegis of 'second-order cybernetics' which was inaugurated by Margaret Mead and Heinz Von Foerster in 1968 during the first symposium of the newly formed American Society for Cybernetics (Mead 1968). Prominent figures in the movement included Gregory Bateson, Stafford Beer, Gordon Pask, and to a certain extent Chilean neuroscientists and theoretical biologists Humberto Maturana and Francisco Varela. The core claim in Mead's paper, entitled *The Cybernetics of Cybernetics*, was that the newly formed Society should organize itself according to cybernetic principles (1968). This 'reflexive turn' in the history of cybernetics was in many ways a natural development for a group of thinkers fascinated by causal circularity. It was accompanied by an interest in a slew of epistemological questions concerning the meaning of observation and the construction of reality by the observer. The movement was particularly interested in the possibility that cybernetic systems under study should be treated not merely as *observed* systems, but as *observing* systems which construct and change their maker in turn (Glanville 2002; von Foerster 2003, 283; 303).[37]

There was never an explicitly articulated account of 'second-order cybernetic teleology' equivalent to Rosenblueth et al.'s 1943 account. In fact, second-order cybernetics became enamored with epistemological problems which, while they

---

[37] 'Second-order cybernetics' implies a radical break with its first-order predecessor which many deem inaccurate (Froese and Stewart 2010; Glanville 2002, sec. 3.2; Krippendorff 2008).

certainly had possible consequences for the issue of teleology, were never worked out in a systematic way. That being said, the movement *did* have a significant (though indirect) influence on the work of Maturana and Varela, particularly their notion of autopoiesis which, by the time of Varela's death, would kickstart a resurgent interest in the question of teleology (Weber and Varela 2002).

3.3.2 From Second-order Cybernetics to Autopoiesis

The notion of autopoiesis was first introduced in Maturana and Varela's 1972 book *Autopoiesis and Cognition: The Realization of the Living*. And while the theory is evidently strongly inspired by cybernetics, it also broke with it in several ways.

The continuities can be attributed to Maturana and Varela's close connections to cyberneticians such as Von Foerster, Stafford Beer, and others in the cybernetics group (Keller 2008). This continuity is threefold. First, it can be seen in the insistence that the theory is "mechanistic", by which they meant: i) that their theory was naturalistic, in the very broad sense that "no forces or principles will be adduced which are not found in the physical universe"; and ii) that theirs would be a systems-approach characterized by abstract principles which have multi-scale applicability from single cells to nervous systems, and which was agnostic about the way these systems are materially realized (Maturana and Varela [1972] 1980, 89).[38] Second, the idea that organisms are 'homeostatic machines' indicates a strong continuity with Ashby's work in particular (see also: Froese and Stewart 2010; Bich and Arnellos 2012). Finally, the authors showed a keen interest in the driving questions behind the second-order cybernetic movement concerning the nature of observation and observers, i.e., what it takes to be the kind of entity which has a perspective on the world.

[38] "It is important to emphasize that when they apply this view to biology, Varela and Maturana conceive of the organism holistically as one mechanism, rather than a system made of distinct mechanisms. Given that the organism is characterized as a self-maintaining system whose activity and existence coincide, it follows that in a living system the phenomenon produced by the mechanism is the organism itself" (Bich & Bechtel 2021, 53).

While some of discontinuities paralleled the evolution from first to second-order cybernetics, others represented a break with cybernetics altogether. One of the clearest distinctions with first-order cybernetics was Maturana and Varela's abandonment of the behaviorist stance, instead opting for an 'organizational' one, aimed at elucidating the "relations that define a machine as a unity, and determine the dynamics of interactions and transformations which it may undergo as such a unity".[39] Maturana and Varela's goal was to specify abstract organizational principles from which the characteristic, 'autonomous' behaviors of organisms are derived, and from that clarify the 'ontological grounding' for the problem of observation by specifying the necessary and sufficient conditions for a system to *be* an observer in the first place. This led to one of their most influential theses: "Living systems are cognitive systems, and living as a process is a process of cognition." (1980, 13). Underpinning this was the claim that life and cognition emerge from the same kind of abstract organizational principles, those which define a given system as 'autonomous' or an 'observer'.

They argued that an observer (or 'cognizer') is a system that generates itself and its own boundaries through its own operations, thereby also making it a self-creating 'individual'. One way to characterize this idea in relation to Ashby's approach is as a homeostatic system *whose essential variable is the preservation of its own unity* (*ibid*. 78). While such an idea was implicit in much of Ashby's work, Maturana and Varela's innovation was to take such homeostatic systems to be of a privileged kind which characterized living systems. Therefore, despite taking up the mechanistic framework of cybernetics, they used it to draw the opposite conclusion: that there *is* a fundamental difference between living (autopoietic) machines, and non-living (allopoietic) machines. They defined autopoietic machines as those organized as self-producing networks of processes which:

[39] In contrast to 'structure', which consists of the "actual relations which hold among the components which integrate a concrete machine in a given space." (Maturana & Varela, [1972] 1980, 77).

> (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network.
>
> (ibid., 79)

Though this description was purposefully abstract to be applicable across substrates and scales, they supplied this definition with certain concrete considerations by stating that "the cell is a molecular embodiment of autopoiesis" (*ibid.* 90).

As the criteria quoted above suggest, Maturana and Varela put great emphasis on the organizational principles underpinning the 'autonomy' of living systems, thereby giving a secondary role to the various 'interactive' processes which occur between organism and environment. This attitude was particularly evident in Maturana's recounting of the early development of the theory: "I had to stop looking at living systems as open systems defined in an environment", which led him instead to think in terms of 'closed' systems "that could only be characterized with reference to themselves" (*ibid.* xiii). This marked yet another departure from the original cybernetic account, which put primacy on the system-environment relationship. According to Maturana and Varela, closure was the necessary condition for autonomy; the system could 'specify itself' through its own activity (*ibid.* 92). This highly 'internalist' notion was thought sufficient to derive any of the behaviors of the system with its environment, where any change of state would be wholly determined by the autopoietic organization itself rather than external factors (*ibid*; Varela 1979, 55; see also Godfrey-Smith 1996; Kirchhoff 2018).

3.3.3 Autopoiesis and the Dispensability of Teleonomy

What then of teleology? Maturana and Varela argued that the concept ought to be situated within a distinction between the human "domain of description" (also called "the domain of observation") and the "domain of the system" in question. The former consists in those features which we use to describe the system but "do not pertain to the constitutive organization of the unity" (Maturana and Varela [1972] 1980, 75), while the latter consists in those physical processes and relations through which the system is realized (ibid). The motivation for this distinction was to emphasize their thesis concerning the autonomy of autopoietic systems, namely, that their autonomy and identity is specified *by their own* operations, not by virtue of anything we say about them. The existence of autonomous systems in the universe in turn establishes the very possibility of a domain of observation or description, which is predicated on the existence of such perspective-taking autopoietic systems, including ourselves.

While certainly a provocative thesis, it is unclear how exactly such a distinction gets around the fact that one can never fully exit the domain of our own descriptions; there is no 'view from nowhere' (ibid, v). At any rate, it led to an extreme caution on the authors part not to use any language that may be even vaguely metaphorical, including concepts like information and, important for this present discussion, concepts of function and teleonomy. For these reasons, they that such concepts are "dispensable", given that "the relations implied in the notion of function [and teleonomy] are not *constitutive of the organization of an autopoietic system*, they cannot be used to *explain* its operation." (ibid, 78; emphasis added). They go on to claim that "the notions of purpose and function *have no explanatory value* in the phenomenological domain which they pretend to illuminate, because they do not refer to processes indeed operating in the generation of any of its phenomena…if living systems are physical autopoietic machines, teleonomy becomes only an artifice of their description which does not reveal any feature of their organization" (ibid, 86). Unsurprisingly, this argument follows the same broad outlines of those discussed over the course of the last two chapters: explanations advert to causal goings-on and are complete—so explanations in terms of goals are either redundant or must find their place in another "method of study" or are somehow only descriptive rather than explanatory.

As I see it, this argument points to a crucially important fact that is easily overlooked. Autopoiesis represents a theory of biological organization which seems to fulfill many of the criteria which the continental critics used to define agency, particularly Jonas'; autopoietic systems are self-producing, self-maintaining entities that actively keep themselves alive; in so doing, they have a stake in their own existence. Moreover, one might think that an autonomous system's perspective (which, as Maturana and Varela claimed, emerges from its organization) consists in seeing the world in a normatively polarized way: some things are good for the realization of its existence, and some are bad. And yet, teleonomy is dispensable. Why? On the one hand, it might just be a by-product of their division between the "domain of the system" and "the domain of description". But no such radical epistemological stance is needed to come to this conclusion. Even without it, it would be entirely consistent for them to both affirm that living beings have a particular autonomous mode of organization distinct from that of machines *and* that talk of purpose and function is dispensable. Afterall, it seems that one can exhaustively describe the dynamics of the system in causal mechanical terms. The same case could be made of Mayr's teleonomy: if one can give a complete causal explanation of genetic codes—both in their origination by natural selection and operation in ontogeny—then one can dispense with talk of teleonomic talk entirely. Why would one explain the behavior using goals or norms when an equivalent causal explanation will do? What Maturana and Varela have shown, then, is that *making sense of the distinctiveness of biological organization does not ipso-facto provide a naturalization of teleology*.

One possible response to this is to rethink scientific naturalism or dispense with it entirely (A. Cooper 2018; Kee 2018; Macarthur 2019; Rouse 2023). Interestingly, this was the path chosen by Varela later in his life when he encountered the work of Hans Jonas, which led him to revise his initially anti-teleological stance. In a paper written with Andreas Weber and published a year after Varela's death, the authors noted the remarkable convergence of the notion of autopoiesis and Jonas' arguments concerning the teleology of living beings (Weber and Varela 2002, 97). The outcome of this convergence between Jonas and autopoiesis was not a shift in the concrete formulation of autopoiesis, but rather a shift in the epistemological and metaphysical assumptions underpinning the theory that came from endorsing Jonas' phenomenological argument for the existence of teleology. In a nutshell, our first-person experience provides undeniable evidence of our own intrinsic purpose as agents, and by extension the reality of agency and purpose in nature. It is through this experience that we recognize that living things (or in Varela's view, autopoietic systems) are also kindred agents with an intrinsic purpose orientated toward their persistence.

Today, such a position is difficult to square even with the most expanded forms of scientific naturalism (Villalobos and Ward 2016; Gambarotto 2020; Sachs 2023). For Jonas, the existence of purposiveness in the universe is a matter of phenomenological certainty; it is not to be explained in terms of something else. But those who have tried to integrate phenomenology into a scientific research program have found that, in practice, "it is just as much abstract, objectifying (neglecting the subjectivity both of the scientist and the target of modelling), and physicalistic (since it borrows its mathematical framework from physics) as the research criticized by Canguilhem and Merleau-Ponty" (Mazviiita Chirimuuta 2020; see also: Feiten 2020; Meyer and Brancazio 2022; Nahas and Sachs 2023). Similar worries about the compatibility of Jonas' thought with contemporary science have also been expressed (Villalobos and Ward 2016; De Jesus 2016; Kee 2018; Sachs 2023; Gambarotto 2020). And if Jonas' phenomenological grounding of intrinsic purpose is a necessary condition for autopoiesis to offer "a naturalized, biological account of Kant's notion of a natural purpose" (E. Thompson 2007, 140), then the success of such a project remains unclear at best (Wolfe 2015).

Such concerns make the Jonasian solution to the apparent dispensability of teleology unattractive. Thankfully, while it may be true that teleology has no place other than as a projection or heuristic within a highly reductive (and somewhat outdated) scientific naturalism, one need not go as far as Jonas in rejecting scientific naturalism all together (Gambarotto and Nahas 2023). Rather, more work must be done to consider the relationship between biological agency and teleological explanations. This will be the focus of Part 2 of the dissertation.

3.4 Conclusion

Despite the differences in their view, one clear theme we find across Jonas, Canguilhem, and Simondon's critiques of cybernetics is the idea of the individual or self as a key, irreducible component of understanding organisms as purposive beings, both as the originating locus of these purposes and the direct beneficiary of their accomplishment. It is precisely this understanding of the organism which they saw as lacking in the cybernetic account of teleology as negative feedback, and life as a complex servomechanism. What this gives us is a distinction between intrinsic and extrinsic purpose where the latter is always secondary and ontologically derivative. This highlights the logical impossibility underlying the artifact model of teleology, which at best attempts to account for intrinsic purpose *from* external teleology or dispenses with the distinction altogether and is no naturalization of teleology at all. That being said, the CC was also lacking in some respects. It showed the importance of taking biological agency seriously, and in so doing advanced two of the three key issues that have framed the discussion so far, namely, the organism-artifact analogy and the intrinsic/extrinsic distinction. However, it also left the issue of causation and teleological explanation unaddressed. Worse still, as Maturana and Varela showed, taking the distinctive character of biological organization seriously does not necessarily mean that talk of purpose, function, and norms is indispensable. It therefore remains to be seen whether an account of teleology based on biological agency can succeed where artifact models fail.

II

BEYOND THE ARTIFACT ANALOGY

Part 1 of this dissertation traced part of the intellectual lineage of the artifact model of teleology and showed its shortcomings. In Chapter 1, I argued that the cybernetic account of teleology is continuous with a philosophical tradition which struggled to make sense of purpose purely in behavioral terms. These difficulties are evident in debates over whether goals should be characterized as internal states inferred from behavior, or physical entities—so-called "goal objects"—in the environment. What the continental critique suggests is that the problem lies in their assiduous attempts to avoid characterizing the behaving entity in question. The 'black box' ontology of cybernetics, upon which servomechanisms could be seen as adequate models for purposive behavior, was a natural continuation of this attitude. It was also precisely this bracketing of the agent as the locus of purposiveness that was at the heart of Pittendrigh and Mayr's teleonomy discussed in Chapter 2. For both, teleology is a property shared by artifacts and organisms. For Pittendrigh, it has to do with 'organization'; for Mayr, with being 'programmed' by natural selection. In sum, the purposive behaviorist, cybernetic, and teleonomic accounts variously used artifactual models, analogies, and metaphors to account for teleology. I argued that this approach is doomed to fail because all an artifactual model can provide is, by definition, an account of *extrinsic purpose*. In the last chapter, I integrated and made explicit the views of various continental critics of cybernetics to argue that they correctly identified this as a general and fundamental shortcoming in all artifactual accounts of teleology. What one needs instead is an account of biological agents, to whom we can ascribe goals and any normative facts they entail, to act as a ground for intrinsic purpose.

While biological agency is a definitive step in the right direction, it remains to be seen how such an account can make sense of the relationship between causal and teleological explanations. At the heart of this apparent conflict is the problem of backward causation: if teleology is to be a veritable mode of explanation that cites goals as the explanans, and to explain a phenomenon is to cite its causes, then goals need to be causes. The proposed solutions to this problem surveyed in Chapters 1 and 2 have encountered a dilemma. They either: i) fail to distinguish teleological explanations adequately from causal explanations, and thereby fail to capture what is distinctive about teleology, or ii) fail to show in what sense teleology is genuinely explanatory. Allport and Holt's conceptualization of teleonomy in terms of laws of behavior and Mayr's in terms of the genetic program falls to the first horn. In both cases, it is not clear how teleology is any different from other kinds of nomological explanations in physics and causal explanations respectively. The cybernetic account of purpose and Pittendrigh's teleonomy fall to the second, since both claim that purpose or teleonomy is non-causal and merely descriptive rather than genuinely explanatory. It now remains to be seen how this dilemma might be overcome, and the role that biological agency ought to play. As I interpret Maturana and Varela's argument regarding the "dispensability of teleonomy", to make sense of the organizational principles that distinguish organisms from artifacts is not sufficient for motivating the use of teleology as a mode of explanation.

Today, it is widely believed that true progress on the issue was only made when philosophers of science, tackling the issue in the broader context of scientific explanation, showed that teleology can be made both explanatory and distinctive as a historical mode of explanation (Ruse 1971; Wimsatt 1972; Wright 1973; Boorse 1976; Brandon 1981). Wright provided the most famous formulation of this when, in a challenge to the dominant deductive-nomological model of explanation, he proposed a more liberal form of explanation which he dubbed the *etiological* model. Such explanations cite factors that bring about (or make more likely) the explanandum, i.e., cite an *etiology* (Wright 1972; Godfrey-Smith 1993). On this basis, he characterized and distinguished causal from teleological explanations, where the former explain what brought about the explanandum by citing preceding factors, whereas the latter explains by citing consequences. On this basis, Wright argued that artifacts and biological traits not only exist for a particular purpose, but that *their purpose explains why they exist*. In this case, the existence of the phenomenon in question is explained by a consequence it brings about (1972, 205).

As we've seen, one of the major challenges of teleology is to make sense of how this could be without invoking the specter of backward causation. Around the time of Wright's writing, a consensus began to emerge that functional statements in evolutionary biology were in fact both admissible and teleological because they explain in virtue of a *past* consequence (Shea 2018, 59). In this view, "hearts are for pumping blood" is equivalent to "hearts exist because they pump blood"; and the latter is an explanation that adverts to consequences that occurred in the evolutionary past. In this respect, teleology qua consequence etiology went directly against Mayr's position on teleonomy, which attempted to distinguish its ultimate and proximate senses to then solely endorse the former. This led Mayr to claim that etiological questions concerning the existence of a trait are *only* questions about how a trait came to be, *not* about what it is for (1961, p. 1502). Wright's etiological conception of teleological explanation sunders this distinction.

That functional explanations can be both legitimately teleological and naturalistic by adverting to *past* consequences further illustrates the continuity between the way teleology was treated by behaviorist psychologist and evolutionary biologists. In fact, a behaviorist conception of teleology that associates purpose with learning and an evolutionary conception that associates function and purpose with adaptation share an underlying logic. Just as learned, adapted behavior can come about through a process distinct from its genesis (random trial and error), so too can adaptations emerge through randomly occurring heritable variation (Scheffler 1959, 270; Wimsatt 1972; Millikan 1984; N. S. Thompson 1987). Moreover, both are after-the-fact processes—hence Thorndike's appropriately named "Law of *Effect*" (1898)—in which the occurrence of a learned behavior or adaptation is explained by a consequence it brought about in the past (Deacon 2011, 111).

Perhaps the most important upshot of this way of framing teleological explanation is that it renders teleology perfectly compatible with causal explanations, given that it casts consequences as causes. In the next chapters, I intend to challenge this proposed solution. Before doing so, however, I will suggest a way in which the discussion on teleological explanation might be usefully framed.

Content and Conditions

As I see it, Wright was correct to characterize a teleological explanation is one in which a state of affairs is explained in virtue of a consequence, in the form of a goal, that it achieves. This is both what renders such explanations distinctive and problematic. And yet, this mode of explanation is familiar to us from our day to day lives, in which we are constantly explaining our own behavior and that of others based on goals. What exactly are we doing when we do this? What kind of understanding are we conveying about the world around us? Considering these questions reveals that the content of a teleological explanation contains far more than information regarding the dependence of a phenomenon on a consequence. By 'content' I mean the information contained in the structure of teleological reasoning and explanation, and what it tells us about the phenomenon it's purporting to explain (Walsh 2014a). The aim of naturalizing teleology, broadly construed, is to make sense of this content within a scientific worldview.[40]

---

[40] That being said, there remains substantive yet implicit disagreement on what is at stake in this project (Nahas and Sachs 2023).

The most obvious thing teleological explanations tell us is why a particular course of action was taken *given* the available possibilities. The answer is that the actions were an appropriate means toward pursuing the goal in that context. It follows that had the same goal been pursued under different conditions, the means would have differed systematically according to their appropriateness in the new context (Walsh 2012; 2015; 2018; Fulda 2017). The notion of appropriateness carries with it normative implications: the actions were selected by the agent in virtue of the fact that they were a *good* way to do so, given the circumstances afforded to the agent, and possibilities of action available to them (ibidem). Appropriateness is also to be evaluated according to the extent to which the actions cohere together. Given that the appropriateness of actions is a path-dependent evaluation, in the sense that the actions which came before constrain and enable the space of possible appropriate actions in the future, explaining the behavior teleologically tells us about the ways the various actions are unified toward the attainment of the goal. This unification produces a particular kind of understanding of the behavior in question.

Important to this analysis is the modal profile of teleological explanations, i.e., what they reveal about what didn't happen but could have. This point is colorfully illustrated by William James, who compares the path taken by Romeo to reach Juliet in comparison to the path taken by iron fillings to a magnet: "With the filings the path is fixed; whether it reaches the end depends on accidents. With the lover it is the end which is fixed, the path may be modified indefinitely" (James [1890] 1983, 20). While the idea that the lover's end is "fixed" is somewhat of an idealization—as we know, agents have the capacity to not only persist in their goals but also to flexibly change them, or reorder which goal takes priority based on the changing context—the fact that Romeo's "path may be modified indefinitely" is accurate. It speaks to the fact that we treat agents in explanations as what Jason Winning calls "open-ended" controllers of their behaviors who flexibly select appropriate actions *because* they are appropriate, and who will modify them in the event that it becomes appropriate to do so in changing circumstances (Winning 2023). Given this capacity, a teleological explanation of

Romeo's behavior implies the numerous paths that he would have been able to take but didn't; or how he might have acted if he'd loved Tybalt rather than Juliet, or if Juliet lived in different part of the city. In contrast, the behavior of "closed-ended" systems such as the iron fillings is explained deterministically by the initial conditions and the laws of physics. Putting this in terms of Wright's contrast between teleological and causal explanations, we could say that while the latter is highly sensitive to initial conditions, the former is relatively insensitive to the causal-mechanical details of the situation; a description of Romeo's behavior in terms of goal-directedness may be accurate and robustly predictive *despite* the fact that it is entirely agnostic about the causal-mechanical details of the situation (Fulda 2017; Walsh 2018). In this way, the precise molecular arrangements of Romeo, Juliet, and the wall in space, including the surrounding context of the city of Verona, is irrelevant to the explanation.

All this is part and parcel of the *content* of a teleological explanation. To take seriously the content of everyday teleological explanations is to take on board the idea that we have a substantial pre-theoretic grasp of how to effectively explain the behavior of agents. But few would follow Jonas in thinking that our phenomenological evidence of agency is necessary and sufficient to deem it real, or fully capture what is right about it, in large part because this fails to meet standards of naturalism. The problem is how to make sense of this content *naturalistically*, or more precisely, how we make sense of what the world must be like, given the constraints of our best science, for this content to be an explanatorily adequate account of it. In the words of Ralph B. Perry, "'Purpose,' 'means and end,' 'in order to,' 'for the sake of,' 'with a view to'—these and many other kindred forms of speech are evidently applicable to the same context. *There is something in our world to which they serve to call attention. What is it*?" (1917, 352; emphasis added). As Perry began to recognize, and the continental critics made even more explicit, the answer to this question is the agency of living beings. If this is right, then an account of the structure of *teleological explanations* must be grounded by an account of agents as *intrinsically purposive systems*. Intrinsically purposive systems are the *condition* which is presupposed by the explanatory *content* of teleology. The distinction between content and conditions addresses a persistent equivocation in the philosophical literature between teleology as a phenomenon to be explained and teleology as a mode of explanation (Toepfer 2011; Dresow and Love 2023; Bich 2024; see also: Gambarotto and Mossio 2022).

So far, I have tried to use the term 'teleology' to refer to a mode of explanation and 'purposiveness' to refer to the phenomenon, though the inconsistent usage of these terms by authors discussed has made this difficult. The terminological distinction will be adhered to strictly from now on, because the content/conditions distinction will be central to Part 2 of this dissertation.[41]

Before moving on to my positive contribution in the forthcoming chapters of this dissertation, it is worth considering the usefulness of the content/conditions distinction in framing the issues discussed in the previous chapters. Consider Perry's argument, discussed in Chapter 1, that adjustment behaviors which maintain equilibrium between system and environment do not require an explanation in terms of means and goals. This is because the response of a servomechanism "*is* complementary; but it means nothing to say that it occurs *in order to* be complementary, or *for the sake of* the complementary outcome." (Perry 1918, 1). And because an appeal to means being for the sake of an end is part and parcel of the content of a teleological explanation, Perry is effectively stating that a characterization of the phenomenon of intrinsic purpose in terms of a maintenance of equilibrium does not provide the appropriate conditions for the application of such an explanation. For similar reasons, Richard Taylor critiqued the cybernetic account on the grounds that it "endow[s] the word 'purpose' with a meaning having no similarity to any meaning it has customarily been taken to possess." (Taylor 1950a, 317). In other words, it distorts the *content* of teleological explanations. Similarly, Mayr's invocation of the genetic program appears to be

[41] The distinction between conditions and content also emerges from the recognition, found in the work of many analytical philosophers such as Larry Wright (1968; 1976) or Ernest Nagel (1977), that to unpack what we are saying when we explain something teleologically is a different exercise than making sense of what the world must be like for such explanations to apply. For one thing, teleological explanations may be appropriate to explain many phenomena which are related but distinct from intrinsic purposiveness.

an attempt at providing the conditions under which biological processes can be explained teleologically. Yet he fails to explain why meeting those conditions makes it appropriate to offer a genuine teleological explanation. The continental critique, in turn, shows why these various attempts fail: the necessary conditions for a full teleological explanation to apply, with all its normative content, is organismal agency. At the same time, the continental critique failed in providing an adequate characterization of the content of a teleological explanation.

But on the face of it, the fact that our world contains agents capable of acting in open-ended and contextually appropriate ways based on goals is a veritable "ontological surprise" (Jonas [1966] 2001, 79). There is therefore a story to be told about what the world must be like for such systems to have emerged in the first place (Deacon 2011; 2018). Framed in terms of content and conditions, the problem that remains is: i) whether these conditions can be made sense under an adequately naturalistic framework, and if they can, ii) whether articulating these conditions is sufficient for making sense of the content of teleological explanations. Indeed, one might think that a causal explanation of the organizational principles underlying agents is ultimately what should fix the content of teleological explanations; and any gaps that currently exist between content and conditions will eventually be filled as we gain a better understanding of what produces agency. In the following chapters, I will argue that there are reasons to think that the content of teleological explanations cannot be reduced to its conditions.

In what follows, I will evaluate the prospects for a contemporary account of intrinsic purpose that moves beyond the artifact model. Following the continental critics, the first step in this project is to treat biological agency as the necessary condition for teleological explanations. But conceptualizing intrinsic purpose in this way is a theoretically arduous task which may be attempted in different ways. Chapter 4 will illustrate this fact by putting two contemporary approaches to naturalizing teleology, called the Autonomy Theory (AT) and the Ecological Approach (EA), side to side. I will show how their different ways of conceptualizing agency as a condition for teleological explanation impacts the ways they conceive of the content of such explanations. Since the publication of Weber and Varela's (2002) paper, the AT has carried on this research program, attempting to—among other things—ground an account of teleology on the distinctive organization of the living, and in doing so, bringing the continental critic's vision to fruition (Sachs 2023). The AT aims to derive the content of teleological explanations from a theory of the minimal conditions required for biological organization, which in their view grounds intrinsic purpose in this 'causal regime'. In this regard, it treats teleological explanations as ones in which consequences explain *qua* causes. In contrast, the EA characterizes agency in terms of the gross dynamics of biological systems and does so by helping itself to the concepts it deems required to make sense of behavior. In turn, it uses those very concepts in its formulation of the content of teleological explanations, where consequences explain *qua* goals rather than causes. Each strategy makes a set of tradeoffs. While the AT explains rather than assumes the emergence of agency, the content of its teleological explanations is highly restricted; it does not easily generalize beyond their account of the minimal conditions for intrinsic purpose. In contrast, the EA assumes concepts that the AT aims to explain, but in doing so arrives at a conception of teleological explanation whose content is more broadly applicable and ostensibly truer to the content of such explanations as they are used in everyday speech.

In Chapter 5, I will build on the EA's framework by arguing that it needs to expand its characterization of intrinsic purposiveness beyond behavior, to what I will call *way of life*. This distinction provides a two-fold conception of purpose as *goals* and *ends* respectively. I will then show how taking teleology as a form of explanation in which phenomena are explained by consequences *qua* goals provides a novel account of biological and artifact function. I will show how my account of function makes sense of the ascription of purposes to artifacts in a way which shows why artifact models of teleology will always fail.

In Chapter 6, I will argue that function ascriptions ought to be located within goal-explanations. Doing so breathes new life into the so-called 'goal-contribution account' of functions, which is typically thought to fail because it cannot distinguish between proper functions and accidental functions. I will show why the Ecological account of goal-explanations resolves this problem. The upshot of this discussion is the way that it illustrates the priority of teleological explanations over function attribution. And since teleological explanations depend on agency as their necessary condition, this shows the artifact model of teleology to be completely wrong-headed: artifact function is to be derived from biological agency, not the other way around.

CHAPTER FOUR

Two Approaches to Teleological Explanation

In this chapter I critically evaluate two recent attempts to make sense of teleological explanation based on an account of organisms as intrinsically purposive agents, which go by the Autonomy Theory and Ecological Approach respectively.[42] The AT is a recently developed theoretical framework which seeks to elucidate the organizational principles underpinning the structure of organisms. The articulation of such principles is thought to have broad theoretical and philosophical implications for biology. Among other things, supporters of the account take themselves to be vindicating the idea, defended by Hans Jonas among many others, that the intrinsic purposiveness of organisms is grounded in their precarious self-maintenance, or the fact that their 'being' is their own 'doing'. In this respect, the AT takes after and builds upon many of the insights from the autopoietic tradition, as well as the work of various scholars who have theorized biological organization (e.g., Piaget 1967; Rosen 1991; Kauffman 2002; Toepfer 2023). The EA takes a distinct approach which, partly through the influence of James J. Gibson, is more closely aligned with the American pragmatist and

[42] While a number of other frameworks have been proposed, including enactivism (E. Thompson 2007; Barandiaran 2017), teleodynamics (Deacon 2011; 2020; García-Valdecasas 2021), and the field theory (McShea 2012; Babcock and McShea 2023), a more comprehensive evaluation of these theories is beyond the scope of this work. I choose to focus on the EA and AT here because their contrast clearly brings to the fore the underlying philosophical problems I am concerned with.

purposive behaviorist approach discussed in the first chapter of this dissertation.[43] According to this approach, the agency of organisms is characterized by the gross, observable dynamics of their behavior, characterized in terms of goals, affordances, and a behavioral repertoire.

The way the AT and EA respectively characterize intrinsic purpose goes some way in explaining the different ways in which they articulate the content of teleological explanations. The AT's characterization of intrinsic purpose in terms of the constant and precarious self-maintenance is thought to provide justification for the idea that a teleological explanation is one in which "the existence of an entity (such as an organism, trait, or behavior), [is explained] in terms of some effect the entity tends to bring about." (Garson 2016, 18).[44] What is distinctive about this mode of explanation is that it involves a kind of (non-vicious) circularity: if an organism is in some sense 'cause and effect of itself'—a formulation which mirrors Jonas' claim that organisms are entities whose 'being is their own doing'—then its existence can be explained by what it does. The AT takes this circularity to characterize teleological explanations as a particular species of causal explanation. A causal explanation is one in which causes explain effects; in teleological explanations, consequences are (also) explanatory, and moreover explain a particular fact about the entity in question: its existence as a persisting entity over time. In contrast, the EA characterizes teleology as a form of *non-causal* explanation that tracks modal regularities between goals and means. As a result, this definition is agnostic with respect to the particular goals that a given organism may pursue.

This chapter aims to compare the EA and AT to bring to light what I take to be the major philosophical issue underlying naturalizing teleology, namely, the standards by which one might evaluate the success of such a project. Particular

---

[43] This connection will be discussed in more detail in the subsequent chapter.

[44] Given that this is a fairly common articulation of what teleology is about—encompassing the so-called 'etiological account of functions'—my analysis extends beyond the AT. I nonetheless choose to focus on the AT as I consider it the best account within this broader family of views.

attention will be paid to the way each approach makes sense of the way such explanations are distinguished from (regular) causal explanations, in virtue of the fact that it treats consequences as the explanans. In this regard, I will argue that each approach involves particular tradeoffs. Whereas the AT attempts to make sense of consequence-based explanation by construing consequences (goals) as a special kind of cause, the EA argues that goals can figure in genuine teleological explanations only insofar as they are not causes, but rather a legitimately different kind of explanans. As a result, the AT's conception of teleology conforms to the standards of mainstream scientific naturalism. But it is also highly restricted, which has problematic consequences when applied to *actions, artifact functions*, and *normativity* respectively. In each case, there is a need to define existence in a way that makes its role as explanandum unproblematic; but doing so leads to strange results. In contrast, the EA's conception of teleology does a better job of making sense of the content of our pre-theoretic use of such explanations, but in doing so it may underspecify the conditions under which it applies.

The structure of this chapter is as follows. In the next section, I characterize the AT's account of intrinsic purpose, and its conception of the content of teleological explanations. Section 4.2 provides the same analysis of the EA. Section 4.3 will highlight and clarify the contrast between the approaches in terms of whether teleology can or should be a form of causal explanation. The final two sections examine the tradeoffs which each approach needs to contend with. In 4.4, I show that while the AT's conception of teleology is broadly causal, it does not easily extend to actions and artifacts. In 4.5, I argue that while the EA's gains in the breadth and distinctiveness of its account of teleology, it also must contend with issues concerning the status of non-causal explanations.

4.1 The Autonomy Theory: Teleology as a Causal-Consequence Etiology

4.1.1 An Organizational Account of Intrinsic Purpose

The AT's characterization of organisms as intrinsically purposive beings follows on from a tradition, often claimed to have begun with Immanuel Kant's third *Critique*, in which he argued that biological organization is distinctively self-organizing (Gambarotto and Mossio 2022; Toepfer 2023).[45] The final section of Chapter 3 described how autopoiesis formalized this in terms of *closure*. The AT attempts to bring together insights from these and numerous other scholars while anchoring their theory in a thermodynamic perspective (see also: Deacon and García-Valdecasas 2023a).

Given that organisms are far-from-equilibrium systems, they can only maintain this state by harvesting matter and energy from their environment. It is only by virtue of this activity that they can exist, and so biological organization must be "self-organizing" in the sense first proposed by Kant: it must be "reciprocally cause and effect of itself" (Kant [1790] 2000, sec. 5:373). The AT details the theoretical requirements for such an organization in terms of *closure of constraints*. Constraints highlight the connection between an entity's organization and its capacity to effect change in the world. To illustrate, a car engine's structure can be described in terms of limitations, or *constraints*, imposed on the ways that the parts move and interact. It is through the constrained ignition of gas in the engine, which pushes a piston, that the engine can harness the energy in the fuel to do thermodynamic work.

Highly organized or constrained structures, such as that of an engine, are highly improbable phenomena in the universe, because there are so many more ways for something to be disordered than to be ordered. Kant famously

[45] The extent to which the AT is veritably "Kantian" is an open question (Gambarotto and Nahas 2022; Wolfe 2023).

articulated this idea by asking us to imagine our surprise if we were to find a regular hexagon drawn in the sand of a deserted beach (ibid, sec. 5:370): "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" (ibid, sec. 5:360). For this reason, the existence of artifacts such as the engine require an explanation in terms of the intentions of a designer. But the same can only be true of organisms if we accept that they were designed by God, which is not admissible in natural science. How then could biological organization emerge spontaneously, given its improbability?

Today, the improbability of life's emergence is usually framed in terms of the second law of thermodynamics, which began to be formalized in the decades after the publication of Kant's Third Critique (Chang 2017). The law states that the entropy, which is the measure of the disorder of a system, always (or almost always) increases in a closed system. Any local increase of order requires an equal or greater increase of disorder somewhere else. This is because creating order requires channeling energy to do useful work, but this conversion of energy to work is never perfect; it always involves losing energy as heat. We now recognize that creating order, or constraints on possible states, requires thermodynamic work: channeling energy in useful ways to make things happen that wouldn't otherwise (Kauffman 2000). This means that constraints are both products and enablers of work, which provides the clue to the AT's formulation of biological organization: a closed network of constraints that channel energy so as to create or maintain the other constraints (Montévil and Mossio 2015; Moreno and Mossio 2015). More precisely, the authors define a constraint as an entity which acts on a process while being unaffected by it at the relevant time scale at which the process occurs (Montevil and Mossio 2015, p. 182). Closure, in turn, is defined in the following terms:

> A set of constraints C realises overall closure if, for each constraint C_i belonging to C:

1. C_i depends directly on at least one other constraint belonging to C (C_i is dependent).
2. There is at least one other constraint C_j belonging to C which depends on C_i (C_i is generative).

(ibid, p.186)

But while this network of constraints must be "closed" (in a variant of the logico-mathematical sense) to maintain itself, the useful work done by each constraint requires harvesting available energy from the environment; what the AT calls "thermodynamic openness" (ibid p. 180). As such, biological organization must be thermodynamically *open* at the same time as it is constraint-*closed*.

The self-maintenance of biological order in terms of closure of constraints is, according to the OA, the key to making sense of intrinsic purposiveness. Because a constraint-closed system is structured to actively resist dissipation, one can legitimately say that the goal of any such system is to keep itself going, to keep existing. More precisely, "because of the dependence between its own existence and the effects of its activity, biological organisation is legitimately and meaningfully teleological" (Mossio and Bich 2017, 1090).

While the AT takes itself to be giving an account of intrinsic purposiveness, it is also intent on distinguishing this from extrinsic purpose, which is crucial for its account of biological functions. Given that biological organization consists in a closed "network of mutually dependent entities, each of them making different yet complementary contributions to the self-determination of the system", these subsidiary entities can be said to have a function based on the role they play in this self-determination or self-maintenance (Mossio and Bich 2017, 1105; Mossio et al. 2009). Those effects which the parts have in contributing to the maintenance of the whole are its function, as opposed to other effects which are by-products. In this sense the heart is for pumping blood rather than making thumping sounds because only the former helps maintain the system as a whole; and as discussed in the previous section, we can call this subset of its effects its purpose or function because it is that activity which tells us why the heart exists.

Because we need to appeal to this broader organizational context to specify the function of the heart, its purpose is extrinsic rather than intrinsic. More precisely, in the AT's view, the distinction between intrinsic and extrinsic purpose is cashed out "topologically" (to borrow a term used by Robert Rosen and Maturana and Varela). The purposiveness of life is *intrinsic* to the closed organizational regime because one can trace a chain of direct dependence between constraints back to where one started. The causal chain that we have thereby traced constitutes a *whole*, in virtue of being a closed network of "efficient causes" (to use Rosen's phrasing; see Mossio and Montevil 2015, 180). In this case, one needs to appeal to the organization of the system to explain how its activity explains its existence. The opposite is true of extrinsic purpose, where one must go beyond the system in question—say, a heart—to the whole system it is part of—a body—to explain how its activity explains why it exists. In this sense, functional parts of organisms can be said to be self-maintaining, i.e., cause and effect of themselves, but only *indirectly*, by contributing to a larger causal chain they are part of, and in virtue of which they exist (Toepfer 2008).

It is worth dwelling on what is distinctive about this view. In Kant's third *Critique*, he offers a similar conception of 'natural' intrinsic purposiveness as self-organization. According to some scholars, this account of purposiveness is important as it does not call for explanation in terms of intentions, which relates means and goals, or actions and intentions, and would necessarily lead to the idea of a designer outside of nature (Huneman 2014, 188; Illetterati 2014). The self-organization of organisms is an account of "purposiveness without an end", where 'end' refers to the intentional model of teleology (later I will use the term 'ends' in a particular sense different from this one).[46]

[46] Though whether this is veritably an account of teleology is unclear. As Angela Breitenbach and others have argued, Kant has no room "for a notion of purposiveness that is entirely divorced from the end-setting intellect." (Breitenbach 2016, 406). Similarly, Paul Guyer claims that Kant's argument suggests that "we must conceive of the purposiveness of organisms as an exception to the ordinary laws of nature because we can only conceive of purposiveness itself on the model of our own free purposiveness" (Guyer 2001, 263). But here too one must consider whether biological

Furthermore, an upshot of rendering the intrinsic/extrinsic purpose in terms of parts and wholes is that it ties intrinsic purpose to a circular causal regime that defines its own boundary, and in doing so grounds the notion of an individual to whom we can ascribe intrinsic goals (Mossio and Moreno 2015, xxix; 101). In this sense, supporters of the AT have a basis for claiming that a closed system's "being is its doing" (in a causal sense), and that it has a very minimal sense of identity. Moreover, an organizationally closed system which includes a particular set of constraints that have the function of regulating interactions with the environment would have the capacity to discriminate among relevant features in it (ibid 99-101), in the sense that those constraints will have the capacity to modulate the internal organization of the system in response to particular features or perturbations in the environment (Bich et al. 2016). This gives the closed system some adaptive capacities, and on this basis provides a minimal criterion for a system to have something like a Jonasian "perspective", albeit not in a phenomenological sense. Indeed, Mossio and Moreno are careful not to follow enactivism in couching these phenomena in phenomenological terms. Despite it being intuitively plausible that regulated interactions have a significance or valence for the organism, leading to a "world full of significance" for the system, such anthropomorphic terminology is scientifically inadequate and superfluous (ibid 101, fn. 5). As Carl Sachs argues (2023), the AT effectively takes the stance that Jonas' critique of cybernetics does not require moving beyond the bounds of an adequately scientific naturalism.

Now that the basic outline of the AT is in place, the next section will show how constraint closure provides the foundation for a particular way of thinking about teleological explanations.

---

purposiveness *qua* phenomenon and teleology *qua* mode of explanation are being conflated. It may be that Kant's account of the former is non-intentional, whereas his account of the latter is.

4.1.2 Explaining Existence with Causes

As discussed in Chapter 2, Pittendrigh's conception of teleonomy has the concept of organization at its heart. For him, "the objects of biological analysis are organizations […] and, as such, are end-directed." (Mayr 1974, 115). Such organization calls for a special kind of explanation which mere order does not: "the origin of that information which underlies and causes the organization" (1958, 396). I claimed that this connection between organization and teleology is never fully worked out, and I suggested that the most plausible interpretation of Pittendrigh's claim is that such an account effectively works backward from a particular way of thinking about artifactual function, as grounded in the intention of the designer: watches are for telling the time because they are designed for that purpose, and hearts are for pumping blood because they are designed for that purpose, albeit by the 'blind watchmaker' of natural selection.

The connection that Pittendrigh sees between organization and teleology (qua mode of explanation) is nothing new (Mossio and Bich 2017, 1092-1099). Indeed, since at least the 18th century, the problem concerning the apparent purposiveness of life has been framed thus: organisms, like artifacts, display a high degree of organization that one cannot account for in terms of the powers of matter alone. In the words of Kant, the "contingency of their form with respect to all empirical laws of nature" ([1790] 2000, sec. 5:377-378) renders organisms and artifacts *mechanistically inexplicable*, and leads us to consider them possible only as products of design. 'Mechanism' in this case means a particular kind of explanation through decomposition where the activity of the whole is explained reductively in terms of the powers of its parts.[47] Since organization is inexplicable in these terms, one must appeal to a different kind of causality, which—depending on how it is characterized—may be called teleology, a vital force, or some associated notion. The problem that brings teleology to the fore is therefore first and foremost about

[47] As the AT sees it, the contrast between mechanism and autonomy is in terms of the direction of determination: for mechanism determination is oriented from parts to whole; in autonomy it is reciprocal.

the very *existence* of certain kinds of phenomena which display a level of intricate organization that cannot be accounted for by our best physical theories.

The idea that explaining the existence of an organized entity is closely tied to teleology was perhaps given its most famous formulation as an account of functions by Larry Wright in the 1970s. As discussed in the introductory section which preceded this chapter, Wright framed his discussion in terms of a general characterization of scientific explanation as one which cites factors that bring about the explanandum, or in his words, cite an *etiology* (Wright 1972). Causal and teleological explanations are two kinds of etiological explanations, where the former explain what brought about the explanandum by citing preceding factors, whereas the latter cite consequences. To identify the function of traits and artifacts, according to Wright, is also *to explain why they exist*.[48] And this explanation is teleological because it explains in virtue of a consequence: the function of a heart is to pump blood because pumping blood is a consequence of what it does, and the consequence is also the reason why the heart *exists*. Wright provides the following example for artifacts: "The reason the sweep-second hand is there is that it makes seconds easier to read. It is there because it does that. Similarly, rifles have safeties because they prevent accidental discharge." (Wright 1973, 158). In these contexts, human intention is part of the etiological explanation, but only implicitly, as can be seen in Wright's reference to the "effort" that was put into constructing the artifacts in this way (ibid). For biological functions, Wright suggests that we refer to natural selection instead, and today the etiological account is almost synonymous with this view (which is often referred to as the 'selected-effects' view; Garson 2018). But as we shall see, tweaking the conditions for such explanations allows for the explanation to be cashed out without invoking a process that happened in the past. Regardless, it is important that in both cases the purpose or *function* of these entities is taken to be explained causally. For that reason, I will dub these explanations "causal-consequence etiologies".

---

[48] As Wright admits, the meaning of "existence" or "is there" is left ambiguous by the account (1973, 158). One can therefore fill in this basic scheme with more or less robust notions of existence.

Since Wright's seminal work, many have taken such causal-consequence etiologies to capture the distinctive content of a teleological explanation. For reasons that will be addressed explicitly in section 4.3, I do not think Wright would have necessarily agreed with this given a significant ambiguity at the heart of the notion of 'etiology'. That aside, what makes causal-consequence etiologies teleological is the fact that the explanandum is about *existence* and the explanans is about some effects or *consequence* of the entity in question. Such a characterization of teleological explanation is endorsed by the AT. The existence of biological (constraint-closed) systems is explained by what they do; the persistence of the system is a causal consequence of its activities. This circular logic precisely mirrors the reasoning which they use to claim that biological organization is intrinsically purposive.

To view teleological explanations as causal-consequence etiologies has three important implications: it avoids the problem of backward causation; it allows for distinguishing functions from side-effects; and it identifies purposes with (at least some of) the causes of existence. I will now review these in turn.

First, with its focus on existence as an explanandum, teleological explanations form a particular kind of causal circularity (Toepfer 2008; 2012). If the heart's existence is explained by something that it does, then that effect is also its cause of existence: a heart keeps existing over time thanks to the fact that it pumps blood, keeping both the organism and itself in existence. This causal circularity is of primary importance for evading the primary criticism which all accounts of teleology face, namely, backward causation. In an ordinary (efficient) causal model, it appears impossible for the effect to explain the cause without the future somehow causing the present, given that effects always *follow* their causes in time, or at least do not precede them (refs- esp. from authors discussed in other chapters). But with a notion of causal circularity, consequences *are explanatory in virtue of being causes*, and therefore do not raise the specter of backward causation.

Second, framing teleology in terms of existence is crucial for distinguishing functions or goals from other effects. Goal-directed entities have many effects, but only a subset of those are its goals. A heart may make thumping sounds, but that is not the goal of heart pumping. So too for functions, artifactual and biological: knives and hearts make sounds, but that is not their function. Existence therefore plays a role in selecting or filtering some of the effects from the others. It tells us why some effects are the function or goals of an entity, and not others. And in all cases, it is only the effects that play a causal role in the entity's existence that are its goal or function.[49] In so doing, causal-consequence etiologies are thought to capture the entire set of normative implications of goals and functions, where something fails to fulfil its goal or function just if it fails to do that thing which explains its existence.

Finally, causal-consequence etiologies identify goals with the causes that explain existence. For selected-effect accounts, those purposes *qua* causes of existence refer to something that the type of entity did in the evolutionary past which increased the fitness of the organisms it belonged to. But nothing in the formulation of causal-consequence etiologies dictates temporal conditions for assigning purpose. Indeed, the AT takes the relevant causes of existence to be those relating to the constraint closure of the autonomous system in the present. This appears to be a significant advantage of the AT's account of functions, since it allows it to evade many critiques which have been leveled at natural selection-based accounts (Mossio, Saborido, and Moreno 2009). But it also has particular implications. If the goals are determined by causes of existence, and those causes *qua* goals have to do with what an organism or organ *currently* does, *then there can only ever be one kind goal: continued existence*. A crucial implication of this is that the criteria by which we identify an intrinsically purposive entity are the same as the ones which we use to identify what kinds of goals they have. To say that an entity is intrinsically purposive is also, *ipso facto*, to say what its purpose is, namely, its continued existence or persistence.

---

[49] This filtering of goals or functions from mere effects is also what Ruth Millikan (1989) and Karen Neander (2017), among others, have sought to do by appealing to past selection (see also: Garson 2024).

For reasons that will become evident in section 4.3, it is crucial to notice how the AT's way of characterizing teleological explanations very closely tracks its account of intrinsic purposiveness. Both causal-consequence etiologies and closure of constraints share a common (non-vicious) circular logic. A constraint-closed system is intrinsically purposive insofar as its existence is dependent on its activity. From these conditions is derived teleology as a mode of explanation with a particular content: one in which an entity's existence is explained by one or more of its effects. It should also be noted that the AT takes constraint-closure to strictly delimit the necessary conditions for teleological explanations. As Mossio and Bich (2017) argue, cyclical phenomena such as the weather cycle are not the appropriate conditions for a consequence etiology, because cycles do not exist as a coherent unity. "The idea of intrinsic purpose, we submit, does not merely point to the realization of a circular relation between causes and effects but, rather, to the situation in which the *activity of a system*, by producing some effects, contributes to specifying the conditions under which the circular relation as such can occur. It is in this precise sense that the connection between teleology and self-determination is to be understood." (Ibid, 1107; emphasis added). In what follows, I will show why it is a problem to characterize the content of teleological explanations in this way.

4.2 The Ecological Approach: Teleology as a Non-Causal Explanation

4.2.1 The Ecological Approach to Agency

The ecological approach takes a distinctively different approach to conceptualizing the phenomenon of agency as the necessary condition for teleological explanations. As we've seen, the AT characterizes the phenomenon in minimal terms as closure of constraints, which it uses as a scaffold for other concepts such as regulation (Bich et al. 2016; Bich, Mossio, and Soto 2020). Closure is itself built out of thermodynamic concepts which are familiar to the physical sciences, which allows the AT to highlight the way their theory is continuous with those sciences while also capturing what is distinctive about biological systems.

In contrast, the EA makes a more radical distinction between so-called "object theories" and "agent theories", given that "citing the mechanisms by which agents are realized might not be sufficient to account for the difference that agents make to the world." (Walsh 2018, 170). Walsh's definition of the term 'mechanism' needs to be carefully distinguished from the sense in which the AT uses it. In this context, mechanism means an explanation that cites the causal or physical realization of a system in a general sense (Fulda 2017). This would include the kind of productive relation between constraints as (thermodynamic) causal processes that make up a constraint-closed system. Fermin Fulda illustrates this distinction with an example from fluid dynamics. Properties like viscosity are causally realized by molecules that constitute the fluid, but viscosity itself "is not specified in terms of the properties of its molecular realizer but in terms of other concepts at the same gross dynamical scale, such as density or surface tension." (2017, 82). In other words, those gross dynamical properties are characterized empirically in a way that is agnostic to the way they are physically realized. In the same way, the EA characterizes agency in terms of its empirically observable gross properties. What we require to make sense of the behavior of an organism are a suite of concepts that are proper to agent theories, including "goal, means, affordance, repertoire, salience, reciprocal constitution, normative requirement, hypothetical necessity, teleology" (Walsh 2018, 174). The distinction between a system's causal-mechanical realization and its gross properties is of course reminiscent of purposive behaviorism, and especially the cybernetic distinction between behavioral and functional modes of study. But it is not only to be found there; in fact this very distinction was suggested by Jonas himself, who gives the reader the "Aristotelian reminder" that "*finalism is in the first place a dynamic character of a certain mode of existence*, coincident with the freedom and identity of form in relation to matter, *and only in the second place a fact of structure or physical* *organization*, as exemplified in the relation of organic parts to the whole and in the functional fitness of organism generally" ([1966] 2001, 86; emphasis added).

Goals, repertoire, and affordance in particular form a cluster of inter-defined concepts that characterize agency (Walsh 2012b). According to the EA, "agency is the gross behavioral capacity of an organism to bias its repertoire in response to what its conditions afford for attaining its goals. This set of ecological concepts goal, repertoire, and affordance [...] specifies the ecological profile of the system" (Fulda 2017, 83). A goal is a state of affairs that can be reliably achieved across a range of circumstances, given the organism's behavioral repertoire. In turn, the way that behavioral repertoire is deployed will depend on the situation that the organism finds itself in, or more specifically, what the situation *affords* from the organism's perspective. This account is particularly well-suited to conceptualizing goal-directed activities given the way that it conceptualizes them in a context-sensitive way.

This way of characterizing agency represents a particular approach to naturalism, which closely tracks what Wright called 'empirical' approaches to teleology. Such empirical approaches justify the use of concepts on the basis of empirical evidence and explanatory indispensability, and the example of fluid viscosity is intended to show that this is in fact a common practice in science (Batterman 2015; Walsh 2018, 172–73; Fulda 2017). In the words of Wright, "In a large number of cases, the goal-directedness of a bit of behavior is obvious on its face. Many of our teleological judgments are as reliable and intersubjective as the run of normal perceptual judgments. Occasionally there simply is no question about it: the rabbit is fleeing, the cat stalking, the squirrel building a nest. Certain complex behavior patterns seem to demand teleological characterization" (1972, 206). By the standards of an account like the AT, such an approach may seem to be helping itself to concepts that may be deemed dubious and unearned; there is simply no reason for us to take such everyday intuitions seriously. But the idea that these are mere intuitions may obscure a stronger pragmatic claim which the empiricist may make, namely, that we have a substantial pre-theoretic grasp of how to effectively explain the behavior of agents from our day to day lives. As I shall argue in the remainder of this chapter, the AT's concerns are justified; but while the approach fails to conform to particularly stringent forms of naturalism, it also gains in empirical adequacy and scope. Moreover, an empiricist feels far more strongly that our account of teleology should share the same fundamental structure as every-day explanation of actions. Wright put this well: "if we accept human action explanations as a model of legitimate teleology, then other teleological explanations should be expected to share *their salient logical features*; and explanations in terms of reasons, motives, and intent are plainly etiological: they help us understand what brought about the behavior in question" (ibid, 205).

4.2.2 Explaining with Goals

Much like the AT, the way the EA approaches characterizing the phenomenon of intrinsic purpose as a condition for teleological explanation has a significant impact in the way that it conceptualizes the content of such explanations. In this view, teleological explanations track the relation between goals and means in much the same way that causal explanations track the relation between causes and effects, and this is what makes both explanatory. At the same time, the differences between the cause-effect and means-end relations substantially distinguishes teleology from causal explanations.

More precisely, Walsh argues that what makes a causal (or causal-mechanical) explanation adequately explanatory is the fact that it involves an *invariance relation* and an *elucidating description*. The elucidating description appeals to the particular activities (such as "unwinding, bonding, and breaking") which parts of mechanisms perform in order to produce a particular effect (Machamer, Darden, and Craver 2000, 21–22). The importance of such activities, "lies not so much in the fact that they *identify* the causal relation between explanans and explanandum, but that they '*elucidate*' that relation" by showing *how* the activities of the mechanism produced the effect (Walsh 2012a, 175). In doing so, such elucidating descriptions also produce understanding of the phenomenon in question. One might think that mechanisms end up being explanatory because they are what cause the effect; but the relation may be characterized more generally in terms of *invariance*. An invariance relation is a robust counterfactual relation which describes how the change or difference in one variable X, would lead to a change in another variable Y (Woodward 2005). Variable X can be said to cause Y, allowing for these variables to act as explanans and explanandum respectively.

According to Walsh, teleological explanations share the same structure as causal explanations insofar as they also involve an invariance relation and an elucidating description. The invariance relation in a teleological explanation is between the means and goals, where means counterfactually depend on goals. A change in the goal of a behavior would, all else being equal, lead to a change in the means pursued for that goal in a systematic way. The various alternate means are related by conduciveness to the goal (Walsh 2012; Fulda 2017). The key here is the way that teleological explanations carve up the modal space differently to causal explanations, and therefore tell us something significantly different about what *could* have happened. Recall William James' contrast discussed in the introduction which preceded this chapter: if the path between iron filings and a magnet is blocked by a card, then the fillings will remain pressed to the card indefinitely. But if the path between Romeo and Juliet is blocked, then Romeo will find a way around the obstacle if it is within his power to do so (James [1890] 1983, 20).

In this view, the content of a teleological explanation also has implications for normativity. The elucidating description provided by a teleological explanation lies not in the way activities produce their effects, but in the way goals normatively *require* their means (Walsh 2012; Fulda 2017). This requirement is one of hypothetical necessity in Aristotle's sense, according to which an axe's hardness might be said to be hypothetically necessary for it to carry out its purpose (Cooper 1987). In virtue of their hypothetical necessity for attaining a goal, means can be said to *conduce* to their end. Whereas causes explain their effects by producing them, means are explained by their ends because they conduce to them, i.e., by elucidating why they are required (Walsh 2012, 178).

Walsh's attempt to distinguish causal explanations and teleology is notably different from the AT's attempt to derive teleology from causality. By taking consequence etiologies to be sufficient for teleology, the AT has been able to identify a particular kind of causal relation that meets this criterion, namely, a circular one. From the EA's perspective, the AT's causal-consequence etiologies are not veritably teleological because the fact that self-maintenance is a goal is incidental to the explanation. It *explains* in virtue of being a cause only (see also: Deacon and García-Valdecasas 2023b; García-Valdecasas and Deacon 2024). For this reason, to be "cause and effect of oneself" may not be as close as it seems to Kant's notion of a natural purpose. On one interpretation, Kant's formulation, like Aristotle's, is one that assumes a notion of goals that is not translatable to the causal language of the first statement (Koutroufinis 2016). Moreover, the relation of hypothetical necessity between means and goals is taken to be intrinsically normative, and equally underivable from the causal necessity which relates causes to their effects (Fulda 2016; 2017).

4.3 Consequence-Based Explanations: Two Interpretations of Wright's Etiological Account

The previous two sections have shown how the AT and the EA attempt to formulate the content of a teleological explanation such that it can be distinctive, while also being explanatory by the standards of causal explanations. Both agree that what typifies a teleological explanation is the fact that it cites consequences as the *explanans*; though this far from exhausts the full content of such explanations. In this section I highlight the differences between the AT and EA's conception of teleological explanation by way of a certain ambiguity in Wright's conception of teleology, which opens it to two interpretations that closely map onto each of the two approaches.

As Wright explains, the intention behind his choice of the term *etiological* was to highlight what causal and teleological explanations have in common (1968). Wright did this to show that teleology could be distinctive without being completely "sui generis": "The causal/teleological contrast is among etiologies, not between etiologies and something else" (1972, 217; see also 205, fn. 3).[50] For Wright, an explanation is etiological and thereby properly explanatory if it can show "what brought about the behavior in question" (1972, 205). But these etiologies in question can be either antecedent conditions (in the case of causal explanations) or consequences (in the case of teleological explanations): "When we say that teleological etiologies are consequence-etiologies we are saying that the consequences of goal-directed behavior are involved in its etiology: it [the behavior] *occurs because it has certain consequences*." (ibid, 217; emphasis added). One way of interpreting this claim, which goes in the direction of the AT, is that Wright is showing how an explanation that appeals to consequences can be made sense of *within* a broadly causal framework. All that is required for a causal explanation to be a consequence etiology is either i) that the existence of the entity in question be explained by something it did in the past, or that ii) there be a circularity between causes and effects. The first option is pursued by the selected-effects theory of etiological functions which, as I briefly described in 4.1.2, explains the cause of existence of a particular trait (such as the heart) in virtue of the fitness-enhancing *consequences* it had in the past. In this case, the explanans is a consequence, but the explanation itself is an ordinary causal explanation. The second option is pursued by the AT. As we've seen, constraint closure involves a circularity of causes and effects, whereby a consequence can be rendered explanatory precisely because it is also a cause. Because the consequences of self-maintenance make the very existence of the self-maintaining system possible, the act of self-maintenance is both a cause and an effect of the system. It is in this sense

---

[50] What Wright means by "sui generis" is not completely clear, which contributes to exegetical ambiguities regarding how broadly the notion of "etiology" is to be understood. In an article published four years prior, Braithwaite uses "sui generis" to describe vitalistic, "non-physical" accounts of goal-directed behavior (Braithwaite 1968, 327).

that Mossio and Bich can claim that biological organization makes possible a particular "*circular* causal regime that grounds intrinsic purpose" (Mossio and Bich 2017, 1106; see also: Toepfer 2008). In sum, *existence qua explanandum is relevant for the AT's teleological explanations because it leads to circularity, which makes it possible for effects, qua causes, to be explanatory.* As shall be argued in the next section, the need for teleological explanations to be circular in nature, and particularly to treat existence as an explanandum, causes some problems for extending this as an explanation of action and artifacts.

An alternative way of interpreting Wright's argument regarding etiology is *not* as an account of the conditions under which the terms of a causal explanation (causes or effects) can be treated as goals, i.e., in those special cases when consequences *qua causes* are explanatory.[51] Rather, Wright is giving an account of the conditions under which goals might explain in an etiological fashion, i.e., explain by showing how goals bring something about—say, a behavior—which would not have occurred if it were not for the sake of the goal. According to this interpretation, a goal is different from a cause. For Wright, like the EA, goals are not something to be derived from mere effects but something that is "justified if goal-directedness (i.e. our formula) is the best account of that something's behavior…in some cases, the goal-directedness of behavior is obvious, palpable, instantly recognizable, and unmistakable to the normally sighted individual. Which is to say it is observable… I clearly want to be able to say we can explain the behavior of the system by appeal to its having a certain goal." (Wright 1972, 216). Here it seems that teleological explanations are not teleological merely in virtue of the fact that they explain via consequences; because the consequence must be a goal, not just a cause or an effect, and *its being a goal must be explanatory*. For this reason, "final causation, anthropomorphism and teleological explanation [are] not reducible to an underlying, deterministic causal mechanism." (ibid, 206).

[51] I take this to be a better way of making sense of Wright's apparent enthusiasm for Charles Taylor's work and some key passages in Wright's 1968 and 1972 papers which are otherwise hard to interpret.

Or: "Teleological expressions are not 'translatable into' or 'saying the same thing as' some causal expression. They do not 'have the same *cognitive content*', and they differ in more than mere 'emphasis'." (Wright 1968, 223; emphasis added). On this interpretation, the EA is much closer to Wright than might have been expected (see especially Walsh 2012). Not only that, but Wright's notion of etiological seems to be closer to Aristotle's *aitia*, which—on one interpretation—refers broadly to the four kinds of facts in the world that an explanation may appeal to as the *explanans* for the phenomenon in question, of which causes and goals are two (Stein 2011; Toepfer 2021).[52]

In sum, the first interpretation of etiology explains the AT's dependence on existence to make consequences—but qua causes, not *qua* goals—be explanatory. More precisely, the identity between goals and causes in this context is precisely thought to be the key naturalizing move. Whereas the AT identifies teleology with circular causality—made possible by the fact that the explanandum of teleology to be existence, and the explanans some effect of the entity in question—the second interpretation of Wright, which is closer to the EA, is really all about the sense in which goals explain. In other words, the AT and EA have different conceptions of what it is to be a goal. For the former it is when an effect acts as a cause in a regime of circular causation. For the other it is to be a state of affairs that brings about some event because that event was hypothetically necessary for the attainment of the goal.

This marks a crucial difference. If goals are a type of *aitia* or cause which does not need to be derived from—in Aristotle's terms—efficient or material causes, then there is also no need for circularity, and thereby no need for making existence the explanandum. For this reason, a teleological explanation, on this second, Aristotelian-like view, *is not restricted to explaining existence, nor does it identify goals with the continuation of existence*. For Aristotle teleological explanation is about

---

[52] Indeed, when Wright states that "the term 'etiology' has impeccable *etymological credentials* for this use", it appears that he can only be referring to the ancient Greek *aitia* (1972 205, fn. 3; emphasis added).

goals being the *explanantia* (and again this is an argument about the *content* of teleological explanation, not the nature of goals per se, i.e., an account of what they "are" in more metaphysically basic terms).

The above carves out an important difference between two ways of thinking about teleology as a consequence-based explanation, which will be crucial for understanding the respective tradeoffs for the AT and the EA which occur as a result of this choice in the next two sections.

But before turning to this discussion, it is important to note that there is a third way in which a consequence can be explanatory, which appears *neither* teleological *nor* causal: equilibrium explanations. These are non-causal forms of explanation in which a resultant equilibrium state explains why something happened. Elliot Sober argues that R. A. Fisher's explanation of the 1:1 sex ratio in many species is a kind of equilibrium explanation: wherever sex ratios diverge from 1:1, and mating is random, there will be a fitness advantage to producing a preponderance of the minority sex, which in turn will move the sex ratio closer to 1:1 (Sober 1983). If the ratio changes from this point, it returns to it *because* it is the equilibrium point. If one assumes that only causes can be explanatory—a position which has become increasingly difficult to square with scientific practice—then this is either false, or the explanation is a causal one in disguise. But let us grant for the moment that this is a valid, non-causal form of explanation. What renders it non-causal is the irrelevance of the initial conditions and selective forces to the explanation: "whatever the actual initial sex ratio had been, the selection pressures that would have resulted would have moved the population to its equilibrium state. Where causal explanation shows how the event to be explained was in fact produced, equilibrium explanation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired." (ibid, 202). At the same time, this example illustrates how the non-causal explanation is "made true" by causal facts, though it is the non-causal counterfactual relationships that are doing the explanatory work (Rice 2021, 100).

To conclude, there appear to be *three* ways for consequences to be explanatory: i) qua cause; ii) qua equilibrium; and iii) qua goal. Evidently, given that the idea of a goal appears somewhat mysterious, many will insist that the third type will necessarily be translatable into the first or second; and we can take the AT to at least implicitly be making this claim. To be more precise, from the AT's perspective a bona fide teleological explanation requires that a goal be nothing other than a cause. In section 4.5, I will evaluate whether the EA's conception of teleology is in fact an equilibrium explanation in disguise.

4.4 Tradeoffs of the Autonomy Approach

In this and the next section, I articulate the tradeoffs that I see in each approach.[53] In a nutshell, the AT abides by a stricter form of scientific naturalism which attempts to derive or explain teleological concepts from non-teleological concepts, very much to its credit. But what it gains here, it loses in the scope of its teleological explanation. In this section I articulate some of these tradeoffs which the AT must contend with, focusing on the underlying commitment to a causal-consequence etiology framing. I will show how the AT's formulation of teleology as a circular explanation which treats existence as an explanandum encounters some problems when applied to actions and artifacts. My aim is not to use these cases as knockdown arguments against the AT, but rather to articulate its respective strengths and weaknesses relative to it. In section 4.5, I will give the EA a similar treatment.

4.4.1 The Problem of Action

In the introduction that preceded this chapter, I spelled out what I take to be the content of an everyday teleological explanation. I took William James to be right

[53] Much of this section is the product of long discussions with Louis Virenque, and this chapter as a whole is far richer for his contribution and his careful comments on earlier drafts. The issues discussed in section 4.4.1 in particular were presented by him on two occasions (Virenque 2022; 2023).

when he characterized Romeo's behavior as one which is not fixed but in which the "path may be modified indefinitely" (James [1890] 1983, 20).

To illustrate a core problem with the AT's account of teleology, it is worth reviewing how it fares when extended to a theory of action, as has been recently done under the banner of enactivism (Di Paolo, Buhrmann, and Barandiaran 2017). Since the AT has no room for there being any goals other than continued existence, all actions must be directed toward their persistence to be goal-directed; a prima facie unintuitive result. There are two ways to make sense of this. First, to render actions as extrinsically purposive functions in much the way organs are; second, to show how actions form closed loops that are themselves intrinsically purposive. I will evaluate these in turn.

The first, perhaps more obvious solution for the AT is to render all action a kind of function, since the AT can admit to a vast array of ways of achieving the goal of continued existence, especially when the entity in question is a complex one. Activities such as play may have no immediate connection to survival, but one can see how such activities may have long term effects that do pertain to that goal.

A potential problem for this solution is that the link between the immediate function of an activity and its long-term promotion of survival may not always be clear (Virenque 2022; Virenque and Mossio 2024). Moreover, there seem to be many activities that seem goal-directed but do not contribute to the persistence of the system in question—call these "bad goals". Obvious cases are self-harm and suicide, but this also extends to activities that are neutral with respect to self-maintenance. Examples abound when we look to humans, who routinely do things that undermine their long-term persistence, such as smoking, or are neutral, such as whistling. But we also find cases of this in the non-human realm too, such as the death-inducing efforts of salmon to swim upriver to spawn or self-harming behaviors of animals under extreme stress. These are far from irresolvable problems for the AT, and a solution to the former issues has already been proposed (Artiga and Martínez 2016; Mossio and Saborido 2016). But they begin to illustrate the challenges of indexing goals to existence.

The second solution to applying causal-consequence explanations to actions can be found in a recent proposal by Ezequiel Di Paolo, Thomas Buhrmann, and Xabier Barandiaran (2017). They argue that actions form part of sensorimotor loops, or habits, that display a logic analogous to that of closure of constraints; close enough, indeed, to be intrinsically purposive themselves. The habitual deployment of actions to recurring affordances becomes the building blocks of self-sustaining action patterns. A habit can be understood as "a self-sustaining pattern of sensorimotor coordination that is formed when the stability of a particular mode of sensorimotor engagement is dynamically coupled with the stability of the mechanisms generating it'' (Barandiaran 2008, 281). In other words, habits are the natural outcome of self-sustaining sensorimotor schemes that integrate the system with its environment (Di Paolo, Buhrmann, and Barandiaran 2017). Distinctive norms emerge because habits enable a way of life that is always precarious and open to disruption (Egbert and Barandiaran 2014). It takes the adaptive capacities of sensorimotor agents to plastically reconfigure their sensorimotor schemes to deal with novel challenges encountered in their environments.

Assuming that one can in fact talk of sensorimotor loops as closures of constraints—which is by no means established—this strategy involves decoupling the intrinsic purposiveness that lies at the core of biological organization from higher "levels" of intrinsic purpose that emerge through sensorimotor coupling (Virenque 2022). But this has two troubling consequences. First, it means that the goal or function of any action is ultimately to be able to perform that action again. And though it is true that actions involve sensorimotor adaptations that make their future occurrence easier and more efficient, it is very odd to say that this is their *goal*. Clearly, the goal of waving down a waiter is to get their attention, not to be able to do the same action on a future occasion. The habitual nature of this action just makes it conducive to achieving that purpose. Moreover, it is not clear what we are to make of the fact that many everyday actions aren't habits at all.

Second, since closure is supposed to provide us with the identity of the agent to whom we might attribute such goals, then it the locus of intrinsic purpose would be the sensorimotor closure itself. The goals of actions therefore belong to the action itself, rather than to the agent deploying them. While the idea that a habit may be so deeply ingrained as to have a quasi-autonomy over us is not completely far-fetched, it is very odd to take this as a *paradigmatic* case of action, or indeed to make it a requirement for the action to be goal-directed at all. In fact, the most representative case of an autonomous and self-maintaining sensorimotor loop would be a pathology called alien hand syndrome, in which one of a person's hands will act on its own, against a person's will. A more common example may be that of addiction, which is antithetical to the notion of autonomous agency. I submit that even in such extreme cases, attributing the locus of intrinsic purpose to the sensorimotor scheme is highly empirically unsatisfactory: when we see agents act, it is the agent that is acting and to whom we attribute the goal, *not* the sensorimotor action loop itself. This is effectively a version of the disappearing agent problem in the philosophy of action (Velleman 1992; M. E. Bratman 2001; M. Bratman 2007; Hu 2024).[54]

4.4.2 The Problem of Artifact Function

Given that the AT is intended to be an account of *biological* purposiveness and teleology, one may think it unfair to critique it for failing to make sense of artifact functions. Artifacts as distinctively human creations may simply not be relevant to an account of biological agency and teleology. That being said, there are several phenomena that the AT ought to encompass which share key features with human artifacts, including cases of niche construction which we find in many animals,

[54] This problem of misallocating the locus of intrinsic purpose recurs in other contexts. The process of evolution itself could also realize intrinsic purpose if we ascribe functions to adaptive traits. If these traits are functional they must be, from the perspective of the AT, part of a circular system that realizes intrinsic purpose; but this circular system is not necessarily the organism, as discussed above—it could also be an evolutionary process itself, or a lineage, where the trait's fitness explains its existence. This is, in effect, the AT's variant on a selected effects theory of function, with the added twist that we are committed to the trait being functional in virtue of a closed loop that it contributes to maintaining. In both the action and evolution case, since teleology is grounded in a notion of circularity that is more general than organizational closure, there is the possibility of both actions and evolution being *intrinsically purposive*.

such as dam and lodge building in beavers, and tool use found in chimpanzees or crows (Toepfer 2008). These structures could be said to act much like external organs, being created and maintained by the activity of the organism and in turn contributing to maintaining the life of the organism. In this case, these artifacts can be rightly said to exist because of something that they do, namely sheltering and protecting the organisms that in turn maintain them. Such artifacts self-maintain, i.e., are cause and effect of themselves, but only *indirectly*; just as parts of organisms, such as organs, maintain themselves indirectly by contributing to a larger causal chain they are part of, and in virtue of which they exist. For this reason, artifacts are extrinsically purposive rather than intrinsically purposive.

While this seems to work for structures that require maintenance to persist, it is not so clear how the account fares when faced with non-constructed or non-designed artifacts, i.e., found natural objects that come to have an artifact-like existence (which following Devitt 2023 I will call "implements"). Examples of implements include sticks and rocks which may fulfill functions such as digging holes, hammering stakes, or prying windows open. In this case, such non-designed artifacts clearly have a function; and yet the existence of such entities has no connection to their function. A rock that has played its role in hammering a stake may be thrown away and cease to have a function while persisting—indeed, its use as a hammer may *contribute* to its ceasing to exist faster than it would otherwise. The problem therefore generalizes to cases of gain and loss of function.

What the AT account needs to make sense of implements (non-designed artifacts) is a broader notion of 'existence'. The notion of existence that we have used so far has primarily been that of continued existence in virtue of constant *maintenance*. This is because organs, or whole organisms for that matter, as well as beaver dams would eventually cease to exist if they are not continuously maintained. But this sense of existence does not work for implements that do not require maintenance to persist. Instead, it seems that existence means a particular spatio-temporal location: to exist "in my hand" rather than "on the ground". While being used for hammering, the rock is indeed moving in a highly improbable, orderly fashion; and it will keep doing so as long as it serves its purpose. But this response, while plausible, raises serious questions about the scope of "existence". Indeed, one may worry that this is a gross overextension of the concept of existence. Perhaps sufficient precision is gained if existence means "existence as part of a closure", meaning that the rock has a particular spatiotemporal existence as part of the sensorimotor loop that it contributes to indirectly maintaining. Again, we are mapping the logic of organs and the ways they gain functions onto artifacts, though in this case the artifact can persist on its own, but a heart cannot. But to say an artifact's function is to do whatever it does which explains its 'existence as part of (sensorimotor) closure' suffers from the problems outlined in the previous section, namely, that it can *only* have a function if the whole sensorimotor loop establishes intrinsic purpose.

As stated at the outset of this section, there is no reason to think that the AT, as a biological theory, ought to provide a good account of artifact function. However, it is worth considering what shortcomings would be encountered if that was attempted, as it further illustrates the issues discussed thus far. As Mossio and Bich say, the goal of such artifacts "does not coincide with their own existence", since "the goal of a knife is not to maintain itself although, of course, it has something to do with its existence since the knife is designed for a certain use" (Mossio and Bich 2017, 1090). What exactly is this relation with existence? There are two possibilities.

The first is to interpret existence in terms of *genesis.* In this sense, a particular artifact would never have been created were it not for the fact that, once created, it would have a particular effect. The artifact exists—i.e., was created—because of an effect that its creator *anticipated* it would do. This effect is therefore the artifact's function because it exists due to that function. That being said, this sense of existence (as applied to artifacts) requires a capacity for intentional forethought, given that it is a yet-unrealized effect or consequence that is supposed to do the explaining;[55] and it is unclear how intentional goal-directedness can itself be cashed out in terms of the AT's conception of teleology. Furthermore, this interpretation also assumes that creators always have a good idea of what their creations will do, which makes accidental discovery hard to account for; and it doesn't work for implements and co-opted functions.

The second possibility is to consider the class of particular artifacts, say knives. While it does not make sense to say that a token knife's function is to maintain itself (albeit indirectly) in virtue of cutting (a knife that ceases to cut does not cease to exist), one might say that the type's existence is maintained by a society in virtue of the function they served. In this view, a society would be a circular self-maintaining system which partly relies on artifacts for its persistence. Therefore, particular classes of artifacts persist, and are functional, in virtue of something they do.

Finally, consider the case of artificial organs. Again, how we make sense of artifactual function depends on what one means by existence. If one means genesis or generation, one runs into the same issues articulated above. Alternatively, if existence means "maintenance", then one can say that the artifact continues to exist because of something it does. If it ceases to do that thing—its function—then it ceases to exist. Either option leads to a strange result for artificial organs. Once implanted, an artificial organ's function is the same as a natural heart, but not because it contributes to maintaining the closure of the organism—even though it evidently does that too.

The foregoing indicates some notable ambiguities regarding the meaning of existence, and the risk that its meaning may be overly open to interpretation. It is this ambiguity, I would suggest, that underlies many classic counterexamples to the etiological account of function, such as the idea that a leaky gas hose's existence can be explained by the fact that it knocks out anyone who tries to fix it,

[55] For this reason, one might think the AT's conception of teleology is not equipped to teleologically explain cases of genesis in the organic realm, say, organogenesis. One would need to revert to a backward-looking account of etiological function, such as the selected effects theory.

or that a stick floating in a river creates a backwash which keeps it stuck in a position that maintains the backwash. In both the action and artifact examples, the core difficulty seems to be the way teleological explanations are tied to existence. Existence plays a key role in simultaneously providing the criteria for the identification of intrinsically purposive systems, and what counts as a teleological explanation. In other words, it tells us both what kinds of things have goals—those that are "cause and effect of themselves", i.e., whose existence is explained by what they do—and what those goals are, namely, to keep on existing. As a result, self-maintenance becomes a defining part of *both* the *conditions* and the *content* of teleological explanations.

4.4.3 The Problem of Normativity

One of the key components of a teleological explanation is its normative content. When we explain why a particular course of action was pursued by citing a goal, we are also stating why these actions were normatively appropriate for achieving it. The AT claims to be taking such normative aspects of teleological explanation seriously; and yet it is not clear if the causal-consequence explanation framework does it justice. A case for this has recently been made by Luca Corti (Corti 2023; see also: Moosavi 2019), and again a key issue appears to be the role of existence as explanandum. According to the AT, the heart has a function because its existence is explained (albeit indirectly) by something it does: pumping blood. Its function is to do that, and it therefore *ought* to do that. If it fails to do so, then we say that it malfunctions. Similarly, the AT holds that an organism's goal is to persist because that explains why it exists, and it therefore *ought* to persist. In Corti's words:

> normative ascriptions (what an item "ought" to do) derive their justificatory force from the fact that the performance of constraint is essential for the very existence of the organization. To use a standard formulation employed by its proponents: "the system must behave in a

> specific way, otherwise it would cease to exist" […] According to this line of reasoning, what grounds functionality and normativity is the "very existence" of the system and its parts (Moreno & Mossio, 2015, p. 70). More precisely: normative and functional ascriptions derive their justificatory force—and their objectivity—from a core counterfactual argument regarding the "very existence" of the system.
>
> (2023, 10)

Corti then argues that while the AT's account of biological organization may well be correct, in the sense that they rightly identify "objective, counterfactually robust conditions" regarding the conditions of existence specific to living systems, those conditions are neutral with respect to normative ascriptions (ibid, 11). This is effectively the age-old problem of deducing normative from non-normative facts. "There is nothing in the notion of "condition" or "dependence" as presented by AT that enables such a logical step." (ibid, 12).

Corti concludes that what the AT needs is to show that the "existence of organizationally closed systems has a value built into itself" (ibid, 18). But Corti rightly points out that this is an additional claim that "must be separately argued for: nothing in the AT conceptual apparatus considered thus far entails this assumption." (ibid, 18). The peril of this strategy is that any attempt to inject values in this way would risk being purely stipulative and non-naturalistic. Nor can the AT lean on its use of agential language in describing how organisms must act in certain ways to stay alive, because this talk can be "entirely translated into causal terms without loss of explanatory import" and therefore "one might end up with a quite deflationary interpretation of its normative implications." (ibid 19, fn. 35). In this light, Corti's critique of the AT's account appears to be about the possibility of deriving normativity, but a whole suite of teleological concepts from purely causal terms. In this sense, "[t]he decisive question, however, is to understand how strongly AT theorists want to commit to the idea that all notions belonging to the framework are ultimately causal" (ibid, 19).

4.5 Tradeoffs of the EA

Much like the AT, the EA must contend with certain tradeoffs which come with the ways that it formulates the content and conditions of teleological explanations. In this section, I articulate three closely related issues related to the EA's defense of teleology as a form of non-causal explanation.

4.5.1 Assuming and Explaining

Perhaps the clearest tradeoff of the EA is that it assumes what the AT aims to explain. Given that the core concepts associated with teleology, particularly goals and norms, are somewhat mysterious from a physicalist point of view, it seems natural to assume that one ought to explain how they could emerge from such an ontology. To simply postulate these concepts may, in the words of Bertrand Russell, have "all the benefits of theft over honest toil" (B. Russell 1919, 70).

Much of this issue hangs on often unstated assumptions about the standards by which one deems a concept 'natural' or a legitimate part of our metaphysical picture of the world (Sellars 1963; De Caro and Macarthur 2004; 2010; Fulda 2016, chap. 3; Macarthur 2019). Perhaps the real problem with assuming rather than explaining purpose—and its associated concepts—in terms of more basic physical concepts is that it opens the floodgates to metaphysical profligacy (Quine 1948; c.f. Wimsatt 2007). That being said, the EA's attempt to justify helping itself to the concepts it needs are ostensibly grounded on their indispensability to a science that would include agents. Such a science would need to contend with the observability of goal-directedness, the projectability of teleological regularities, and so on.

Moreover, the EA makes the case that scientific concepts can have an indispensable explanatory role without being directly reducible to more fundamental concepts. As discussed in section 4.2.1, defenders of the view hold that the notion of 'goal' is much like that of 'fluid viscosity', in the sense that it is not derived or explained in terms of its underlying causal realization (Batterman 2015; Fulda 2017). But this example, while illuminating, may only go so far given that it is taken from physics rather than biology (c.f. Sultan, Moczek, and Walsh 2021). Even so, such an argument would still rely on substantive assumptions about the role of scientific practice in populating our ontology, assumptions which are different than, but not entirely removed from, those which the AT seems to operate under. Both the EA and the AT conform to a kind of scientific naturalism which, writ large, states that science is our best guide to ontology. Both accounts assert that biology needs its own set of concepts to make sense of what is distinctive of life, and for the EA these are goals, repertoire, normative requirement, affordances and hypothetical necessity. But where the accounts diverge is in how they conceive of the limits on how these concepts are constructed, and what relation these concepts should hold to those of the physical sciences (if any).

But even if we assume that the EA can legitimately help itself to the teleological concepts it deems necessary for explaining organismal behavior, it needs to do more to clarify the metaphysics status of goals and norms. For these seem to belong to a broader category of phenomena which, like affordances, exist at the boundary of perspective-taking agents and the world. In Gibson's words, "an affordance is neither an objective property nor a subjective property; or it is both if you like. An affordance cuts across the dichotomy of subjective–objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither" (Gibson [1979] 2014, 121). This dual-faceted nature of affordances also appears true of goals and norms, and also seems to be a genuine metaphysical category whose precise nature needs further elaboration.

4.5.2 Are There Really Non-Causal Explanations?

A second tradeoff for the EA concerns the status of non-causal explanations. At the outset, it should be made clear that the EA is not necessarily challenging the idea that every event is caused. Rather, it is claiming that even if this is true, it doesn't mean all explanations are causal. Teleological explanations which consist in non-causal counterfactual dependencies are made true, in a certain sense, by causal relationships; but it is the counterfactuals between goals and means that are doing the explanatory work, not the additional fact that they are grounded in causal relationships (Rice 2021, 100). In this sense, the teleological explanation contains information that is irreducible to a causal-mechanical description of the same state of affairs. But in attempting to firmly establish teleology as a distinct form of explanation irreducible to causal explanations, it opens itself to the criticism that teleology is either not explanatory, or a much weaker form of explanation. As discussed in Chapter 1 and 2, these were precisely the problems facing purposive behaviorism, cybernetics, and teleonomy.

According to the causalist view, which was widely accepted from at least the mid 20th century into the early 2000s, a scientific explanation must represent or identify the causes of the explanandum (Carl Gustav Hempel 1966; Salmon 1998; Woodward 2005; Strevens 2011; Skow 2014; Glennan 2017). But the difficulty with this argument is that 'causal' admits of very broad interpretations under which the EA's teleology *would* in fact be causal, and it is precisely these ambiguities that are at the heart of the two diverging interpretations of Wright's etiological account discussed in section 4.3. To articulate the sense in which it is non-causal, we must therefore look to more specific conceptions of causality. Such an example can be found in the work of neo-mechanists, who argue that a causal explanation must show how the explanandum is produced by a *mechanism*, which consists of a combination of part or entities which perform activities that, when organized in the right way, bring about the phenomenon to be explained (Machamer, Darden, and Craver 2000; Bechtel and Abrahamsen 2011). Mechanisms can be structured hierarchically, such that each part of a mechanism is itself a mechanism. As such, this neo-mechanist framework does not assume that all causal power is resident in some ground level of reality. In fact, the importance of *organization* as a concept for recent neo-mechanist work has led to the proposal that a proper understanding of organisms as integrated systems requires an integration between the AT and neo-mechanist perspective, though this is not a consensus view among supporters of the AT (Bich and Bechtel 2021; Bechtel and Bich 2021; Bich and Bechtel 2022a; 2022b).

If neo-mechanists or some alternative versions of causalism are right, and we hold that *only* such explanations are legitimately explanatory, then the EA's conception of teleological explanation may be thought to gain its breadth at the cost of being non-explanatory, or as was discussed in Chapter 2, a mere *descriptive* gloss on what will eventually be translated into a causal explanation (such as the AT's version of a teleological explanation). But the very idea that the EA's teleology cannot be explanatory because it is non-causal would have to confront the recent voluminous defenses of non-causal explanation (Reutlinger 2017a; 2017b).

As several scholars have argued, the core of an explanation lies in elucidating the relation of dependence between explanans and explanandum. What defenders of non-causal explanations challenge is the claim that this dependence must necessarily be causal in nature. A recent definition of non-causal explanation is provided by Alisa Bokulich: "A non-causal explanation is one where the explanatory model is decoupled from the different possible kinds of causal mechanisms . . . such that the explanans is not a representation (even an idealized one) of any causal process or mechanism" (Bokulich 2018, 149). Though this definition may be overly stringent, it suggests that an explanation can pick out *non-causal* features of a system, which figure as an explanans by showing how the explanandum depends on those non-causal features. Such cases are often found in scientific explanations that depend on idealizations and abstractions that abstract away from causal details, precisely because it is the non-causal features of a system that explain (Batterman and Rice 2014; Huneman 2018; Lange 2016; Potochnik 2017; Suárez and Deulofeu 2019; Rice 2015; 2021). I follow Rice (2021, ch. 4) in thinking that what makes such non-causal explanations explanatory is the way they identify a relation of counterfactual dependence between explanans and explanandum, and that this counterfactual dependence requirement is *independent* from a causal dependence requirement (c.f. Woodward 2003).

While there may be different ways to taxonomize non-causal explanations, they appear in various forms including mathematical, statistical, computational, network, optimality, and equilibrium explanations (Lange 2016; Reutlinger and Saatsi 2018, 4; Reutlinger 2017, 4). Each is explanatory yet distinct based on the kind of modal pattern that it picks up.

Of particular interest here are equilibrium explanations since, as mentioned in section 4.3, they also seem to be a case in which a consequence is explanatory.[56] In addition to Sober's case of sex-ratios in a population, one might explain the reason why a ball rolls toward and eventually comes to a rest in the lowest point of a bowl on the basis that it is the point of equilibrium for the system. In such cases, the final state is explained in a way that is agnostic toward the causal trajectories of the ball, but rather appeals to a set of possible causal trajectories and initial conditions which all converge on the same point (Rice 2015, 596-598). The *lack* of information about initial conditions and actual causal trajectories provides crucial modal information that would be lost by including these details; and the absence of such details shifts our focus to what is relevant for explaining the ball's final resting state (ibid). The next section will detail how teleological explanations end up differing from equilibrium explanations. For the movement, I wish to highlight that if one accepts that equilibrium explanations are both non-causal and legitimately explanatory, then it one cannot disqualify the EA's teleological explanations on those grounds. Indeed, the explanations in both cases seem to share some key features; though as we shall now see, this may pose yet another problem for the EA.

4.4.5 Teleological and Equilibrium Explanations

The content of the EA's teleological explanations seems to share some key features with equilibrium explanations, particularly the fact that both are insensitive to

---

[56] As Rice makes clear, some equilibrium explanations such as optimality models may *also* contain idealizations (2015).

initial conditions and certain causal details about the precise path which was taken toward the equilibrium or goal. But while this helps the EA make the case that its teleological explanations are not merely descriptive, it also threatens its claim that teleology is a *distinct* form of (non-causal) explanation. Indeed, the proximity of teleology to equilibrium explanations (via the principle of least action) has a long history since at least the time of Leibniz (Mcdonough 2009; McDonough 2020). And in fact, the purposive behaviorist, cybernetic, and teleonomic accounts of teleology seemed to have often blurred the lines between equilibrium and causal explanations. The various analogies which purposive behaviorists made between organismal behavior and equilibrium processes offer a clear case of the former, particularly Holt's analogy between purposive behavior and orbiting planets. The same is true of the cybernetic account of purposive behavior as one in which terminates in "a definite correlation in time or in space with respect to another object or event"; a definition which ostensibly includes equilibrium points. And as Bill Wimsatt has argued, the same is even true of the subset of behaviors dubbed "teleological", i.e., those governed by negative feedback: from a behavioristic mode of study, behavior controlled by negative feedback is indistinguishable from the behavior of an open system tending toward equilibrium (1971; Bedau 1992). More recently, Dan McShea and colleagues have argued for the 'field theory' account of goal-directed behavior that takes goal-directed behavior to be driven by fields in a way that is intentionally analogous to gravitational fields that might be appealed to in equilibrium explanations (McShea 2012; Lee and McShea 2020; Babcock and McShea 2021; 2023). On this basis, one might argue that the EA's teleology is also a kind of equilibrium explanation. A ball rolling toward the bottom of a bowl may be said to behave robustly and plastically toward its goal, reacting to 'affordances' in the form of bumps and divots in the bowl, to which it can 'adapt' to by flying and bouncing.

Proponents of the EA may respond to this by denying that equilibrium explanations are veritable consequence explanations: the equilibrium point is not a consequence of the dynamics of this system, but rather prefigured from the outset. But even if it is true that equilibrium explanations are consequence-based, that would not be sufficient to call them teleological. Such a claim would underestimate how demanding the content of a teleological explanation is; though asserting this simultaneously brings the issues discussed in 4.5.1 to the fore. For the EA, there are good reasons to treat the explanation of a mouse's behavior making its way toward a piece of cheese at the bottom of a bowl differently than a rolling ball. Much like William James' contrast between the behavior of iron filings attracted to a magnet and Romeo's behavior toward Juliet, the modal profile of each explanation is markedly different. The ball's behavior, while insensitive to initial conditions, is always fixed; it will always take the path of least resistance toward the bottom of the bowl. In contrast, a teleological explanation of the mouse' behavior would include the fact that it could have done otherwise; indeed, the mere fact that it will move toward the cheese rather than attempt to escape the bowl is not guaranteed. Granted, one may insist that if determinism is true, then this particular mouse in that particular context was always going to pursue the goal of eating cheese, and that this goal was in some sense 'fixed' at the outset. But even so, given the kind of entity which the mouse is—an agent—it remains nomologically possible that it could have had another goal. In contrast, given the laws of physics as we know, there is no way the ball could have had any other "goal" than to roll toward the equilibrium point.

While the behavior of the ball is deterministic and 'closed-ended', the mouse's behavior is 'open-ended' in the sense that it will take normatively appropriate actions based on the circumstances (Winning 2023; see also: Longo, Montévil, and Kauffman 2012; Montévil 2019). It is precisely because the mouse has a wide range of possible behaviors that we can normatively judge its actions: it *could* have done otherwise, and perhaps it *should* have done so. None of this is appropriate in the case of the ball, because it simply could not have done otherwise. For this reason, James' assertion that Romeo's goal is "fixed" is a useful albeit misleading idealization when explaining behavior. For while agents may show remarkable perseverance in their pursuit of a goal, they also have the capacity to modify their goals—or reorder their priorities, if they have multiple—based on a changing situation. And these changes may well have been brought about precisely by the original pursuit of a goal. Indeed, this seems part and parcel of what it means to *achieve* a goal: to change a situation to the agent's satisfaction. This is relevant in the case of the mouse: while a ball can do nothing but fall to the bottom of the bowl, a mouse may well stop in the middle of its search for the cheese and jump out of the bowl if it becomes frightened. The open-endedness of agential behavior is therefore not only a fact of the way it pursued goals, but also about the way it can switch *between* goals. The next chapter shall argue that agents are not merely goal-directed systems, but ones that pursue many goals that cohere into a way of life.

4.6 Conclusion

In this chapter I contrasted two of the most well-developed approaches to naturalizing teleology based on an organism-centered view of agency. Both the AT and EA take the intrinsic purposiveness of organisms to provide the necessary condition for teleological explanations but conceptualize those conditions differently. However, the AT does much more to flesh out the causal conditions—closure of constraints—under which those conditions are to be met. This, in turn, impacts the ways they conceive of the content of such explanations. AT sees teleology as a form of circular causality where causes and effects are inextricably linked, if not conflated, such that consequences explain *qua* causes. The EA sees teleology as a distinct form of explanation in which goals explain. I showed how each approach must deal with a set of tradeoffs. The AT's conception of teleology is causal, which is a virtue according to an orthodox conception of naturalism. But as a result, self-maintenance becomes part of the *content* of its teleological explanations; and it can therefore only apply teleological explanations to actions and artifacts if it shows how those are directly or indirectly self-maintaining too. In contrast, the EA assumes a different kind of naturalism that may seem weak by the standards of the AT, though it allows the EA to formulate a notion of teleology that is more broadly applicable. It is perhaps unsurprising to notice that the disagreements at the heart of these difference—naturalism and what counts as proper scientific explanation—are the same philosophical problems which the purposive behaviorist, cybernetic and teleonomic accounts struggled with, as discussed in the first two chapters of this dissertation.

If I am right in framing the AT and EA's views in terms of a series of tradeoffs, then there are two ways for each approach to advance on these issues: one can either attempt to resolve the issues and thereby minimize the costs or inflate the gain by further elaborating on the benefits of each approach. In the next chapter, I will pursue the latter strategy with respect to the EA. I will show how its account of goal-directedness would benefit from being conceptualized in terms of how goals contribute to the *way of life* of an organism. I will then show how the notion of 'way of life' provides the metaphysical foundation for a unified conception of artifact and biological function.

CHAPTER FIVE

Ways of Life

The previous chapters have argued that the existence of organisms as intrinsically purposive systems is a necessary *condition* for a teleological explanation. However there remains disagreement among scholars over how that condition should be conceived and how that impacts the *content* of a teleological explanation. The Autonomy Theory (AT) assumes that a proper explanation picks out causes in the world, and that a teleological explanation must therefore be a species of causal explanation. The AT does this by characterizing teleological explanations as a kind of circular causal explanation where existence of an entity is explained by some of its effects. I dubbed these 'causal-consequence' explanations, because the consequence or effects explain *qua cause*. In subscribing to this view of teleology, the AT avoids standard objections to the use of teleological explanations, including the problem of backward causation. An alternative strategy, defended by the Ecological Account (EA), is to consider teleology a *non-causal* explanation that tracks empirically observable counterfactual dependencies. In this view, the status of teleology as a naturalistic form of explanation is justified by its explanatory indispensability, rather than its status as a species of causal explanations.

My aim in this and the next chapter is to make a case for the EA by building out the metaphysical foundations of its conception agency and demonstrating the potential payoffs of the view. My motivation for doing so is two-fold. First, I wish to highlight the fact that the EA shares much in common with purposive behaviorism and inherits many of its best features. At the same time, it also avoids the major pitfalls of those past attempts at naturalizing purpose. Most notably, the EA shows that a behavioral view of purpose need not be wedded to an artifact model. Second, the EA stands apart from the AT and other accounts in its focus on making sense of the rich content of teleological explanation. The primary importance of this task has remained underappreciated. In what follows, I will argue that the tendency to characterize purposive behaviors in isolation from the concrete way of life within which they originate is a useful albeit misleading idealization. In Chapter 6, I will argue that despite failing to do so in the past, an account of teleology based on a behavioral account of goal-directedness can produce a satisfactory account of artifact and biological functions.

If we are to take seriously the purposive behaviorist's methodological claim that theorizing teleology begins with the phenomenon of biological agency, then it we must expand our conception of what this phenomenon consists of. Rather than being about singular goal-directed behaviors—which experimental behaviorist psychology intentionally isolated—biological purpose is about the ways in which a plurality of mutually supporting goal-directed behaviors bring about a *way of life*. But such a *methodological* commitment also implies certain *metaphysical* commitments regarding the nature of agency. To pursue a way of life is not merely to display goal-directed behavior but also to individuate or constitute oneself as the kind of creature that pursues this particular way of life.

In making such a claim, I am effectively suggesting that we reconsider Hans Jonas' idea that organisms (or more generally, agents) are goal-directed in that their "being is of their own doing". Contrary to the AT's interpretation, I suggest that Jonas' notion of "being" seems to commit him to more than mere existence, but a notion of "*existence as*" a particular kind of living thing. As I see it, Jonas is right to think that intrinsic purposiveness is tied to a metaphysically substantive notion of being, though I will also follow the AT in thinking that we need not appeal to a phenomenological grounding for such a view.[57] Rather, one should

[57] For Jonas, this more specific notion of "being" is fundamental and derived from phenomenological reflection on our own experience as such beings. As we have seen, he holds the

appeal to the concepts that one needs to make sense of the latest findings in biology. Specifically, I will take the position that "being" or "existence as" is a far more metaphysically loaded concept than mere "existence", as the AT would have it, insofar as it presupposes some substantive facts about the entity in question, specifically, that it is an entity with a particular *way of life.* Roughly speaking, the enriched conception of "existence as" *qua* way of life serves as an answer to the question: "What kind of entity is it?" In turn, one sees that this is an altogether different—and by no means incompatible—question than the one the AT seeks to answer: how is the existence of biological organization causally possible? In this regard, the contrast between the EA and AT will be further clarified.

I will argue that the notion of way of life allows for a distinction between two senses of purpose, which I will dub "goals" and "ends" after Edward Stuart Russell (1945). Goals refer to the aims of goal-directed activities pursued by an agent on the basis of its behavioral repertoire and affordances, which is already detailed by the EA. Ends, in contrast, refers to the aim of pursuing a particular way of life. Consider that of a squirrel, which includes activities such as shivering, licking wounds, burying acorns, fighting off rivals, seeking shelter in trees from passing dogs, and so on. Taken together over a long enough period of time, these bundles of goal-directed activities cohere and reliably bring about a type of end, namely that typical of a squirrel way of life. In this manner ends i.e., ways of life, individuate entities, and in doing so give us a reliable indication—though not a determination—of the kinds of goal-directed activities that such entities might be likely to be found doing.

Goal-directedness and end-directedness will be further characterized on explanatory grounds. I shall argue that goals and ends, while both teleological, figure in explanations in different ways. Goals may be a legitimate *explanans* for questions pertaining to why a particular process, change, or behavior is occurring. Ends, on the other hand, may be appropriate for answering why it is that an

---

view that this primordial experience provides the grounds on which organisms show up to us as kindred perspective-taking, teleological agents.

organism has a particular coordinated set of features, and why those *constitute* the agent as the particular kind of agent that it is. I argue that such 'constitutive explanations' are a particular kind of teleological explanation. And while the strongly essentialist flavor of this mode of explanation may seem anathema to contemporary biology, recent neo-Aristotelian accounts of essentialism have argued that their position is *supported* by—rather than in conflict with—contemporary evolutionary developmental biology. In this view, ways of life are characterized as the *realization* of the teleological dispositions that are essential to an organism. This argument is intended to clarify the fundamental difference between the intrinsic purposiveness of organisms, and extrinsic purposiveness of artifacts. The purposes and functions of artifact are metaphysically dependent on organismal essences.

The structure of this chapter is as follows. In section 5.1, I briefly discuss the ways in which the EA might be located within the purpose behaviorist tradition, while highlighting its potential for overcoming these problems. In section 5.2, I will then argue for a characterization of the phenomenon of intrinsic purpose in terms of the pursuit of a *way of life*. In section 5.3, I will distinguish goals and ends on explanatory grounds. Finally, 5.4 will argue that insofar as the notion of a way of life and an end-explanation implies an essentialist commitment, this need not be in conflict with contemporary evolutionary biology.

5.1 Goal-directed Behavior and Ways of Life

In the previous chapter, I hinted at some ways in which the EA can be situated within a broader empiricist and pragmatist tradition within analytic philosophy of science and philosophy of mind. This is most evident in the way that the EA, like the purposive behaviorism, characterizes the phenomenon of purposiveness first and foremost in terms of behavior. Moreover, one of the major inspirations from the EA comes from the work of James Gibson, whose work can be situated in the confluence between the realist conception of perception which he inherited from the New Realist movement that comprised purposive behaviorists such as Edwin Holt—who was Gibson's PhD supervisor—and Ralph B. Perry (Shaw 2011); the Gestalt psychology of Kurt Koffka; and the philosophy of embodiment of French philosopher Maurice Merleau-Ponty (Chemero and Käufer 2016 see also; Baggs and Chemero 2018).

At the same time, I take it that the EA provides the conceptual resources to overcome many of the problems with purposive behaviorism discussed thus far. The one which I will focus in this section is the way purposive behaviorism tended to characterize the phenomenon of purposive behavior by isolating behaviors into 'units' of goal-directed behaviors, themselves built up, as Holt suggested, from the coordination of individual reflex arcs. In this regard, artifact analogies and models such as the thermostat provided a simplified albeit misleading picture of purposiveness, which abstracted it away from the concrete way of life of the organism. For one obvious way in which a thermostat or torpedo is quite *unlike* an organism is that its goal (if it even has one) is singular and fixed (Canguilhem [1952] 1992). As discussed in the previous chapter, a mouse's goal-directed behavior toward a piece of cheese at the bottom of a bowl may be superficially like the behavior of a ball rolling down a bowl. But what this comparison misses is what the mouse *could* have done, but didn't, such as decide to do something else entirely. In evaluating this analogy, one must not only look at what actually happened but what *could* have. I will argue that doing so requires that we attend to the space of possibilities carved out by a way of life which, following E.S. Russell, I will call an *end*.

5.2 Self Constitution, Unity, and Way of Life

5.2.1 Aristotle on Bios and Self-Constitution

As we've seen, the AT favorably cites Jonas' claim that intrinsic purpose consists in the fact that an organism's "being is its own doing", which they take to mean that an organism "exists because of what it does" or the Kantian notion of being "cause and effect of oneself". An alternative interpretation of this claim adverts to a compelling idea that goes back at least to Aristotle, namely, the idea of *self-constitution* as a fundamental aspect of the purposiveness of living beings. In this view, organisms are self-constituting in that their end is to continue being the kind of entity which they are (Korsgaard 2009; Schapiro, Ebels-Duggan, and Street 2022). It is from this vantage that we should interpret Aristotle's claim that "...in the case of living things, their being is to live", or Jonas' claim that an organism's being is its own doing (Barnes 1984, vol. 1, l. 415b13; Jonas [1966] 2001; Toepfer 2021). In both cases, "being" and "living" are metaphysically located concepts that point to the idea that organisms do not merely strive to go on existing in any viable way, but to persist in existing *as* the particular entities that they are. In the words of Jonas, the question is not "duration as such", but "duration of what?" ([1966] 2001, 106). This particular relation between identity and agency is at the heart of the idea that the intrinsic purposiveness of agents constitutes them as the kinds of entities that they are.

Acknowledging the centrality of self-constitution gives some credence to the AT's claim that there is a particular teleological activity that is more fundamental than the others, while avoiding the AT's restriction of goal-directedness to the realization or perpetuation of one's existence. From the EA's perspective, it does not necessarily follow from the fact that organisms are self-organizing and self-maintaining that the meaning of intrinsic purpose "just is" self-maintenance and nothing else. It is in this sense that organisms can be said to have the goal of persisting qua *self-constitution* rather than qua *self-maintenance*: existing *as* the kind of thing that one is. Consider, for example, that to be a squirrel is not merely to look a certain way or to behave in any one distinctively squirrel way but to have the capacity to pursue a whole range of particular goal-directed activities that cohere in a particular way. These activities begin in embryonic development and include the myriad activities that we commonly observe such as shivering, licking wounds, burying acorns, fighting off rivals, seeking shelter in trees from passing dogs. The ways in which those activities cohere and constitute the agent as a particular kind of thing is best understood in terms of a *way of life.*

According to Aristotle, the particular aspects of ways of life that are usually most salient are those relating to where, how, and when organisms develop, feed, and reproduce (Lennox 2010a; 2010b). Way of life or *bios* is a somewhat underappreciated concept in Aristotle's thinking, but is nonetheless philosophically rich and remains relevant to this day. According to Lennox, ways of life play a conceptual role in "tying many different modes of activity to an animal's overall way of interacting with its environment", and play a key role in the way that we identify different kinds of organisms (Lennox 2010a, 337). Of particular importance is the unificatory role that ways of life play (ibid; c.f. Russell 1924). A way of life is not merely one or many particularly distinctive activities, nor an entire list of all an organism's activities over its lifetime, but the way those activities are unified and deployed in a coordinated and coherent way at different stages of their life cycle.[58]

This coherence has a teleological and normative profile, in the sense that the different behaviors and traits of an organism acts as means of each other. Hawks are able to fly high above the ground to identify their prey thanks to numerous adaptations that allow them to do so, such as keen eyesight. The various traits that enable the hawk to identify prey from far away are means to that goal; but that goal is itself a means toward hunting and feeding, and can be explained as such (Moosavi 2022, 81–82). There is, as Moosavi argues, a pattern of counterfactual dependence between these traits such that they can be explained as a means—or, as shall be argued in the next chapter, as serving function—toward the realization and persistence of other goals that are part of an organism's way of life. In the words of Lennox, attending to ways of life enables on to notice "the way in which

[58] Though the fact that a way of life is unified does not mean that we cannot subdivide it along several axes, such as the way that E. S. Russell distinguishes between the ends of maintenance, development, and reproduction. "the fact that a way of life is divisible along a variety of axes that make reference to activities does not entail either that it is reducible to those activities or that it is non-unitary." (Lennox 2010a, 337).

the many exquisite adaptations found in a particular kind of organism, at whatever level of generality one is investigating, are not only adaptations to particular functions, but adaptations to one another" (Lennox 2010b, p. 257).

5.2.2 Bios and Contemporary Evolutionary Biology

As we have learned from contemporary developmental biology, the manner in which traits are adapted to each other is not prespecified by a 'genetic program' but the realization of an active and plastic process of development that is orchestrated by the organism as a whole. This is most clearly highlighted by abnormal cases, such as that of a goat born without functioning forelimbs that learned to move with its two hind legs (West-Eberhard 2005, 6545). The goat in question developed a series of unexpected behavioral and physiological adaptations to walk on two legs, including changes to the structure of the hind legs, pelvis, and spine with corresponding changes in its musculature and tendons (ibid). This process, known as phenotypic accommodation, allows novel phenotypes produced in response to particular internal and external conditions that have potentially catastrophically disruptive effects on the organism as a whole. But the plasticity of development buffers the organism against these. This adjustment to phenotypic change is crucial to development insofar as it reduces the functional disruption caused by developmental novelties (West-Eberhard 2003, 147; Pfennig et al. 2010; Levis and Pfennig 2020; Uller et al. 2024). But it also promotes the kind of adaptedness of traits to each other that enable them to coordinate their activities in a way that is conducive to the realization of various goals. As such, the two-legged goat's "anatomy is simply not a means toward walking on four legs, just as it is not a means toward flying or swimming. It's rather a means toward walking and hopping on two legs" (Moosavi 2022, 97). In doing so, the goat has found a way to live a goat's way of life; and this discovery should make us wary of thinking that such a way of life is pre-given (I shall discuss this further in 5.4.2). Important in this context is the fact that various adaptations which the goat developed are explained by the goal of walking and feeding on two legs because those traits counterfactually depend on those goals and the way of life they conduce to. But importantly, the goat did not adapt to its bipedal lifestyle in an arbitrary way; rather it did so in a way that is deeply tied to its myriad other goals. Here we are not only interested in asking how the goat's organization exists in a causal sense—though the AT would give us an account of that—but how it continues to exist *as* a goat. Indeed, one shouldn't think that bipedalism is contrary to a goat's way of life—rather, it is evidence that there are many surprising ways for goats to pursue it (a point I will further elaborate in section 5.4.2).

The coordinated integration of various structures and activities make organisms suited to living in particular habitats. Bats, flying insects, and birds are all adapted to a similar air-born way of life, but due to morphological constraints in their evolutionary history have adapted to it in different ways (Lennox 2010b, 256). In order to distinguish these ways of life, or even more fine-grained differences in the ways of life of closely related species, we must bring both of these facts about organism morphology and habitat together. The fruits of this kind of broad integration are a particular way in which we gain understanding. A way of life "explains why that is the kind of food these birds must be adapted to eating, and thus why they must have this specific sort of beak. Not all birds have *flat* beaks; not even all swamp-dwelling birds do, since for those who eat frogs and small fish a flat beak would not be useful" (Lennox 2010b, 250; see also: E. S. Russell 1934b, 15).

Given the current orthodoxy in evolutionary biology, one may think that a way of life implies the outdated notion of biological essentialism or, in line with Pittendrigh's teleonomy, is synonymous with adaptation and therefore fully explicable in terms of natural selection. On the latter view, organism's activities are integrated because this integration is fitness-enhancing insofar it increases the chances of survival and reproduction. Samir Okasha, for instance, assumes that the core of 'agential thinking' in biology lies in an analogy with rationality, where the organism displays a 'unity of purpose' "in the sense that its evolved traits contribute to a single overall goal", namely, enhancing fitness (2018, 5; 32; see also;

Kennett and Matthews 2003). As he argues, there are many cases in biology in which such a unity fails to obtain, often due to intragenomic conflict. For instance, normally hermaphroditic flowering plants will sometimes fail to produce any pollen (which produces male gametes) and therefore solely produce the ova that makes female gametes. This occurs because of certain mitochondrial genes, which are only transmitted maternally, and persist in populations because they favor their own transmission. In such a case, we ought to think of those genes as having the function of facilitating their own transmissions, since they were selected for that purpose, and yet "we cannot sensibly think of the plant itself as having this as its goal, for it is detrimental to it, preventing it from selfing, and the plant has other traits which conflict with this goal." (Okasha 2018, 31). As such, agency characterized in terms of various activities that display a 'unity-of-purpose' conflicts with contemporary evolutionary biology (see also Ågren 2016).

While it seems true that unity-of-purpose is a regulative ideal for rational agency, and perhaps agency more broadly, it is unclear why one must assume that agents must meet such a steep criterion to be agents at all. Rather, unity-of-purpose describes *effective* agency, not agency simpliciter. The integration and coordination displayed by organisms need not be fully unified, or completely free of conflict, to be viewed teleologically. Indeed, internal tradeoffs and conflicts in most organisms show up against the background of their otherwise remarkable multiscale integration, which is a veritable achievement of every single developed organism.

What cases of conflict highlight is how much more conflict, disunity and sheer chaos there *could* be, but isn't. If it were true that organisms follow a predetermined and deterministic developmental program, then even small errors in the 'code' (beyond what can be buffered by redundancy) would lead to disastrous outcomes. But as discussed above, developmental biologists have shown that organisms display a remarkable capacity to adapt to unpredictable internal and external conditions. This work also highlights the many ways in which biologists have attempted to integrate ecology with the study of the structures and functions of organisms at various life-stages, such as in ecological morphology, or more recently, ‘eco-evo-devo’ (see e.g., Abouheif et al. 2014; Gilbert et al. 2015; see also Russell 1924). It is now becoming evident that organisms have the capacity to develop a wide range of adaptive phenotypes according to variations in their environment, and that this plasticity may even be a driver of evolution (Sultan 2017; Perry et al. 2018; Levis and Pfennig 2020). In this way, the agency of organisms is a *precondition* rather than a result of their evolution (Toepfer 2012; Walsh 2015).

5.2.3 Goals, Ends, and the Autonomy Theory

The foregoing indicates that while the notion of a way of life may not be familiar to contemporary biology, it is not in conflict with modern evolutionary theory. On the contrary, much recent work lends credence to it. In the following section, I will argue that the way organisms are purposively directed toward their way of life is importantly different to the ways that activities are directed toward goals. To make this explanatory difference clear, I will borrow the distinction between ‘goals’ and ‘ends’ from the organicist E. S. Russell (1945). Goals are those states of affairs that activities are directed towards, and which can be adequately characterized by the EA’s existing framework, namely, as part of a conceptual cluster with affordances and repertoire (see e.g., Russell 1945, 80). Ends, in contrast, correspond to the directiveness toward a particular way of life; they identify an overarching goal of the system that explains the coherence of its parts and activities. These parts and activities cohere to bring about that way of life, typified by patterns of survival, reproduction, and which constitute the organism as the kind of thing that it is.

The relationship between goals and ends marks an important contrast with the AT. As discussed in the previous chapter, the AT defines goals in causal-etiological terms, i.e., in terms of an entity’s causal powers that explain its existence. This means that continued existence is, by definition, the only goal there can be—even if there are innumerable ways to achieve it. In this way, the relationship between goals and existence, according to the AT, is definitional. One may think the same is true of goals and ends: to be a goal is to be a means to an end or way of life. This assessment would be incorrect. It is not my intention to reduce the EA's definition of goals in terms of behavior, repertoire, and affordances to ends. Rather, my aim is to *situate* that definition of goals within the pursuit of a way of life. In this sense, the notion of goals laid out here is definitionally independent of ends even if, as a matter of fact, when one observes the goal-directedness of organisms, it always appears directed toward a particular end or way of life.

Another contrast between the AT and the version of the EA I am pursuing here is that, in my view, an activity may conform to a way of life but may not be in the service of self-maintenance. This is because a way of life consists of the particular life history strategies of an organism, which includes particular ways of ageing, reproducing, and dying. Consider the case of anadromous fish, such as salmon, that migrate from saltwater to freshwater to spawn. This process is accompanied by significant physiological changes which includes storing of energy reserves and redirected energy to the development of reproductive organs. These processes are directed toward the goal of reproduction, which typically ends with the salmon's death shortly after spawning. Such a series of behaviors are ultimately self-undermining for the fish in question, and yet they are purposive. This is where a shift of perspective from self-maintenance to a way of life is crucial: from the point of view of a fish's life cycle, this particular manner of dying is a natural part of the salmon's way of life—i.e., what it is to exist as a salmon—and for reasons that will soon become clear, the directiveness of this process of self-constitution can be explained teleologically.

To summarize, organisms have as their fundamental end to persist as the entities that they are: squirrels strive to remain squirrels, and being a squirrel is therefore not a state but a continuous activity. To say the squirrel pursues a way of life rather than the mere persistence of its existence implies a complex suite of capacities necessary to carrying out the activities that make up that way of life. At the same time, the notion of way of life vindicates the AT's claim that there is a particular purpose which is more fundamental than the others, though I've interpreted this in terms of self-constitution rather than self-maintenance. This adds an important ingredient to the EA that is currently missing at present, and moreover, is at least one way that I take the EA to have not fully overcome one of the shortcomings of purposive behaviorism, namely, its failure to account for ends as that which metaphysically grounds goal-directedness.

5.3 How and What Ends and Goals Explain

The foregoing suggests important ways in which the pursuit of a way of life, or *end*-directedness, is irreducible to the *goal*-directedness of activities. In this section I will do two things. First, I will further clarify the distinction between goals and ends by highlighting the differences in both *how* and *what* they explain. Second, I will argue that despite being different from goal-explanations, end-explanations are genuinely teleological.

5.3.1 Ends and Coordination

As we've seen, ends consist in the *coordination*, rather than the mere summation, of myriad goal-directed activities. This has implications for the understanding that we gain when we situate a goal-directed activity in the larger end-directedness of an organism. While an activity can be understood as an isolated response to an environmental stimulus, we learn something new about this activity when we learn how it contributes to a way of life. This closely tracks Alejandro Fábregas-Tejeda's recent suggestion that biological agency, for Russell, is indispensable for understanding living beings:

> to understand the action of the wasp on the gate post you have to integrate this action into the whole directive cycle of activity […]. This is the sort of knowledge which we get by studying animals from the [agential] point of view, and it is extremely valuable knowledge too.

> Without this knowledge we simply cannot make sense of behaviour; even the most extreme believer in mechanism must use this sort of knowledge to make behaviour intelligible at all.
>
> (Russell 1934a, 15; cited in Fábregas-Tejeda 2024, 53)

Granted, Russell may be overstating his case here. It seems implausible to suggest that any information we have gained from mechanistic biology either implicitly presupposed ends or provided no knowledge at all. Alternatively, Russell may mean that one can only recognize or classify an activity *as* goal-directed behavior if we recognize that the behaving agent in question has an end. I take this second interpretation to be more plausible. Indeed, such knowledge would be extremely valuable because it is unifying. It tells us why certain activities belong together in virtue of their conduciveness to an end. In so doing, it also tells us why an organism has a particular set of features. Indeed, as discussed in the previous section, it is not clear if it would be possible for organisms to display robustness and phenotypic accommodation in the face of perturbation if they were not integrated in this way. And while a teleological explanation in terms of ways of life may well be greatly enriched by information pertaining to the selective pressures which an organism's ancestors faced, end-explanations do not reduce to such evolutionary explanations for the reasons outlined in the previous section: coherence may be fitness enhancing, but it has its origin in the agential dynamics of developmental systems.

In one respect, the kind of understanding produced by the end-explanations is the same as that produced by any kind of teleological explanation. This is a kind of understanding which we gain from all teleological explanations, insofar as they tell us why what otherwise appears to be a purely contingent collection of facts have a particular kind of necessity. This point was emphasized by both Aristotle and Kant (J. Cooper 1987; D. Charles 1988; Ginsborg 2001, 248; Quarfood 2006, 738; Huneman 2014; van den Berg 2013, 4). As discussed in the introduction to Part 2, teleological explanations produce understanding in part by showing how the various contributing factors to the achievement of the goal coordinate in the way that they do. But if this is true, then it might seem odd to attribute to ways of life a particular form of understanding on that basis. One difference between ends and goals is that ends unify a particular set of phenomena, namely, goal-directed activities. This unification produces understanding of the *identity* of the organism in general, in a way that the explanation of a particular behavior does not—even if it does produce a kind of unifying understanding of why the movements which constitute that behavior coordinate in the way that they do. We learn something substantive when we are told that an activity occurred because it contributes to a particular way of life.

5.3.2 What Ends Explain

The foregoing establishes one way in which goal and end explanations are different: while both are unificatory, they unify different sets of phenomena and in virtue of that confer different kinds of understanding. But this does not fully account for the explanatory differences between goals and ends. To see this, consider what happens when one mistakenly appeals to an end when one should appeal to a goal. For instance, instead of explaining why the Caribbean reef squid, *Sepioteuthis sepioidea,* secreted ink by citing the goal of escaping its predator, one might try to explain this behavior in virtue of the squid's end. In this case, we are explaining that the squid secreted ink 'in order to be the kind of entity that it is, namely, a Caribbean reef squid'. This seems to be problematic in two respects. First, it is totally uninformative. Second, it is not even clear if the locution 'in order to' is even appropriate in such a context. I will address both points in turn.

First, when one asks why an organism performed a particular action, what is explanatorily salient is what did not happen but could have, i.e., different outcomes had the organism had a different goal, affordances, and deployed its behavioral repertoire differently in that same situation. This level of granularity is not achieved by the invocation of a way of life, because it does not tell us why the agent did one thing rather than another. In fact, having a particular repertoire at a particular life-stage is a feature of an organism's way of life; and if so, way of life alone can't answer the question why an organism actuated one element of its repertoire rather than another.

The necessity for such a contrast is not particular to teleological explanations (Dretske 1972). As Peter Lipton argues, the causal history of a phenomenon is "long and wide" and for the most part this history would be uninformative when explaining why the phenomenon happened. Indeed, every event is caused, in a trivial sense, by the big bang. But this amounts to an explanation for very few phenomena (1990, 249). Any causal or teleological explanation requires explaining both what happened and a foil, i.e., what could have happened but didn't (see also: Woodward 2005, chap. 3; Gijsbers 2018). In this sense, the question "why P?" is not the same as "why P rather than Qi,...Qn"; and usually when we are asking the former, we are actually after an answer to the latter. The latter may be informative while the former is not. For those reasons, ways of life are not appropriate for explaining behaviors or actions. To cite another example, many regulatory mechanisms may perform one action or its opposite (e.g., speed up or slow down metabolism) and doing so in the appropriate context is essential to the pursuit of a way of life. Since doing either may conform to an end or way of life, citing such an end does not help explain why one or the other occurred.

The second reason to think that it is inappropriate to explain an activity by an end (i.e., to say an entity acts to be the kind of entity that it is) has to do with the use of the 'in order to' locution. To parallel the squid case above, consider the explanation of a person's act of deliberation. If an agent correctly deduced Q from P, given that "if P then Q" is true, then it makes little sense to say the agent came to this conclusion *in order to be rational.* And yet it is acceptable to say that an agent who rejects those beliefs that do not logically fit with their other beliefs does so because it is rational to do so. In both the squid case and the rationality case, we have explained by identifying behaviors as belonging to a set of coordinated activities that bring about particular ways of life; but we have done so without claiming that the agents acted in order to bring those ways of life about.

That being said, there may still be a way to make sense of the divergent use of the 'in order to' locution. Aristotle suggests two meanings for the phrase: it can either mean “that on account of which” or “that for which” (Aristotle 2016, 415a24-b7). When one explains a particular activity as being for the sake of a goal, one is citing “that for which” it occurs. In contrast, the squid can be said to have secreted ink to escape the predator “on account of” the fact that it is a Caribbean reef squid. Similarly, an agent may have deduced Q from P *on account of* the fact that the agent is a rational being. And in both cases, we have a sort of constitutive relation. Being rational is constitutive of being a thinker but may not be a direct goal of any particular activity; and having a particular way of life is constitutive of being an organism of a particular sort, even if it is not the goal of any of the activities which constitute it. In both cases, one might say that the squid and the thinker acted as they did because they have an intrinsic capacity to do so, and it is that capacity which accounts for the possibility of that behavior. An explanation in these terms cites that activity as an essential part of the kind of agents these entities are.

In sum, goals and ends may be distinguished on explanatory grounds in the virtue of the kinds of questions that they are appropriate for answering. As discussed in the previous chapter, goals may be a legitimate *explanans* for questions pertaining to why a particular process, change, or behavior is occurring. Ends, on the other hand, may be appropriate for answering why it is that an organism has a particular set of features organized or orchestrated in the way they are (Gelber 2021, 98).[59] Ends or ways of life are ill-suited to explain a particular action or behavior for two reasons. First, it would not be particularly informative to say that an organism did X “in order to be the kind of entity that it is”. It may be true in the sense that doing that activity does in fact contribute to the organism’s pursuit of its way of life. But such an explanation would lack explanatory power because it doesn’t delineate the right contrast class. Second, to teleologically explain why a particular behavior is occurring is to say “that for which” it occurs.

---

[59] I am not committed to goals always being the relevant explanation for the occurrence of *any* process, change, or behavior; there may be myriad cases where the most appropriate explanation is of a non-teleological kind, whether it be a species of causal explanation or non-teleological, non-causal explanations like equilibrium explanations.

Ends, in contrast, explain that "on account of which" a behavior occurs. This makes an end-explanation a *constitutive* explanation.

5.3.3 Are End-Explanations Really Teleological?

That end-explanations are constitutive speaks to a recent disagreement between Lennox and Jessica Gelber on whether Aristotle understood way of life to be teleological. Lennox argues that the activities essential to a particular organism are not for the sake of the way of life. He reasons that, at least for Aristotle, such a claim would conflate ways of life ("bios") and the hierarchically organized set of intrinsic teleological capacities that make up the soul ("psyche") (Lennox 2010a, 349–50). But as Gelber rightly argues, this implies a non-teleological interpretation of way of life which renders its "causal and explanatory role [...] somewhat elusive." (Gelber 2015, 282). I would suggest that both are right; activities aren't for the sake of *bios*, but that doesn't mean that bios or end-explanations aren't teleological.

As previously discussed, an end is like rationality: we do not act in order to be rational, but because it is a good way for us to resolve problems given the kinds of entities we are. Another example is health. As Christine Korsgaard argues, "*health is not, strictly speaking, a goal for giraffes* [or other organisms], but rather is our name for the inner condition which enables the giraffe to successfully perform her function—which is to go on being a giraffe" (Korsgaard 2009, 35; see also: Toepfer 2021). As discussed in Chapter 3, Canguilhem also held the view that health is a capacity rather than a state. But if such ends are not goals, how can end-explanations be teleological? As argued in the previous chapter, the most basic feature of such explanations is that they explain a phenomenon in virtue of a consequence it brings about. Ends are also consequences because the way of life that characterizes an agent is always in the process of being actively realized in a processual, future-orientated way. So too with reason and health, which are internal "conditions" or capacities, but being a rational or healthy creature is a consequence of multiple activities cohere in particular ways over time; rationality and healthiness, like the good, are what a particular set of activities have in common. In the words of Jonas, such a "continuous identity of being is not assured by mere inertial persistence" but rather must be actively brought about (1968, 243). End-directedness is teleological in that sense. Moreover, end-explanations have a structure of hypothetical necessity: goal-directed activities are the hypothetically necessary means for the realization of ends. For that reason, they have the same normative implications. Living a healthy life hypothetically requires eating well and exercising often, among other things.

The foregoing adds an important dimension to the EA's account of teleology, which can be framed in terms of the distinction introduced in the previous chapter between the *content* and *conditions* of a teleological explanation. This distinction refers to the ways in which one characterizes the phenomenon of intrinsic purposiveness as a defining feature of biological agency and the logic of the kinds of explanations they require, and it applies equally to goal and end-explanations. Way of life expands what I take to be the necessary condition for such explanations to apply, in that it characterizes the phenomenon of agency not merely as a random assortment of directed behaviors but a coordinated suite of activities and structures. Furthermore, it distinguishes the content of goal and end-explanations as two kinds of teleological explanations. The latter tells us something about how an activity coheres with others, and how those bring about the way of life that constitutes the agent in question.

That goal-explanations and end-explanations appeal to two different, complimentary kinds of consequence suggests that teleology is a particular class of explanations that may include further types. The AT's account of teleology, discussed in the previous chapter, is one of these: in virtue of being an etiological, i.e., causal-consequence explanation, it supplies answers to the question of how and why a system exists in a causal sense. This is an altogether different question than that answered by goals and ends.

I take this to be an important advance over the purposive behaviorist accounts which repeatedly bracketed such considerations. In Chapter 1, I argued that the use of artifact analogies and metaphors were a way for scholars to show their naturalist colors; a way to show that one could talk about purposiveness without implying that there was anything other than the kind of mechanical forces that one would find in a centrifugal governor. This strategy carried over to the cybernetic account, which shied away even more from any metaphysical commitments. Indeed, the repeated assertions that purpose is merely "useful" and that it has a place within a particular "method of study", gave it a strong heuristic flavor. While showing how purpose and teleology may be useful for science is a legitimate philosophical project, it does not exhaust the scope of the problem (Nahas and Sachs 2023, 202). Whether one likes it or not, at stake in these debates is also a question about metaphysics: about whether agency and purposiveness have a place as part of the metaphysical furniture of our world. And if this is true, then it remains unclear what the adequate metaphysical foundations for organisms as agents with ways of life might be. I will turn to this issue in the next section.

5.4 Way of Life and Essentialism

This section sketches out the metaphysical groundwork which would support a view of organisms as intrinsically purposive beings. The pursuit of a way of life, I contend, emerges out of the realization of a set of intrinsic dispositions or capacities which one might call *essences*. While the very mention of essentialism may seem heretical, the notion remains highly misunderstood. My argument will rest on the kind of neo-Aristotelian essentialism that has recently been defended by Walsh (2006) and Austin (2017; 2018) among others on the basis of the kinds of findings in evolutionary developmental biology that have been discussed so far.

In the first section, I will clarify the relationship between teleology, typology, and essentialism to show why the standard criticisms of essentialism are based on a misunderstanding of Aristotle's biological project. The second section will flesh out what such a neo-Aristotelian essentialism might look like.

5.4.1 Typological and Teleological Essentialism

Biological essentialism has been much maligned by contemporary biologists, in no small part due to the influence of one of the architects of the modern evolutionary synthesis and proponent of teleonomy, Ernst Mayr. As he famously put it, Darwin's truly revolutionary contribution was to shift biology from "typological thinking" to "population thinking" (Mayr 1955; 1982, chap. 6; see also: Hull 1974; Sober 1980; Chung 2003; Witteveen 2015). The distinction is based on two different conceptions of biological species. The former categorizes species into natural kinds based on unchanging essential properties. Any apparent variation within kinds is deemed illusory or accidental (this was not Aristotle's view). In contrast, the neo-Darwinian conception of species sees variation as fundamental and species as evolving over time (Mayr 1982, 256). Not only that, but many argue that there is no single set of criteria by which we individuate species (Dupré 1993; Ereshefsky 2010).

Those who have since sought to rehabilitate essentialism have done so in ways that rethink the relationship between essence and typology. Broadly speaking, this has been pursued in two ways. On the one hand, some have attempted to rehabilitate essentialism by reinterpreting the concept to make it compatible with the evolution of species and pluralist views of taxonomy (Griffiths 1996; Boyd 1999). Such novel conceptions of essentialism, while interesting, have deemed it necessary to decisively move away from Aristotelian essentialism insofar as essential properties are taken to be extrinsic (relational and historical) facts, rather than unifying and intrinsic facts (Griffiths 1999; Okasha 2002; LaPorte 2003). On the other hand, scholars have aimed to show that Aristotle's views of biological essentialism have been badly misconstrued and do not carry a commitment to typological thinking (Balme 1987; Gotthelf and Lennox 1987). Aristotle was not a *typological* essentialist, but rather a *teleological* essentialist.

Perhaps the most serious misconception of Aristotle's project is that it is first and foremost a taxonomic one. This confusion has come about due to a conflation between two kinds of questions concerning taxonomy (Devitt 2008; 2023). First is a question about taxon-membership: "in virtue of what is organism *x* a member of species *S*?" Second is a question about taxon-categories: "in virtue of what is *S* a species?". In line with other essentialists, Austin contends that debates about species are predominantly about taxon-categories, i.e., they are debates about "what species is *qua* taxonomic rank" (Austin 2018, 29; see also: Grene 1972). But neo-Aristotelian essentialism is interested in the former notion of taxon membership rather than taxon categories; in the words of Christopher Austin, it is concerned with *telos* rather than *taxon*, i.e., *that in virtue of which an organism is a member of a kind* (Austin 2018, 22; Walsh 2006, 432) Crucially, neo-Aristotelians take such taxon-membership questions to require appealing to a set of properties that are both *intrinsic* to organisms and *causally responsible* for the morphological profile which typifies members of this kind (Austin 2018, 31-32). Christopher Austin takes such properties to be the intrinsic goal-directed behaviors which, I've argued, typify the agency of living beings. It is the coherence of goal-directed activities that make up an organism's nature: that in virtue of which it is a member of a kind in the Aristotelian sense (Walsh 2006). This strict requirement disqualifies by Aristotelian standards any taxon-membership criteria that one might derive from common taxon-categories, such as the capacity to interbreed, because those criteria are not sufficiently causally informative.

Ironically, Mayr appeared to have been close to understanding this himself. While he never explicitly linked his discussion of typological thinking with his discussion of teleonomy, he made many implicit connections between these issues. In what seem to have been an attempt to clarify his earlier comments regarding typological thinking (Mayr 1959), he argued that Aristotle's notion of *eidos* or form is fundamentally different to Plato's (Mayr 1972). And as discussed in Chapter 2, he followed Delbrück in claiming that the genetic program provides the material basis for the form-giving principle or *eidos* of an organism. And yet Mayr would never go so far as to claim that the genetic program provided a notion of essence that is intrinsic and yet makes room for evolutionary change. This may be because he was unable to see how essentialism could be extricated from typological thinking, or because his conception of teleonomy was an attempt to naturalize teleology without having to rely on essence. Indeed, it is a widespread belief that rehabilitating teleology means extricating it from any essentialist commitments, and almost all recent attempts to naturalize purposiveness from a biological point of view have done this.

As the above has shown, the connection between essentialism, teleology, and typology is by no means self-evident. What matters here is that a commitment to essentialism and teleology does not necessitate a commitment to typology. My aim in the remainder of this section is to elaborate a particular relationship between essentialism and teleology which provides the metaphysical foundations for 'way of life', which will in turn clarify what it is that we are adverting to, metaphysically speaking, when we cite ends in a constitutive explanation. I will not, however, argue 'way of life' *necessitates* this essentialist interpretation. That organisms have species-typical ways of life is empirically evident, but there may be multiple ways to metaphysically cash out this fact. And needless to say, one may use way of life while remaining agnostic to the metaphysical commitments it implies. In this view, the concept might imply a non-essentialist, heuristic form of typological thinking that represents just one of many possible typological "strategies" that biologists could use (Amundson 1998; Love 2009). Indeed, this would be precisely in line with the AT's way of thinking. In this sense, a way of life could be a useful epistemological tool or idealization rather than an ontological thesis about biological essences, and under this guise may still have some use for proponents of the AT benefits in their theorizing about teleology.[60] These possibilities are well worth exploring, but I will not do so here.

5.4.2 Evo-Devo and Teleological Essentialism

The starting point for neo-Aristotelian accounts is their appeal to the *natures*, or capacities intrinsic to organisms (Walsh 2006; Austin 2018). Such powers or

[60] Indeed, it might be tempting to interpret the way in which activities make up a coherent and *unified* way of life as a loose kind of 'closure of activities' that supervenes on the organizational 'closure of constraints'.

capacities are taken to be the metaphysical bedrock from which scientific explanations are constructed. As Walsh argues: "[i]n Aristotle's essentialist biology the nature of an organism is manifested as a goal-directed disposition to produce and maintain a living thing capable of fulfilling its vital functions in ways characteristic of its kind." (Walsh 2006, 427; see also: Haldane, Russell, and Mackenzie 1923, 77–78). As discussed above, this entails *neither* a commitment to the fixity of species, nor to the idea that types are rigidly unvarying. While it is true that Aristotle did not see biological kinds as evolving, there is nothing about his biological project that necessarily commits one to this view. Moreover, natural variation was not something which Aristotle attempted to do away with, but rather to *explain* with his account of essences.

What kind of properties *would* be causally informative? Austin (2018) and Walsh (2006; 2021) argue that it would be precisely those properties that evolutionary developmental biologists have been uncovering over the past few decades. Organismal kinds ought to be understood in terms of the particular morphological profile which is caused by a collection of dispositional properties inherent to its developmental modules (ibidem; Austin and Nuño De La Rosa 2021). While Austin argues that the best candidate realizers for these dispositional properties are gene-regulatory networks, such realizers are not identical to the dispositions. In other words, Austin may turn out to be wrong about the way these powers are materially realized or may be that other life-forms have those properties materially realized differently; but this would not be a particular problem for his view (2018, 107). What really matters for the Neo-Aristotelian essentialist position are that there are intrinsic, teleological dispositional properties which are causally responsible for its morphological profile and provide the developmental modules with the requisite robustness and plasticity (Walsh 2006).

One place in which Austin and Walsh part ways is in their conception of purposiveness. For Austin, all dispositional properties are purposive (Austin 2018, 16). In contrast, Walsh takes purposiveness to be distinctively biological. This disagreement mirrors an exegetical debate within Aristotelian scholarship regarding the way final cause is to be interpreted. While some see the normativity of purposiveness to be part and parcel of the realization of a potential or disposition, others take this to collapse the distinction between efficient and final causes (Gelber 2021, 100-101). While I side with Walsh and Gelber in this respect, the way one interprets the notion of "disposition" is an issue that requires further elaboration that is beyond the scope of this chapter (see Brigandt et al. 2023 for a recent discussion in the context of evolutionary biology).

Beyond these differences, neo-Aristotelians agree that taken together, developmental modules confer onto the developing organism the simultaneous stability and mutability that make it possible for it to develop into and pursue a way of life appropriate for members of its kind, over a wide range of conditions. In this sense, essences are equally about observable variation as they are about the evident *constraints* on that variation, and the corresponding constraints on the potential ways of life (ends) that this enables (Nuño de la Rosa and Lennox 2023). According to this research, essentialism has been given new life by recent advances in biological research: those properties that are best suited to be candidates for nature are, ironically, indispensable for explaining development and adaptive evolution.

That variation is part and parcel of a kind's essence, and that many dispositional properties are shared across even distantly related species (Austin 2018), speaks to the fact that essentialism need not subscribe to sharp species boundaries. As Devitt argues, "the evolution of S2 from S1 will involve a gradual process of moving from organisms that determinately have G1 to organisms that determinately have G2 via a whole lot of organisms that do not determinately have either. There is no fact of the matter about where precisely the line should be drawn between what constitutes G1 and what constitutes G2, hence no fact of the matter about where precisely to draw the line between being a member of S1 and being a member of S2. Essences are a bit indeterminate." (2023, 28). This is an important point that defuses one of the central worries with essentialism. Moreover, Austin and Walsh's developmental-centered notion of essence allows us to put a finer point on this, which will prove relevant for the discussion of norms and functions.

Given that essences are defined by sets of intrinsic dispositions that carve out a space of possible developmental trajectories or morphospace, to say that essences are indeterminate is really to say that it is this space itself which is indeterminate. This is not merely an issue of complexity. Rather, it is because development is an active, goal-directed process in which the organism responds to affordances in its environment and in turn changes those very affordances, in a process which has been called 'dialectical' or 'experiential' niche construction (Levins and Lewontin 1985; Chiu 2019; Sultan, Moczek, and Walsh 2021; see also: Oyama 2000; Moczek et al. 2015; Laland et al. 2015; Aaby and Desmond 2021).[61] In this sense, Walsh suggests that developing into an adult organism in an adaptive manner is not like solving a sudoku problem where the answers are fixed ahead of time; indeed, it is more akin to a game of chess in which every move changes the space of possible moves which can be made next (Walsh 2014b). Though arguably even chess does not fully capture what is distinctive about this process. A more radical analogy for this process asks us to consider the set of possibly musically meaningful symphonies (Montevil 2019). A symphony is a sonic artform which must presumably be within an audible range for human listeners. But the mere fact of being audible is not enough to specify whether it is musically meaningful. The range of possible musically meaningful symphonies is highly dependent on the symphonies which came before; and each new symphony—along with a whole range of also evolving cultural and social factors—changes the space of what is possibly heard as musical. The space of musically meaningful symphonies is a constantly evolving space which, while not infinite or unbounded, is indefinite in the sense of being algorithmically *unprestatable* (Kauffman 2000; Longo, Montévil, and Kauffman 2012). An alternative musical

[61] For a recent overview of the debate regarding the way development influences evolution and a proposal for conceptualizing this process, see Ramsey and Villegas (2024). Regarding the more general issue of reciprocal causation between organism and environment, see Baedke et al. (2021); see also Hazelwood (2023) for a critique.

analogy is to be found in jazz. The genre is immediately identifiable by a shared vocabulary of riffs, motifs, and lines, and a set of standard tunes. The performance of any standard will use this vocabulary and follow the general structure outlined by the chord progression of the tune (though even that is liable to change through reharmonization). And yet, the improvisational quality of a performance will lead it to uniquely chart a course through a space of possibilities shaped by the dynamic response of players to each other. Each is continuously responding to what came before and anticipating where it will go next as the performance continues. It is precisely this dynamic of continuously evolving possibility spaces that is thought to characterize both biological development and the evolutionary process (Montevil 2019). The remarkable capacity of agents to navigate such a space—from relatively simple developing organisms confronting environmental stressors to humans confronted by new situations—and particularly to be able to pick out those features of a novel situation that are *relevant* to achieving their goals, is only beginning to be understood (Jaeger et al. 2023).

The above begins to illustrate, I hope, the tenability of a teleological essentialism today, given the necessary specifications that one must make given recent work in developmental, philosophical, and theoretical biology. Essences in this view are comprised of intrinsic dispositional properties that carve out an unprestatably large—though not unconstrained—space of possibility through which an organism may develop. And because this developmental trajectory always occurs within a particular environmental context, morphospaces are intimately tied to the notion of a way of life. The pursuit and realization of a way of life is an achievement of each individual developing organism. And while this process mobilizes and is constrained by their evolved repertoires, which play a decisive role for the lives that would likely unfold, but these do not entirely fix or determine what's possible. For this reason, the two-legged goat discussed earlier should be interpreted as a surprising discovery of the surprisingly wide range of ways of life that are possible for goats; even while these variations may be altered by the process of natural selection.

How an organism develops shapes its landscape of affordances, i.e., the relation between its behavioral repertoire and the way it experiences its environment, which in turn opens up a possible path for pursuing a way of life (Walsh 2018). Even if one makes the idealizing assumption that its goals are fixed from the outset, the way in which the organism develops will inevitably change how it pursues those goals and by extension, its way of life. This is because the morphological features typical of an organism's kind are necessary for its way of life. As Lennox argues, "if a hawk had all its other features, but lacked the crook talons to grasp its prey, it would not be able to live the life of a hawk. That is what it is to be a hawk." (Lennox 2010a, 349–50).[62] Ways of life, in this sense, are the expression of the essential, intrinsic dispositional properties that form the metaphysical bedrock of intrinsic purposes. An end-explanation adverts to such essences, and a goal-explanation metaphysically depends on their existence. If this is right, then the conclusion which purposive behaviorist, cyberneticians, and defenders of teleonomy wished to avoid is true: there is a metaphysical difference between organisms and artifacts. If the wellspring of intrinsic purposiveness is to be found in the essences of organisms, then we now have a firmer grasp on what it means for artifacts to be *extrinsically purposive*: their purposes and functions metaphysically depend upon such organismal essences.

5.6 Conclusion

In this chapter, I have begun to show how the EA may carry forward the lineage of purposive behaviorism while avoiding one of its major pitfalls, namely, the characterization of purposive behavior abstracted away from the concrete way of life in which they exist. This abstraction flowed from a desire to both simplify an

[62] Such a conception of a way of life makes better sense, I believe, of Jonas' invocation of "the ends by which one lives" ([1966] 2001, 120), viz, the ends according to which one is the kind of entity that one is.

otherwise highly complex phenomenon, and to meet certain strict (albeit largely implicit) naturalist criteria by deploying artifact models and metaphors that remained largely agnostic to the metaphysical questions raised by the phenomenon of intrinsic purposiveness. This is not to say that there is nothing to be gained from showing how idealized artifact models of purposive behavior may concretely advance scientific research on numerous dimensions—far from it. Rather, I merely wish to emphasize that this leaves metaphysical questions unanswered. Returning to the phenomenon of biological agency with these considerations in mind, the EA emphasizes that they exist within a constantly shifting landscape of affordances to which agents may respond to with a variety of different means given their behavioral repertoire. The best way of theorizing this, I've suggested, is in terms of the pursuit of a way of life, which manifests in terms of the overall orchestration of goal-directed activities. This suggests a distinction between two kinds of teleological explanations: goal-explanations, which explain a behavior based on a goal being pursued, and end-explanations, which explain by adverting to the intrinsic essence that, expressed through a way of life, constitute the agent as the kind of agent that it is. While the realization of an end is the consequence of the pursuit of goals over time, a system can only have goals and be an agent if it has an end. Finally, the distinction between goals and ends goes some way in further clarifying the contrast and potential compatibility between the AT and the EA. The former conceptualizes teleology as causal-consequence explanations which explain why a system exists in a causal sense. The relevant contrast case seems to be end-explanations, which explain why a system *exists as* the kind of entity it is in a constitutive sense. In this light, it appears that the AT and EA are asking different kinds of questions.

CHAPTER SIX

Functions and Teleological Explanations

In the first part of this dissertation, I identified a common thread which runs through the purposive behaviorist, cybernetic, and teleonomic conceptions of teleology. Despite their differences, they are all artifact models of teleology which attempt to derive a conception of intrinsic purposiveness from extrinsic purposiveness. These problems went unrecognized by the proponents of the artifact model of teleology, for whom it seemed a genuine philosophical breakthrough that resolved the problem of backward causation and made teleology compatible with mechanism. But undoubtedly the most significant legacy of the artifact model of teleology was to be had on the debate over biological functions.

During the 1970s, a near-consensus emerged that function-talk in biology was here to stay, and therefore merited serious philosophical attention. But it also motivated the idea that the debate over purpose and teleology should both begin and end there (Ruse 1971). By the early 2000s Tim Lewens reported on the general sense that teleology was "boring", because the debate was either resolved, fruitless, or amounted to nothing more than being clear about the meaning of function (Lewens 2004, 4). This is ironic, given that function debate which emerged in the 1970s was deeply tied to earlier debates about purposiveness and teleology which I aimed to illuminate in the first two chapters of this dissertation (see: Garson 2016, chap. 2).

Wright's etiological account of function emerged out of an engagement with the issues raised by teleological explanations in general, and particularly what he called the ‘empiricist’ accounts of purposive behavior defended by Ernest Nagel, Richard Braithwaite, Charles Taylor, and the cyberneticians (Wright 1972). For this reason, one of the central issues which Wright discussed was the relationship between function and teleology. He warned against conflating these concepts and treating function as “all there is to teleology” since it is “not even the central, or paradigm, teleological concept” (ibid, 134). At the same time, he claimed that one “should expect functional ascriptions to be explanatory in something like the same way as goal ascriptions” (ibid, 154), in the sense that they are not only *etiological* but also *teleological*. As discussed in the previous chapter, Wright claimed that etiology means “causal” in a broad and highly ambiguous sense intended to accommodate both teleological and ‘merely causal’ explanations. Both, according to this view, are ways of answering questions about how something brought about or produced a phenomenon. What makes teleology distinctive as a species of etiological explanation is the way in which a phenomenon is explained in virtue of the consequences it brings about.

Ambiguities with the notion of causality—which have permeated the various accounts discussed in Part 1—led the etiological account to be used as part of a project of naturalizing function that would simultaneously de-teleologize biology (Walsh 2014). This was a natural outcome of the artifact model of teleology. To say that organisms and artifacts are purposive in the same sense was a way of making purposiveness reducible to a causal-mechanical explanation; usually—though not necessarily—through the idea that evolution is design-like. Etiological or selected-effect accounts of function were then thought to form a promising basis of a naturalized account of intentionality, and all without committing to purposes a feature of the structure of the world. The reliance of these programs on artifact models of teleology doomed them to fail from the outset; but it also created opportunities for rethinking how the problem should be approached.

In this chapter, I aim to provide an account of functions based on an account of organisms as intrinsically purposive agents. More precisely, I aim to resurrect a largely unpopular account of functions which takes them to be the goal-contributing roles performed by processes and entities in the context of purposive activities. This view has been called the goal-contribution approach, or 'GCA' for short, and is usually considered a species of so-called 'dispositional' accounts of function (e.g., Mossio et al. 2009). In doing so I will put particular emphasis on the role that teleological explanations—in the non-causal, goal-consequence sense defended by the EA—play in function ascription. Past attempts at a GCA of functions, such as those defended by Boorse (1976) and Adams (1979) rested on behaviorist and cybernetic accounts of purpose which, for reasons already discussed, fail to adequately characterize the phenomenon of intrinsic purpose. Furthermore, accounts of functions based on them were plagued by a fatal problem, namely the inability to distinguish functions from accidental contributions to goals. I will argue that these problems stemmed from an inadequate conception of teleological explanation which can be overcome by the EA's account. My proposed version of the GCA, which derives a conception of artifact function *from* an account of intrinsic purpose, will further highlight why the artifact model of teleology cannot work: function ascription depends on a notion of teleological explanations with a rich content, including the counterfactual dependence of means on goals. In turn, such explanations have intrinsically purposive agents as their necessary condition, for they are the performers of purposive activities in the context of which functions may be ascribed.

The structure of this chapter is as follows. In the next section, I introduce the GCA and contrast it to Wright's etiological approach, with an emphasis on the different ways they make sense of an accidental function. In section 6.3, I argue that the EA's account of teleological explanation provides the tools necessary for distinguishing cases of accidental and non-accidental goal-contribution. Section 6.4 further expands in this view by showing how the context-dependence on function ascription also vindicates our intuitions that *designer-intentions* can play a role in function ascription. Finally, section 6.5 shows how my account of artifact function could be extended to biological functions.

6.1 The Goal-Contribution Account of Function

According to the goal-contribution account of functions, or GCA, an entity or process' function is to perform in such a way as to contribute to the attainment of a goal (Boorse 1976; Nagel 1977; Adams 1979). While the previously cited works are some canonical examples of this approach, there are many other similar accounts which may be categorized under the same umbrella if the meaning of 'goals' is broadly construed. Boorse, for instance, interpreted Ruse (1971), Canfield (1964), and Sorabji's (1964) accounts as all variants on the GCA, where the goals in question are construed as 'reproductive fitness', 'usefulness', or 'the good' respectively (Boorse 1976, 77).[63] Boorse deemed these competing approaches to be operating under an overly restrictive notion of goal, suggesting that we employ a notion of goal-directedness defended by Sommerhoff instead (1950; see also: Nagel 1979). In this view, "To say that an action or process A is directed to the goal G is to say not only that A is what is required for G, but also that within some range of environmental variation A would have been modified in whatever way was required for G" (Boorse 1976, 78). As he saw it, such an account provides the requisite amount of generality to be unificatory, insofar as it applies equally to both artifacts and organisms, though in the latter case he deemed these goals to be restricted to survival and reproduction (Boorse 1976, 84).

The contrast between Wright and Boorse's approaches provides a useful framing for the foregoing discussion. Wright emphasized three desiderata for an account of functions: 1) First, functions should have an *explanatory* role. As he put it, "functional ascriptions are—intrinsically, if you will—*explanatory*. Merely saying of something, X, that it has a certain function, is to offer an important kind

[63] Typically these three authors are not thought to be proposing GCAs, but rather a distinct kind of 'fitness contribution view'. That being said, both of these accounts, along with a third 'systemic' view, have been characterized as part of a broader family of *dispositional* approaches, where the function of x is its disposition to contribute to a larger system (Mossio et al. 2009). The divisions between these three subsets are not clear cut (Garson 2019). Furthermore, though dispositional views are usually contrasted with etiological views, this contrast has been contested (ibid).

of *explanation* of X" (1973, 154). 2) Second, a function analysis should be unifying, in that it applies equally across artifact and biological cases. 3) Third, any account of function should be able to make sense of the distinction between functions and accidents. While it is true that noses hold up glasses, hearts make thumping sounds, and screwdrivers open cans of paint, it is not their function to do so (ibid, 141). The fact that hearts, noses, and screwdrivers do this is merely *incidental*, and an account of function should not ascribe a function to those incidental effects. Though we may say that heart thumping "functions as" a diagnostic tool, or screwdrivers "function as" can openers, these are functions only in a weak sense for Wright (ibid, 148). This distinction between functions and accidents was so important to Wright that he took it to be "*perhaps the primary aim*" of his analysis (ibid, 142; emphasis added).

All three points featured prominently in Wright's critiques of Morton Beckner (1959; 1969) and John Canfield (1964). The latter two are particularly important: what function ascriptions ought to explain (desideratum 1), according to Wright, is why the functional entity in question exists, rather than merely 'what it does' or its contribution to what is 'useful' or 'good' (c.f. Cummins 1975). These "weaker" interpretations inevitably failed because they could not make sense of desideratum 3, i.e., the function-accident distinction (Wright 1973, 155). Screwdrivers and hearts may do innumerably many things and may be quite *good* for many of those; but their proper function only corresponds to a small subset of those. Treating existence as the explanandum for function statements served three roles: i) as a way to demarcate function and accident, while ii) rendering function statements robustly explanatory, and iii) making sense of the intuitively "obvious role of design" for artifact functions while accommodating biological cases (ibid, 153; 146).[64] As Wright saw it, both Beckner and Canfield's accounts failed to rule out accident cases and account for the role of design (ibid, 153).

[64] Though it may also be true that this is a holdover from Hempel, as Boorse suggests (1976, 75).

Boorse challenged Wright's analysis and is best remembered for coming up with a series of counter examples (Garson 2016, 40). On the biological front, he challenged Wright for assuming that all claims about functional statements are made with reference to evolution. For it seems that physiologists do not necessarily do so, nor should we imagine that function statements made prior to the rise of evolutionary theory were necessarily theological in nature or meaningless (1976, 74).

More important for my purposes are Boorse's critical comments on Wright's distinction between "proper function" and "function as" (or "function as use"). While Wright rejected that an artifact or trait's functions is whatever it might be good for, he admitted that there are cases where entities may usefully *"function as"* something else. Such cases are typically taken from artifacts that are being used for reasons other than they were intended, such as the flathead screwdriver used to open a can of paint. In this case, the screwdriver *functions as* a can-opener, but that is not its "proper function", i.e., what it was designed to do.[65] Similarly, it is thought to be absurd to say that a nose's function is to hold up glasses, even if they "*function as*" a means of supporting them in a particular context. In such cases, we are effectively "pretending" that this was the reason why it was designed, i.e., came to exist (Boorse 1976, 77). Boorse rightly responded to this with skepticism: "It remains obscure why in the world we might nevertheless wish to 'pretend' that their etiology was of the right sort to justify functional language. If the pretense is made only where a trait has useful effects, that is some argument for a connection between functions and usefulness." (ibid).

Boorse maintained that the distinction between 'weak' (Wright's "function as") and 'strong' (proper function) cases is really a linguistic one, and thereby somewhat "illusory" (ibid, 81). A Bible which stops a bullet once, for Boorse, is incidentally "functioning as" a protective shield; but if it does so thirteen times then it becomes *the* function of the Bible. The appropriateness of the terms 'a' and

[65] Though the etiological view may claim that an accidental "function as" becomes a proper function when the functional entity in question persists because of what it does (Wright 1973, 165)

'the' are determined by context: "'The function of X' will be simply that one among all the functions performed by X which satisfies whatever relevance conditions are imposed by the context of utterance" (ibid, 82). In doing so, Boorse effectively denied that there is a substantive distinction between the proper "function of" and "function as", and asserted that cases of incidental function (such as the Bible-stopping bullet) are perfectly legitimate cases of functions being performed (ibid, 80-82). So what is absurd about the claim that noses are for holding glasses is *not* that they perform that function in that context, but rather the inference which one might make from this claim to a justification for existence (as satirically illustrated by Voltaire's Pangloss); and this is not an inference that, according to Boorse, we need to make if we want to attribute that function to a particular nose at a particular time.

While there is much that seems right about Boorse's analysis, his embrace of the fortuitous function cases remains highly problematic. In the next section, I will give my own version of the GCA, which incorporates what I take to be the best aspects of those views while also showing how they can overcome Wright's critiques of them. From Adams I will take up the idea that functions are closely tied to the notion of means (1979, 494).[66] I will also follow Boorse in arguing that function-ascriptions are highly context sensitive. At the same time, I will take seriously Wright's claim that function ascriptions are first and foremost explanatory, and that such explanations are *teleological* rather than *causal*—though I shall use the EA's rendition of that contrast. My aim is to show that a bona fide teleological mode of explanation that takes the phenomenon of purposive behavior as its starting point can be the foundation for an account of function. This account ultimately underscores the shortcomings of the artifact model of teleology.

---

[66] This proximity between means and function is also implicitly recognized by Wright when he states that "the 'in order to' in functional ascriptions is a teleological 'in order to.' Its role in functional ascriptions (the heart beats in order to circulate blood) is quite parallel to the role of 'in order to' in goal ascriptions (the rabbit is running in order to escape from the dog)." (1973, 154). For this reason, my account of function is close to an account of action (Fulda 2016, chap. 6).

6.2 Functions, Accidents, and Teleological Explanations

Given the centrality of goal-directedness for GCAs of function, the way in which function attributions are made greatly depend on the account of purpose and teleological explanation in question. For Boorse, this was a version of Sommerhoff's account of goal-directedness as persistent and plastic behavior. The resulting account of functions is spelled out by Boorse in this way:

> "X is performing the function Z in the G-ing of S at *t*, *means*:
> At *t*, X is Z-ing and the Z-ing of X is making a causal contribution to the goal G of the goal-directed system S."
>
> (Boorse 1976, 80)

As this formulation makes explicit, "all function statements, weak and strong, [are] implicitly relative to system, goal, and time" (ibid). It is also clear from this that this version of the GCA cannot avoid attributing functions to bullet-stopping Bibles, since *any* causal contribution to a goal—whether intended or by chance—will be functional. As discussed in the previous section, Wright explicitly critiqued Canfield and Beckner's accounts for this (Wright 1973, 152).

But it would be incorrect to think that this is an inevitable consequence for any version of a GCA. Indeed, one might think that the reason for this outcome is simply that the notion of a 'causal contribution' is underspecified. An alternative account developed by Adams (1979), for instance, attempted to avoid such cases of 'accidental function' by taking up an explicitly cybernetic account goals such that functions are ascribed in virtue of the way in which they enter "appropriate means-ends relation in a goal-directed system", i.e., insofar as the functioning entity is "part of the causal chain involved in the feedback loop of the goal-directed system" (1979, 511). In this case, one avoids ascribing the function of stopping bullets to the Bible, since it is not part of the causal chain of the feedback loop that governs the soldier's movement.[67]

Putting aside the fact that Adams' cybernetic account of purpose is deficient in its own regard (for reasons discussed in the first two chapters of this dissertation), the real problem with both Boorse and Adams' accounts, in my view, *is that the explanatory role of function in both of their accounts is causal.* Even though Adams claims to center the relation between means and goals in his account, it is never clear how that relation is any different to that between causes and effects. In fact, a means-goal relation for Adams is merely a causal relation which occurs within the context of a servomechanical system. If the goal of a servomechanism is to maintain a particular value, then anything that causally contributes to that is a means, according to Adams. As he sees it, X has a function Y if that activity causally leads to the system it is part of outputting a particular value (ibid, 494).

In these respects, both Boorse and Adams failed to appreciate the importance of Wright's attempt to distinguish between causal and teleological explanation, and its relevance of that contrast for making sense of function. Not so for the EA. According to this view, the distinction between causal and teleological explanations is to be found in their respective modal profiles. While the details of this distinction were spelled out in Chapter 4, it is worth reviewing them here. In causal explanations, there is an invariant relationship between causes and effects, whereby the effects counterfactually depend on the causes, and the activities which produce the effect elucidate that relation. In contrast, teleological explanations involve a counterfactual dependence of means on goals, and we show how the goal comes about through the ways means conduce toward it. So,

---

[67] Adams' way of distinguishing accident from function mirrors the AT in this respect. The previous chapter discussed how the AT renders teleological explanations in causal terms, and ends up with a Wrightian formulation of purpose: an organism's goal corresponds with its causes of existence. Functions, in this view, are attributed to parts of systems that contribute to this goal, and in doing so, indirectly maintain themselves. On this basis the AT can distinguish parts of closed systems that causally contribute to the system (and their own) existence, and those factors that only do so by chance (and therefore do not have a function).

a defining feature of teleological explanations is the robust counterfactual relationship between means and goals. This relationship is both ontologically and explanatorily asymmetric: while there cannot be a means if there is no goal, the existence of a goal necessarily implies hypothetical means for achieving it (even if either the goal or the means are impossible to achieve). Similarly, means are *explained* by goals, but not vice-versa. In fact, this two-fold asymmetric dependency is a good way, I would suggest, to characterize the difference between so-called "intrinsic" and "extrinsic" teleology in general. So if we are to take seriously Adams' claim that "the essential feature of a functional relationship is that of a means-ends relation", then it seems that functions ought to be *explained* by the goals they conduce to. As I shall now show, this provides a way of distinguishing functions from accidents; and because functions *qua* means are part of the explanatory structure of teleological explanations, I will take my account to vindicate Wright's claim that function ascription is explanatory in just the way that teleological explanations are.

If it is a soldier's goal to avoid being hit by an enemy bullet while running across a field, then a bullet-stopping Bible may contribute to the realization of that goal. So may many other things, including the soldier's helmet, his courses of action, and so on. Consider then two soldiers, one of whom is saved by a Bible and the other by a helmet. In either case, we can give a causal story about why, given the initial conditions, bullets hit and were blocked by the helmet and Bible respectively. But there are also relevant features of this story that are to be explained teleologically by the respective soldier's goals. One of these features is why the soldiers were carrying helmets and Bibles respectively. In the former case, the soldier's helmet wearing is explained by their goal of staying alive; so its capacity to contribute to that goal is explained by the goal. It therefore has *the function* of contributing to the goal in that way. In contrast, the fact that the other soldier was saved by the Bible would not be explained by their goal of staying alive.[68]

These cases mirror a famous passage from Aristotle's Physics II.5 (Barnes 1984). To give a variant on the example presented there, consider a creditor who happens on his debtor by chance after encountering him at the market. In this chance encounter, the creditor's purpose of being reimbursed is fulfilled; but his goal of being repaid does not *explain* why he went to the market. Similarly, the Bible's fortuitous bullet-stopping is not explained by the soldier's goal of not being shot; but his wearing a helmet is. In both Aristotle's example and the bullet-stopping Bible, there is a key difference between the ways in which one explains events that occur by chance—irrespective of whether these chance events are beneficial—and events that occur because of a goal. These respective explanations have distinct modal profiles (Walsh 2017, 245–244). Whereas chance events are extremely sensitive to initial conditions, goal-directed or purposive activities are not. Had the creditor left moments earlier or later, he would not have chanced upon his debtor. But if his goal was to find his debtor, then he would have done what was necessary to find him within his power, *despite* contingent factors. In fact, it is precisely because goals are robust across a range of initial conditions that we can cite them in explaining why this means, rather than some other, was used to achieve the goal.

With that in mind, we can now return to the Bible and the helmet. On one level, the fact that both the helmet-wearing and Bible-carrying soldiers were saved is due to chance, which are events that are extremely sensitive to initial conditions. The Bible or helmet would have failed to save the soldier had it been a few centimeters to the right or left, or if the soldier had started running a second before or after they did. But over and above these chance conditions, the fact that soldier A was carrying a Bible and soldier B was wearing a helmet *are explained by different*

---

[68] There is, of course, room for Gettier problem-like cases where the soldier intentionally carried a Bible with them thinking that God would reward them for their faith. They might be mistaken in that belief, yet the Bible *would* serve a function in this particular case. But the goal which explains this series of events is the soldier's intention to have their faith affirmed.

goals. And it is only because soldier B's helmet wearing is explained by the goal of staying alive that we can attribute to it the function of doing so. The same would be true for a large piece of metal that the soldier happened to find on the field and chose to use *as* a shield: it has the function of stopping bullets because it is being used for that purpose. But the same is not true for the Bible. The Bible may well have a function in contributing to one of soldier B's goals, but those are orthogonal to its bullet-stopping capacities.

The foregoing nicely illustrates the difference between an "accidental function" and an accident *simpliciter*, which is a distinction Wright fails to make (1973, 149). Accidental functions, or what Wright calls "function as", are weaker according to Wright because a functional explanation which adverts to the causes of existence is more robustly explanatory. But the EA's account of teleology poses no constraints on *what* goals can or should explain. Nor do we have any reason to think that the nature of the explanandum impacts the strength of the explanation.

Another important aspect of the EA's conception of teleological explanation is its normative component. So it follows that explaining functions teleologically imbues them with the kind of normativity one would expect of function. If a soldier wears a helmet to stay alive, then the helmet's conduciveness to this goal is to be evaluated normatively: a good helmet will stop bullets; a bad helmet will fail to do so. Either way, the normativity of function is fully derivative from the normativity of teleological explanation. And since I follow Boorse in thinking that there is no real distinction between proper functions and "function as", I also take it that the distinction has no normative import either: one can evaluate how good a flathead screwdriver is at driving screws independently to how good it is at opening cans of paint; and previous experience tells us that it is good at doing both.

To summarize, Wright's etiological account of function—which ascribes functions to the causes of an entity's existence—was criticized by Boorse on the grounds that it cannot account for entities "functioning as" something that they were not designed for (or for implements with no history of design). Boorse's solution to this was an overstep in the right direction, insofar as it dissolved the distinction between "function as" and "proper functions", but also made any accidental contribution also functional. In this regard, it failed on largely the same grounds as Beckner and Canfield's accounts. I've shown that one can retain the positive aspects of Boorse's view without its shortcomings. An entity or process can have a function—i.e., be explained by a goal—*without* inevitably attributing functions to all accidental or chanceful contributions to goals. In other words, I've shown that a GCA of function can, on explanatory grounds, satisfactorily distinguish between "function as" and accidental functions. Importantly, this outcome is a direct result of the account of teleological explanation defended thus far, which I have argued depends on the intrinsic purposiveness of biological agents as its necessary condition.

In the next section, I will further flesh out this EA of functions by putting particular emphasis on the context-sensitivity of function attribution. Doing so will clarify how the EA of function can accommodate the intuition that designer intentions do go some way in fixing functions, albeit in particular contexts.

6.3 Function Ascription and the Importance of Context

One of the key features of my account of function is the context-sensitivity of function attribution (a point also emphasized by Boorse 1976, 81). An entity or process has no function outside the context of a purposive activity, and so function statements must always at least implicitly be statements about said activity. Specifically, function statements are about that *for the sake of which* the activity is performed, and by extension the agent performing it. When we say that hammers are for hammering, we are saying that a hammer is or would be conducive to the realization of that activity when it is required for the attainment of a goal. Building a table, for instance, may require fastening four pieces which will become the legs of the table to the tabletop. In the context where this fastening is appropriately realized by hammering nails through wood, a hammer is appropriate and conducive to that end; its *use* is thereby *explained* by its appropriateness to the end.

As we've seen, this explanation follows the counterfactual structure of teleology by telling us that had the hammer not been available, the agent may have chosen from many other objects that share the relevant structural features for hammering in this context. And had the hammering *not* been required because the craftsperson decided to join the legs using joinery (which does not require nails) another tool or set of tools would have been used.

Conduciveness comes in degrees. While hammers are engineered to be highly conducive to hammering, other hard objects could also be used for the job. Perhaps the crudest tool would be a rock, which has no history of design. But I maintain that insofar as a rock succeeds in being conducive to hammering, its function *in that particular context* is rightly to hammer. This would be precisely the implication of any explanation given to an observer asking "Why is Peter hitting nails with the rock? What is he doing that for?". An answer would likely say something about the absence of a hammer; but it would also imply a whole range of objects that Peter could have used but did not, such as a glass bottle, because it would *not* have been conducive to hammering. Nonetheless, insofar as the rock succeeded in performing the task at hand, the relevant functional contrast between it and the hammer is how *good* they are at performing that task *in that particular context*, i.e., given the job to be done and the available options. So the rock really is 'for hammering' when its use is explained by that goal; it "functions as" a hammer, to use Wright's language, and Boorse is right that there is no need to "pretend" that it was designed for that purpose to make that claim (1976, 147). As previously discussed, I take Boorse to be correct in thinking that the distinction between function and "*function as*" is dubious. And yet Wright is *also* correct in thinking that designer intentions matter for ascribing function, and that an artifact can perform functions which are 'accidental' or 'fortuitous' with respect to *that* etiological function. What are we to make of this? As I see it, one can rescue what is correct about this intuition while also showing that the distinction between function and "function as" is not warranted.

To see this, consider that artifact creation is itself a purposive activity that is explained by a goal. A teleological explanation of the screwdriver's *existence* in terms of a maker's intentions or goals to create it is just as explanatory as explaining a screwdriver's *use* in terms of the user's goal. In either case, I maintain, the screwdriver has a different function *depending on the particular purposive activity we choose to focus on*, in other words, depending on which agent's *perspective* we choose. This is because, in this view, identifying a function requires saying *for whom* it is a function, and *to what purpose*. There can be no function ascription outside of one of those contexts, and there is to be no reason to privilege one context over another. So when the creation of an artifact is a means to a goal, it can rightly be said to be serving the function of bringing about a particular goal.[69] To illustrate: a designer's goal may be to make an artifact, but this goal may also be a means toward the goal of using it for another purpose down the line. In the former case, the existence of the artifact is explained by its creator's goal of making it; in the latter case, the use of the artifact is explained by a goal it is conducive to realizing. In this sense, I take it that *creating*, like *maintaining*, *repairing*, *reproducing*, *altering*, or *using* are all teleological activities within which an object may gain a function insofar as it contributes to the goal, *qua* means. As such, design can and often does appropriately ascribe function in one context; but we need to let go of the widely held view that design has a *privileged* role in fixing function that goes beyond that specific context.

This has a crucial implication that distinguishes my account of function from most other accounts: to ask after an entity's "function" in the abstract, as if it is an inherent rather than a purely relational property, is mistaken. To say what an entity's function is, one must at least *implicitly* indicate a particular context. In everyday speech, this implied context is usually either the current context—e.g., the example of Peter above—or the context of the artifact's designer. Asking after designer intentions *is* in fact often what we are implicitly doing when we ask after the function of a thing. But Boorse is right about why it is that we usually do this.

---

[69] Needless to say, nothing precludes the possibility that the creation of an artifact is an goal in itself, and this is also true of some actions, such as dancing. In cases where an action is done or an artifact is made for its own sake, there is no function.

If we enquire about the function of some part of a car, it "is because we want to know what good X is for making our car run, *and the designer is usually the best authority on how the artifact works*. Etiology seems relevant only insofar as it is expected to show how the part can advance the goals we pursue with the car." (Boorse 1976, 82). In other words, designer intentions usually tell us what goals an entity is most conducive to leading towards. And the more sophisticated the artifact, the more important this information is. Indeed, without any information about what an extreme ultraviolet lithography machine is good for, it might be very hard to surmise that it is remarkably well-suited to printing complex designs on microchips. But designer intentions do not restrict the scope of what an artifact *could* be good for in any possible context. Indeed, designers routinely fail to see that their creations may be best used in contexts far removed from those they had imagined. And so, it can be simultaneously true that i) a screwdriver's existence is explained by a particular goal-directed activity, and within that teleological context, it serves a function; ii) a screwdriver is an excellent way to screw a screw; iii) its use in another teleological context is explained by another goal, which gives it another function *in that context*.

The fact that function attributions are indexed to particular activities gives this account a tremendous amount of flexibility. Indeed, functions in this view are highly dynamic and context sensitive rather than a fixed property. This does justice to the way agents routinely achieve their goals by jury-rigging a solution. In this regard, this view strongly contrasts with Wright's claim that "the notion of something having more than one function is derivative" (Wright 1973, 166). I take the opposite view. Jury-rigging is the norm, and functional rigidity (conduciveness to a highly particular goal, as is the case with sophisticated machines) is a special case (Canguilhem [1952] 1992, 56-58).

That being said, such context sensitivity also brings to light a potential problem, given that it implies that an entity ceases to have a function when its involvement in the activity ends. But it may not be clear when activities cease. To illustrate, at some point a hammer surely ceases being part of the activity of building the chair. But does this involvement end when the hammer is momentarily put down? Or when it is put away for the night? Or only once the chair is complete? As I see it, the boundaries of activities are indeed fuzzy, but there is no real need to be worried about this. Again, we must return to the *explanatory* use of function statements. A workshop is conducive to constructing furniture because it contains the hypothetically necessary tools for the accomplishment of that goal. Such explanations, while general, do not fail to explain because of that generality; and by the same token, they do not fail to ascribe functions in an explanatory way. Moreover, if one is asking why one finds hammers in workshops or toolboxes, then an appeal to past, present, or even *hypothetical* purposive activities will usually be appropriate. And nothing precludes talk of a function that an entity once had in the context of a particular purposive activity that is no longer carried out, as is the case with archeological artifacts.

The highly context-sensitive nature of function of this account makes it out to be both specific and open-ended. Given the context of a particular purposive activity, the functional contribution(s) of an entity or process to the attainment of a goal will in most cases be rather clear. But the set of possible functions which an entity or process can have is highly open-ended, and in fact may well be indeterminate or "*unprestatable*" in the sense discussed in the previous chapter. The first person to use a screwdriver to open a can of paint made a genuine discovery of one of its excellent uses. So not only can one not ascribe a function outside of a purposive context, one cannot pre-state all the possible contexts within which a process or entity would have a relevant function, and thereby all the possible functions that a process or entity might have across all those contexts. This goes some way to illuminate the age-old issue of the relation of structure to function (Novick 2023). While one can theoretically specify the exact structure and physical properties of some—though not all—entities, one *cannot* specify the particular way in which a subset of these physical properties become functionally relevant in a given purposive context.[70] So on the (questionable) assumption that structure is fixed, the mapping of structure to function is not merely one or many but one to *unprestatably* many.[71] At the same time, structure *does* constrain the possible space of functions in the sense that it produces an upper bound; unprestatable does not mean unbounded. Opening cans of paint requires objects that have the right physical properties; an object made of an extremely brittle or soft material may never be able to fulfill that function.

One upshot of the view is that it renders function as *designer-intention* (Wright) and *use* (Boorse) perfectly compatible. This is advantageous, as there are numerous well-rehearsed problems with a narrow conception of function as fixed by designer intentions. First, to make sense of exaptation or gain of function, hence the notion of 'accidental function'. There must be a point when this function ceases to be accidental if it comes to be used solely for that purpose, but it is unclear when that line is drawn. Second, it seems unable to account for the fact that many artifact functions are accidentally discovered. Finally, it seems that we cannot rightly say that a natural entity such as a rock has a function, even though the line between designed artifact and found object may be very blurred indeed. A stick used for walking may not be designed, but does it become so if it is cut and trimmed, or if a grip is added to it? Like the artifact "function as use" view, mine avoids these shortcomings while also acknowledging that designer intentions *are* relevant for ascribing function, *if that is the teleological context we are interested in*. My view also benefits from all the strengths of a 'function as use' view, namely, that it does seem perfectly natural to ascribe functions to objects used for a particular purpose they were not designed for (as the notion of an "accidental function" illustrates).

To conclude, the view I am defending here is intended to be contextually specific while also admitting of many possible different contexts. These contexts

---

[70] The structure of some entities, such as some proteins, is not fixed but rather fluctuates depending on context (P. E. Wright and Dyson 1999; see also: Nicholson 2019).

[71] Unprestatability is about the potential activities the objects could be involved in, not the impossibility of inferring the intention behind the creation of something (though that could be true too!).

are the purposive activities of agents in which the involvement of various tools and processes may (or may not) be explained by their conduciveness to achieving a goal. This is because, as previously stated, identifying a function requires saying *for whom* it is a function, and *to what purpose*. An account of function depends on an account of agency.

6.4 Biological Functions

Before concluding this chapter, I would like to briefly discuss the plausibility of extending this account of artifact function to biological functions. While there is no a priori reason to think that a single account of function should apply to both, I follow Wright, and more recently Artiga (2023) and others in thinking that it is a virtue of function accounts for them to be unificatory: "no analysis should begin by supposing that the two sorts [conscious and natural function] are wildly different, or that only one is really legitimate. That is a possible conclusion of an analysis, not a reasonable presupposition" (Wright 1973, 143; see also: Godfrey-Smith 1993; Garson 2018). The account defended here is unificatory because it rests on a conception of teleological explanation which itself depends on biological agency as its necessary condition. The close ties of my account of teleology to my account of function makes it such that since the former is broadly unificatory, the latter necessarily is also (c.f. Toepfer 2008). Furthermore, I deem it important to recognize that our intuitions about artifacts *have* and *do* influence our intuitions about biological function, and therefore that getting the metaphysics of artifacts right can have a positive impact on the way we frame our thinking about biological function. Indeed, it seems likely that the designer-intentions view of artifacts, which is especially intuitive and widespread, at least motivates the idea that natural selection can analogously provide function ascriptions, and it is therefore there that we should look for an account of function. As Michael Ruse claims: "At the heart of modern evolutionary biology is the metaphor of design, and for this reason function-talk is appropriate." (Ruse 2003, 273).

My account of function has important implications for this debate. For one thing, its context-specificity renders it amenable to a restricted form of pluralism consistent with scientific practice. Function statements can rightly be about why some organisms *have* certain features, but unlike selected effects accounts, would not be about the prevalence of a type of feature in a type of organism because such an explanation is causal rather than teleological (Walsh 2014a, 204). Rather, the existence of a token trait in a token organism can rightly be understood as the product of a purposive developmental process, where the generation of said trait was a means toward a particular developmental goal which is robust and plastic across a range of counterfactual conditions. At the same time, a biologist in a different field (say, behavioral ecology) may ascribe different functions to the same trait depending on the kind of teleological explanation they are after (see Boorse 1976, 84-85).[72]

Take the wings of *Egretta ardesiaca*, or the black heron, for instance. While a developmental biologist may ask why they developed in the way that they did, and may appeal to a purposive process of development in the past to answer that question, the ecologist may ask why it is that the heron wraps its wings around its head while standing in a shallow body of water. In this case, the wings function as a way of casting shade onto water in order to both attract fish and for the heron to spot them more easily: the wings contribute to the attainment of that goal in the context of that particular purposive activity. This is what the wings are *for* in that context, in the sense that their use is explained by the goal of catching fish. Neither the developmental biologist nor the ecologist's function ascriptions compete, as they are by nature *local* claims indexed to *particular* activities that are appropriately explained in a teleological manner. By extension, both cases of function attribution are *locally* explanatory. Asking further questions about wing function in flight, or in contexts where it serves as insulation from cold weather, only serve to augment our *overall* understanding of the various functions of wings which are necessitated

[72] Though how we track the identity of a trait across development in practice is no simple matter (DiFrisco, Love, and Wagner 2020).

given the various goal-directed activities that constitute a heron's way of life. All these functional explanations contribute to our understanding of why herons "have" wings, in the most general sense of the term, and why they put them to the uses that they do.

The foregoing makes clear the deep dependence of function ascription on teleological explanation, which in turn require intrinsically purposive agents as their necessary condition. While a process or entity may have a determinate set of physical properties, agents have the capacity to see the relevance of a subset of those properties as being relevant to the attainment of a particular goal. Functions have explanatory relevance in this context: to ascribe function is to explain how that subset of properties, when the object is arranged, positioned, or used in a particular way in a particular context, make it conducive to a goal.

6.5 Conclusion

In this chapter, I've argued that function ascription falls out of the teleological explanation of purposive activities, because functions indicate how a structure or process is a means to the attainment of a goal. In so doing I have shown that a goal-contribution account of function can overcome what Wright—and many after him—saw as its fundamental shortcoming, namely, its inability to distinguish functions from accidents. I've done so by showing how a teleological explanation is different from a chance explanation. And if a course of events is fully explained as the product of chance rather than a purposive process, then no ascription of function will be appropriate. I've also shown how a GCA can meet another one of Wright's desiderata, which is to make sense of the role of design in ascribing functions.

One of the central arguments of this dissertation is that the artifact model of teleology fails: one cannot make sense of the intrinsic purposiveness of the living from the extrinsic purposiveness of the artifactual. I've concluded this argument by showing that the opposite is true: the extrinsic purposiveness or function of an artifact can only be made sense of in the context of an agent's intrinsically purposive activities. Identifying a function requires saying *for whom* it is a function, and *to what purpose*. If I am right that the notion of function is as entwined with teleology as that, then it seems that the reason why other accounts of function fail is because they get the metaphysics of intrinsic purpose wrong or attempt to do away with it altogether.

Conclusion

…the non-recognition of final causation…has been and still is productive of more philosophical error and nonsense than any or every other source of error or nonsense. If there is any goddess of nonsense, this must be her haunt.
— Charles S. Peirce (1903)[73]

This dissertation has shown the shortcomings of the artifact model of teleology and offered a positive alternative based on an account of organisms as intrinsically purposive agents. The artifact model of teleology is founded on the premise that organisms are purposive insofar as they are artifact-like. This conception finds its apotheosis in a conception of evolution as a "blind watchmaker" which designs organisms in much the same way that humans design artifacts. In making such claims, the artifact model collapses the distinction between intrinsic and extrinsic purposiveness. In doing so, it fails to account for agency as the ground of teleology, but also fails to naturalize—rather than eliminate—teleology as a distinctive mode of explanation.

Part 1 of this dissertation aimed to characterize and critique cybernetics and teleonomy as the most prominent artifact models of the 20th century. I argued that the cybernetic account of purposiveness belongs within a broader American tradition of *purposive behaviorism*. I identified three ways in which cybernetics followed on from this tradition. First, it defined purpose and teleology as behaviors, and did so in a way that abstracts away from their causal-mechanical realization. This distinction, characterized by Rosenblueth et al. in terms of

[73] From unpublished manuscript 478, as listed in Robin (1967); cited in Hulswit (1996, 182).

behavioral and functional "modes of study" proved central to their cursory attempt at characterizing the relationship between causal and teleological explanation. Second, purposive behaviorists wrestled with the possibility of making sense of behavior through artifact models and analogies. Cyberneticians strongly pursued this line of thinking, and in doing so abandoned any idea that these models may fail to capture something essential about the nature of purposiveness. Finally, and largely because of these two commitments, purposive behaviorists and cyberneticians struggled to make sense of the distinction between intrinsic and extrinsic purposiveness.

The same is true of teleonomy. Despite the differences in the way the concept was articulated, and the notable differences between it and the cybernetic account of purpose, its proponents struggled against the same set of core problems. Even in its most well developed, Mayrian conception, teleonomy muddled both the intrinsic/extrinsic distinction and the that between causal and teleological explanations.

The shortcomings of the artifact model were made explicit by the continental critics, who also offered a powerful alternative. If organismal agency is the wellspring of intrinsic purpose, then artifacts must be parasitic on and secondary to them. However, while these scholars point the way to solution, they left several problems open.

Part 2 of the dissertation offered a positive account of teleological explanation grounded in the agency of organisms as its necessary condition. I surveyed the AT and the EA as two of them most well developed—though by no means the only—recent attempts at doing so. I argued that while the AT's focus is on spelling out the conditions of teleological explanations in terms of its causal realization, its conception of teleology has a highly restricted content. The EA gives a richer, non-causal account of teleological content, but in doing risks underspecifying the conditions under which it applies. I chose to build on the EA in part because it salvages key aspects of purposive behaviorism but does so precisely because it resolves some of the core tensions in that work. First, it decisively breaks with the temptation to analogize biological purpose with artifact purpose. Second, the EA leverages recent insights about the nature of explanation in the sciences to clarify the distinction between teleological and causal explanations. And third, it opens up new avenues for thinking through the explanatory uses of teleology.

In chapter 5, I argued that the agency of organisms should be characterized in terms of the pursuit of a way of life, and argued that this intrinsically purposive process has its well-spring in a set of dispositions that make up an organism's essence. This led to the view that teleology is a consequence-based explanation, and that on this liberal conception of teleology, it come in at least three varieties: causal-consequence; goal-consequence; and end-consequence. The first of these explains how it is that agents exist in a causal sense, by being causally-mechanically organized so as maintain themselves over time. The second tracks the counterfactual dependencies evident in the gross dynamics of agent behavior. The third is a constitutive explanation that explains why an agent's goal-directed behaviors cohere in the manner that they do, namely, to bring about a particular way of life. And while goal- and end-explanations are forms of non-causal explanation, they are made true by causal relationships. Moreover, I have suggested that all three answer different questions and for that reason need not conflict.

The final chapter suggested that function claims ought to be located within goal-explanations, in so far as attributing function explains why an entity or process is conducive to a goal. Doing so highlights why artifact functions are derivable from biological purposiveness and not vice-versa. This vindicates the idea that any artifact model of teleology was bound to fail: the function of an artifact, and any normative facts it implies, is *fully* derivative from a source of intrinsic purposiveness. This is not to say that artifacts cannot be useful idealized models for organisms. But it remains the case that what makes an artifact purposive is not, and can never be, an intrinsic fact about the artifact; nor any fact, for that matter, which does not appeal to the intrinsic purposiveness of an agent.

Bibliography

Aaby, Bendik Hellem, and Hugh Desmond. 2021. "Niche Construction and Teleology: Organisms as Agents and Contributors in Ecology, Development, and Evolution." *Biology & Philosophy* 36 (5): 47. https://doi.org/10.1007/s10539-021-09821-2.

Abraham, Tara H. 2012. "Transcending Disciplines: Scientific Styles in Studies of the Brain in Mid-Twentieth Century America." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, Special Section I: Assimilating Knowledge – Food and Nutrition in Early Modern Physiologies. II. Psychology, A Science in Crisis? A Century of Reflections and Debates, 43 (2): 552–68. https://doi.org/10.1016/j.shpsc.2012.02.001.

Adams, Frederick R. 1979. "A Goal-State Theory of Function Attributions." *Canadian Journal of Philosophy* 9 (3): 493–518.

Ågren, J. Arvid. 2016. "Selfish Genetic Elements and the Gene's-Eye View of Evolution." *Current Zoology* 62 (6): 659–65. https://doi.org/10.1093/cz/zow102.

Aiken, Howard, John Von Neumann, and Norbert Wiener. 1944. "Letter to H. H. Goldstine," December 28, 1944. Box 4, Folder 66. Norbert Wiener papers, MIT Archive.

Allen, Garland E. 2005. "Mechanism, Vitalism and Organicism in Late Nineteenth and Twentieth-Century Biology: The Importance of Historical Context." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 261–83. https://doi.org/10.1016/j.shpsc.2005.03.003.

Allport, Floyd H. 1937. "Teleonomic Descriptions in the Study of Behavior." *Journal of Personality* 5 (3): 202–14. https://doi.org/10.1111/j.1467-6494.1937.tb02219.x.

Allport, Gordon W. 1921. "Personality and Character." *Psychological Bulletin* 18 (9): 441–55. https://doi.org/10.1037/h0066265.

Amundson, Ron. 1998. "Typology Reconsidered: Two Doctrines on the History of Evolutionary Biology." *Biology & Philosophy* 13 (2): 153–77. https://doi.org/10.1023/A:1006599002775.

Ariew, André. 2003. "Ernst Mayr's 'ultimate/Proximate' Distinction Reconsidered and Reconstructed." *Biology and Philosophy* 18 (4): 553–65. https://doi.org/10.1023/A:1025565119032.

Aristotle. 2016. *Aristotle: De Anima*. Translated by Christopher Shields. Oxford University Press.

Artiga, Marc. 2023. "A Dual-Aspect Theory of Artifact Function." *Erkenntnis* 88 (4): 1533–54. https://doi.org/10.1007/s10670-021-00414-9.

Artiga, Marc, and Manolo Martínez. 2016. "The Organizational Account of Function Is an Etiological Account of Function." *Acta Biotheoretica* 64 (2): 105–17. https://doi.org/10.1007/s10441-015-9256-x.

Ashby, William Ross. 1953. "Homeostasis." In *Cybernetics. Circular Casual and Feedback Mechanisms in Biological and Social Systems, Transactions of the Ninth Conference, March 20 and 21, 1952.*, edited by Heinz Von Foerster, Margaret Mead, and Hans L. Teuber, 73–108. New York: Josiah Macy, Jr. Foundation.

———. 1957. *An Introduction to Cybernetics*. 2nd ed. Chapman & Hall Ltd.

———. 1960. *Design for a Brain: The Origin of Adaptive Behaviour*. Dordrecht: Springer Science & Business Media. http://link.springer.com/10.1007/978-94-015-1320-3.

———. 1962. "Principles of the Self-Organizing System." In *Principles of Self-Organization: Transactions of the University of Illinois Symposium*, edited by Heinz Von Foerster and Zopf W G, 255–78. London: Pergamon Press.

Aumann, Philipp. 2011. "The Distinctiveness of a Unifying Science: Cybernetics' Way to West Germany." *IEEE Annals of the History of Computing* 33 (4): 17–27. https://doi.org/10.1109/MAHC.2011.78.

Austin, Christopher J. 2017. "Aristotelian Essentialism: Essence in the Age of Evolution." *Synthese* 194 (7): 2539–56. https://doi.org/10.1007/s11229-016-1066-4.

———. 2018. *Essence in the Age of Evolution: A New Theory of Natural Kinds*. 1st ed. Routledge. https://doi.org/10.4324/9781351240857.

Austin, Christopher J., and Laura Nuño De La Rosa. 2021. "Dispositional Properties in Evo-Devo." In *Evolutionary Developmental Biology: A Reference Guide*, edited by Laura Nuño De La Rosa and Gerd B. Müller, 469–81. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-32979-6_64.

Ayala, Francisco J. 1970. "Teleological Explanations in Evolutionary Biology." *Philosophy of Science* 37 (1): 1–15. https://doi.org/10.1086/288276.

———. 1972. "The Evolutionary Thought of Teilhard de Chardin." In *Biology, History, and Natural Philosophy*, edited by Allen D. Breck and Wolfgang Yourgrau, 207–16. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-1695-4_14.

Babcock, Gunnar, and Daniel W. McShea. 2021. "An Externalist Teleology." *Synthese* 199 (3): 8755–80. https://doi.org/10.1007/s11229-021-03181-w.

———. 2023. "Goal Directedness and the Field Concept." *Philosophy of Science*, October, 1–10. https://doi.org/10.1017/psa.2023.121.

Baedke, Jan. 2019. “O Organism, Where Art Thou? Old and New Challenges for Organism-Centered Biology.” *Journal of the History of Biology* 52 (2): 293–324. https://doi.org/10.1007/s10739-018-9549-4.

Baedke, Jan, and Alejandro Fábregas-Tejeda. 2023. “The Organism in Evolutionary Explanation: From Early 20th Century to the Extended Evolutionary Synthesis.” Preprint. http://philsci-archive.pitt.edu/21476/.

Baedke, Jan, Alejandro Fábregas-Tejeda, and Guido I. Prieto. 2021. “Unknotting Reciprocal Causation between Organism and Environment.” *Biology & Philosophy* 36 (5): 48. https://doi.org/10.1007/s10539-021-09815-0.

Baggs, Edward, and Anthony Chemero. 2018. “Radical Embodiment in Two Directions.” *Synthese* 198 (S9): 2175–90. https://doi.org/10.1007/s11229-018-02020-9.

Balme, David M. 1987. “Aristotle’s Biology Was Not Essentialist.” In *Philosophical Issues in Aristotle’s Biology*, edited by Allan Gotthelf and James G. Lennox, 291–301. Cambridge; New York: Cambridge University Press.

Barandiaran, Xabier E. 2017. “Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency.” *Topoi* 36 (3): 409–30. https://doi.org/10.1007/s11245-016-9365-4.

Bardin, Andrea. 2015. *Epistemology and Political Philosophy in Gilbert Simondon*. Vol. 19. Philosophy of Engineering and Technology. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9831-0.

Barnes, Jonathan, ed. 1984. *The Complete Works of Aristotle: The Revised Oxford Translation, Volume 1*. Vol. 1. 2 vols. Princeton University Press.

Bates, David W. 2014. “Unity, Plasticity, Catastrophe: Order and Pathology in the Cybernetic Era.” In *Catastrophes*, edited by Nitzan Lebovic and Andreas Killen. Berlin, Boston: DE GRUYTER. https://doi.org/10.1515/9783110312584.32.

Bateson, Gregory. 1979. *Mind and Nature: A Necessary Unity*. New York : Dutton. http://archive.org/details/mindnaturenecess0000bate.

Batterman, Robert W. 2015. “Autonomy and Scales.” In *Why More Is Different*, edited by Brigitte Falkenburg and Margaret Morrison, 115–35. The Frontiers Collection. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43911-1_7.

Batterman, Robert W., and Collin C. Rice. 2014. “Minimal Model Explanations.” *Philosophy of Science* 81 (3): 349–76. https://doi.org/10.1086/676677.

Beatty, John. 1994. “The Proximate/Ultimate Distinction in the Multiple Careers of Ernst Mayr.” *Biology and Philosophy* 9 (3): 333–56. https://doi.org/10.1007/BF00857940.

Bechtel, William, and Adele Abrahamsen. 2011. “Complex Biological Mechanisms.” In *Philosophy of Complex Systems*, 257–85. Elsevier. https://doi.org/10.1016/B978-0-444-52076-0.50009-2.

Bechtel, William, and Leonardo Bich. 2021. “Grounding Cognition: Heterarchical Control Mechanisms in Biology.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1820): 20190751. https://doi.org/10.1098/rstb.2019.0751.

Beckner, Morton. 1959. *The Biological Way of Thought*. Columbia University Press. https://doi.org/10.7312/beck93140.

———. 1969. "Function and Teleology." *Journal of the History of Biology* 2 (1): 151–64. https://doi.org/10.1007/BF00137271.

Bedau, Mark. 1992. "Goal-Directed Systems and the Good." *Monist* 75 (1): 34–51. https://doi.org/10.5840/monist19927516.

Berg, Hein van den. 2013. "The Wolffian Roots of Kant's Teleology." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4): 724–34. https://doi.org/10.1016/j.shpsc.2013.07.003.

———. 2018. "Kant and the Scope of Analogy in the Life Sciences." *Studies in History and Philosophy of Science Part A* 71 (October):67–76. https://doi.org/10.1016/j.shpsa.2017.07.007.

———. 2024. "Explanation, Teleology, and Analogy in Natural History and Comparative Anatomy around 1800: Kant and Cuvier." *Studies in History and Philosophy of Science* 105 (June):109–19. https://doi.org/10.1016/j.shpsa.2024.05.003.

Bich, Leonardo. 2024. "Organisational Teleology 2.0: Grounding Biological Purposiveness in Regulatory Control." *Ratio* n/a (n/a). https://doi.org/10.1111/rati.12405.

Bich, Leonardo, and Argyris Arnellos. 2012. "Autopoiesis, Autonomy, and Organizational Biology: Critical Remarks on 'Life after Ashby.'" *Cybernetics & Human Knowing* 19 (4): 75–103.

Bich, Leonardo, and William Bechtel. 2021. "Mechanism, Autonomy and Biological Explanation." *Biology & Philosophy* 36 (6): 53. https://doi.org/10.1007/s10539-021-09829-8.

———. 2022a. "Control Mechanisms: Explaining the Integration and Versatility of Biological Organisms." *Adaptive Behavior*, January. https://doi.org/10.1177/10597123221074429.

———. 2022b. "Organization Needs Organization: Understanding Integrated Control in Living Organisms." *Studies in History and Philosophy of Science* 93 (June):96–106. https://doi.org/10.1016/j.shpsa.2022.03.005.

Bich, Leonardo, Matteo Mossio, Kepa Ruiz-Mirazo, and Alvaro Moreno. 2016. "Biological Regulation: Controlling the System from Within." *Biology & Philosophy* 31 (2): 237–65. https://doi.org/10.1007/s10539-015-9497-8.

Bich, Leonardo, Matteo Mossio, and Ana M. Soto. 2020. "Glycemia Regulation: From Feedback Loops to Organizational Closure." *Frontiers in Physiology* 11:69. https://doi.org/10.3389/fphys.2020.00069.

Bickhard, Mark H. 2009. "The Interactivist Model." *Synthese* 166 (3): 547–91.

Bickhard, Mark H. 2017. "Information, Representation, Biology." *Biosemiotics* 10 (2): 179–93. https://doi.org/10.1007/s12304-017-9296-5.

Black, Max. (1962) 2019. *Models and Metaphors: Studies in Language and Philosophy*. Cornell University Press.

Boden, Margaret A. 1972. *Purposive Explanation in Psychology*. *Purposive Explanation in Psychology*. Harvard University Press. https://doi.org/10.4159/harvard.9780674188419.

Bokulich, Alisa. 2018. "Searching for Non-Causal Explanations in a Sea of Causes." In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, edited by Alexander Reutlinger and Juha Saatsi, 0. Oxford University Press. https://doi.org/10.1093/oso/9780198777946.003.0008.

Bongard, Joshua, and Michael Levin. 2021. "Living Things Are Not (20th Century) Machines: Updating Mechanism Metaphors in Light of the Modern Science of Machine Behavior." *Frontiers in Ecology and Evolution* 9 (March):650726. https://doi.org/10.3389/fevo.2021.650726.

Boorse, Christopher. 1976. "Wright on Functions." *The Philosophical Review* 85 (1): 70–86. https://doi.org/10.2307/2184255.

Boring, Edwin G. 1944. "Letter to N. Wiener," November 13, 1944. Box 4, Folder 66. Norbert Wiener papers, MIT Archive.

Bowker, Geof. 1993. "How to Be Universal: Some Cybernetic Strategies, 1943-70." *Soc. Stud. Sci.* 23 (1): 107–27. https://doi.org/10.1177/030631293023001004.

Boyd, Richard. 1999. "Homeostasis, Species, and Higher Taxa." https://direct.mit.edu/books/edited-volume/chapter-pdf/2292811/9780262286350_caf.pdf.

Braithwaite, Richard B. 1968. "X: Causal and Teleological Explanation." In *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science (Based upon the Tanner Lectures, 1946)*.

Brandon, Robert N. 1981. "Biological Teleology: Questions and Explanations." *Studies in History and Philosophy of Science Part A* 12 (2): 91–105. https://doi.org/10.1016/0039-3681(81)90015-7.

Brandt, Christina. 2022. "Vitalism, Holism, and Metaphorical Dynamics of Hans Spemann's 'Organizer' in the Interwar Period." *Journal of the History of Biology* 55 (2): 285–320. https://doi.org/10.1007/s10739-022-09682-9.

Bratman, Michael. 2007. "Planning Agency, Autonomous Agency." In , 195–221. https://doi.org/10.1093/acprof:oso/9780195187717.003.0010.

Bratman, Michael E. 2001. "Two Problems About Human Agency." *Proceedings of the Aristotelian Society* 101 (3): 309–26. https://doi.org/10.1111/j.0066-7372.2003.00033.x.

Breitenbach, Angela. 2014. "Biological Purposiveness and Analogical Reflection." In *Kant's Theory of Biology*, edited by Ina Goy and Eric Watkins. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110225792.131.

———. 2016. "Normativity and Purposiveness." *The British Journal of Aesthetics* 56 (4): 405–8. https://doi.org/10.1093/aesthj/ayw051.

Brigandt, Ingo. 2020. "How Are Biology Concepts Used and Transformed?" In *Philosophy of Science for Biologists*, edited by Kostas Kampourakis and Tobias Uller, 1st ed., 79–101. Cambridge University Press. https://doi.org/10.1017/9781108648981.006.

Brigandt, Ingo, Cristina Villegas, Alan C. Love, and Laura Nuño De La Rosa. 2023. "Evolvability as a Disposition: Philosophical Distinctions, Scientific Implications." In *Evolvability*, edited by Thomas F. Hansen, David Houle, Mihaela Pavličev, and Christophe Pélabon, 55–72. The MIT Press. https://doi.org/10.7551/mitpress/14126.003.0006.

Broad, Charlie D. 1925. *The Mind and Its Place in Nature*. The International Library of Psychology, Philosophy and Scientific Method. London ; New York: Kegan Paul, Trench, Trubner & Co., Ltd.

Burbano, Andrés, and Everardo Reyes. 2022. "Capsaicin and Cybernetics: Mexican Intellectual Networks in the Foundation of Cybernetics." *AI & SOCIETY*, February, 1013–25. https://doi.org/10.1007/s00146-021-01337-3.

Calkins, Mary Whiton. 1921. "The Truly Psychological Behaviorism." *Psychological Review* 28 (1): 1–18. https://doi.org/10.1037/h0072853.

Callanan, John J. 2008. "Kant on Analogy." *British Journal for the History of Philosophy* 16 (4): 747–72. https://doi.org/10.1080/09608780802407480.

Canales, Jimena. 2020. *Bedeviled: A Shadow History of Demons in Science*. Princeton, New Jersey: Princeton University Press.

Canfield, John. 1964. "Teleological Explanation in Biology." *The British Journal for the Philosophy of Science* 14 (56): 285–95.

Canguilhem, Georges. (1966) 1991. *The Normal and the Pathological*. New York: Zone Books.

———. (1952) 1992. "Machine and Organism." In *Incorporations*, edited by Jonathan Crary and Sanford Kwinter, translated by Mark Cohen and Randall Cherry, 6:44–69. Zone Books, New York.

———. (1952) 2005. *Knowledge of Life*. Edited by Paola Marrati and Todd Meyers. Translated by Stefanos Geroulanos and Daniela Ginsburg. Fordham University Press, New York.

———. (1989) 2012. *Writings on Medicine*. Translated by Stefanos Geroulanos and Todd Meyers. Fordham University Press, New York.

Cannon, Walter Bradford. 1945. *The Way Of An Investigator, A Scientists Experiences In Medical Research*. http://archive.org/details/in.ernet.dli.2015.115040.

Carroll, David W. 2017. *Purpose and Cognition: Edward Tolman and the Transformation of American Psychology*. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9781316388402.

Chang, Hasok. 2017. "Thermal Physics and Thermodynamics." In *The Oxford Handbook of the History of Physics*, edited by Jed Z. Buchwald and Robert Fox. Vol. 1. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199696253.013.17.

Chardin, Pierre Teilhard de. (1955) 2008. *The Phenomenon of Man*. 1st edition. Harper Perennial Modern Classics.

Charles, David. 1988. "Aristotle on Hypothetical Necessity and Irreducibility." *Pacific Philosophical Quarterly* 69 (1): 1. https://doi.org/10.1111/j.1468-0114.1988.tb00300.x.

Charles, Eric P, ed. 2011. *A New Look at New Realism - The Psychology and Philosophy of EB Holt*. Vol. 1. Transaction Publishers.

Chemero, Anthony, and Stephan Käufer. 2016. "Pragmatism, Phenomenology, and Extended Cognition." In *Pragmatism and Embodied Cognitive Science: From Bodily Intersubjectivity to Symbolic Articulation*, edited by Roman Madzia and Matthias Jung, 57–72. De Gruyter. https://doi.org/10.1515/9783110480238-004.

Chirimuuta, Mazviiita. 2020. "The Reflex Machine and the Cybernetic Brain: The Critique of Abstraction and Its Application to Computationalism." *Perspectives on Science* 28 (3): 421–57. https://doi.org/10.1162/posc_a_00346.

Chirimuuta, Mazviita. 2021. "Your Brain Is Like a Computer: Function, Analogy, Simplification." In *Neural Mechanisms: New Challenges in the Philosophy of Neuroscience*, edited by Fabrizio Calzavarini and Marco Viola, 235–61. Studies in Brain and Mind. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-54092-0_11.

Chiu, Lynn. 2019. "Decoupling, Commingling, and the Evolutionary Significance of Experiential Niche Construction." In *Evolutionary Causation: Biological and Philosophical Reflections*, edited by Tobias Uller and Kevin N. Laland, 299–322. Cambridge, Massachusetts: MIT Press. https://direct.mit.edu/books/edited-volume/4523/chapter/202291/Decoupling-Commingling-and-the-Evolutionary.

Chung, Carl. 2003. "On the Origin of the Typological/Population Distinction in Ernst Mayr's Changing Views of Species, 1942–1959." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 34 (2): 277–96. https://doi.org/10.1016/S1369-8486(03)00026-8.

Churchman, C. W., and R. L. Ackoff. 1950. "Purposive Behavior and Cybernetics." *Social Forces* 29 (1): 32–39. https://doi.org/10.2307/2572754.

Cooper, Andrew. 2018. "Two Directions for Teleology: Naturalism and Idealism." *Synthese* 195 (7): 3097–3119. https://doi.org/10.1007/s11229-017-1364-5.

Cooper, John. 1987. "Hypothetical Necessity in Aristotle's Final Causes." In *Philosophical Issues in Aristotle's Biology*, edited by Allan Gotthelf and James G. Lennox, 243–74. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

Cordeschi, Roberto. 2002. *The Discovery of the Artificial*. Studies in Cognitive Systems. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-9870-5.

Cornell, John F. 1986. "Newton of the Grassblade? Darwin and the Problem of Organic Teleology." *Isis* 77 (3): 405–21.

Corti, Luca. 2023. "Organizational Normativity and Teleology: A Critique." *Synthese* 202 (3): 96. https://doi.org/10.1007/s11229-023-04255-7.

Costall, Alan. 2006. "'Introspectionism' and the Mythical Origins of Scientific Psychology." *Consciousness and Cognition*, Special Issue on Introspection, 15 (4): 634–54. https://doi.org/10.1016/j.concog.2006.09.008.

Craik, Kenneth J. W. 1943. *The Nature of Explanation*. The Nature of Explanation. Oxford, England: University Press, Macmillan.

———. 1966. *The Nature of Psychology*. Cambridge University Press.

Cueto, Marcos. 2016. "An Asymmetrical Network: National and International Dimensions of the Development of Mexican Physiology." *Journal of the History of Medicine and Allied Sciences* 71 (1): 43–63. https://doi.org/10.1093/jhmas/jrv011.

Cummins, Robert. 1975. "Functional Analysis." *Journal of Philosophy*, no. 72, 741–65.

Dawkins, Richard. 1996. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design*. WW Norton & Company.

De Caro, Mario, and David Macarthur. 2004. *Naturalism in Question*. Harvard University Press.

———, eds. 2010. *Naturalism and Normativity*. Columbia Themes in Philosophy. New York: Columbia University Press.

De Jesus, Paulo. 2016. "Autopoietic Enactivism, Phenomenology and the Deep Continuity between Life and Mind." *Phenomenology and the Cognitive Sciences* 15 (2): 265–89. https://doi.org/10.1007/s11097-015-9414-2.

De Laguna, Grace A. 1962. "The Role of Teleonomy in Evolution." *Philosophy of Science* 29 (2): 117–31. https://doi.org/10.1086/287855.

Deacon, Terrence W. 2011. *Incomplete Nature: How Mind Emerged from Matter*. 1st ed. New York: W.W. Norton & Co.

———. 2018. "How Can a Molecule Become about the Relationships between Other Molecules?" June.

———. 2020. "Teleodynamics: Specifying the Dynamical Principles of Intrinsically End-Directed Processes." In *IAISAE International Conference*, 4. Worcester, MA, USA.

Deacon, Terrence W., and Miguel García-Valdecasas. 2023a. "A Thermodynamic Basis for Teleological Causality." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 381 (2252): 20220282. https://doi.org/10.1098/rsta.2022.0282.

———. 2023b. "Teleological Causality as Constraint Propagation."

Delbrück, Max. 1971. "Aristotle-Totle-Totle." In *Of Microbes and Life*, edited by Jacques Monod and Ernest Borek, 50–56. New York: Columbia University Press. http://archive.org/details/ofmicrobeslife0000unse.

Devitt, Michael. 2008. "Resurrecting Biological Essentialism." *Philosophy of Science* 75 (3): 344–82.

———. 2023. *Biological Essentialism*. 1st ed. Oxford University PressOxford. https://doi.org/10.1093/oso/9780198840282.001.0001.

Dewey, John. 1939. *Logic The Theory Of Inquiry*. New York: Henry Holt and Company.

Di Paolo, Ezequiel A., Thomas Buhrmann, and Xabier E. Barandiaran. 2017. *Sensorimotor Life: An Enactive Proposal*. First edition. Oxford, United Kingdom: Oxford University Press.

DiFrisco, James, Alan C. Love, and Günter P. Wagner. 2020. "Character Identity Mechanisms: A Conceptual Model for Comparative-Mechanistic Biology." *Biology & Philosophy* 35 (4): 44. https://doi.org/10.1007/s10539-020-09762-2.

Dobzhansky, Theodosius G. 1968. "Teilhard de Chardin and the Orientation of Evolution: A Critical Essay." *Zygon®* 3 (3): 242–58. https://doi.org/10.1111/j.1467-9744.1968.tb00150.x.

———. 1977. "4. Natural Selection." In *Evolution*, by Theodosius G. Dobzhansky, Francisco J. Ayala, George Ledyard Stebbins, and James W. Valentine. San Francisco : W. H. Freeman. http://archive.org/details/evolution00luti.

Donohue, Christopher, and Charles T. Wolfe, eds. 2023. *Vitalism and Its Legacy in Twentieth Century Life Sciences and Philosophy*. Vol. 29. History, Philosophy and Theory of the Life Sciences. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-12604-8.

Dresow, Max, and Alan C. Love. 2023. "Teleonomy: Revisiting a Proposed Conceptual Replacement for Teleology." *Biological Theory*, January, 101–13. https://doi.org/10.1007/s13752-022-00424-y.

Dretske, Fred I. 1972. "Contrastive Statements." *The Philosophical Review* 81 (4): 411–37.

Dupré, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass: Harvard University Press.

Dupuy, Jean-Pierre. (1994) 2009. *On The Origins of Cognitive Science: The Mechanization of The Mind*. A Bradford Book. Cambridge, Mass: MIT Press.

Egbert, Matthew D., and Xabier E. Barandiaran. 2014. "Modeling Habits as Self-Sustaining Patterns of Sensorimotor Behavior." *Frontiers in Human Neuroscience* 8 (August). https://doi.org/10.3389/fnhum.2014.00590.

Ellis, Brian D. 2002. *The Philosophy of Nature: A Guide to the New Essentialism*. 1. publ. Chesham: Acumen.

Ereshefsky, Marc. 2010. "What's Wrong with the New Biological Essentialism." *Philosophy of Science* 77 (5): 674–85. https://doi.org/10.1086/656545.

Erickson, Paul, Judy L Klein, Lorraine Daston, Rebecca Lemov, Thomas Sturm, and Michael D Gordin. 2013. *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. University of Chicago Press.

Esposito, Maurizio. 2011. "Between Holism and Reductionism : Organismic Inheritance and Neo-Kantian Biological Tradition in Britain and the USA, 1890-1940." Phd, University of Leeds. https://etheses.whiterose.ac.uk/5398/.

———. 2017. "The Organismal Synthesis: Holistic Science and Developmental Evolution in the English-Speaking World, 1915–1954." In *The Darwinian Tradition in Context: Research Programs in Evolutionary Biology*, edited by Richard G. Delisle, 219–41. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-69123-7_10.

Esposito, Maurizio, and Lorenzo Baravalle. 2023. "The Machine-Organism Relation Revisited." *History and Philosophy of the Life Sciences* 45 (3): 34. https://doi.org/10.1007/s40656-023-00587-2.

Esposito, Roberto. 2008. *Bíos: Biopolitics and Philosophy*. U of Minnesota Press.

Fábregas-Tejeda, Alejandro. 2024. "Charting Contrasting Stances on Organismal Purposiveness and Agency in Early Twentieth-Century Biology." In *The Riddle of Organismal Agency: New Historical and Philosophical Reflections*, edited by Alejandro Fábregas-Tejeda, Jan Baedke, Guido I. Prieto, and Gregory Radick, 1st ed., 41–60. London: Routledge. https://doi.org/10.4324/9781003413318.

Feiten, Tim Elmo. 2020. "Mind After Uexküll: A Foray Into the Worlds of Ecological Psychologists and Enactivists." *Frontiers in Psychology* 11 (March):480. https://doi.org/10.3389/fpsyg.2020.00480.

Fiant, Océane. 2018. "Canguilhem and the Machine Metaphor in Life Sciences: History of Science and Philosophy of Biology at the Service of Sciences." *Transversal: International Journal for the Historiography of Science*, no. 4 (June), 149. https://doi.org/10.24117/2526-2270.2018.i4.13.

Foerster, Heinz von. 2003. "Cybernetics of Cybernetics." In *Understanding Understanding: Essays on Cybernetics and Cognition*, edited by Heinz Von Foerster, 283–86. New York, NY: Springer New York. https://doi.org/10.1007/0-387-21722-3_13.

Frank, Lawrence K. 1948. "Foreword." *Annals of the New York Academy of Sciences*, Teleological Mechanisms, 50 (4): 189–96. https://doi.org/10.1111/j.1749-6632.1948.tb39852.x.

Froese, Tom, and John Stewart. 2010. "Life After Ashby: Ultrastability and the Autopoietic Foundations of Biological Autonomy." *Cybernetics & Human Knowing* 17 (4): 7–49.

Fulda, Fermín C. 2016. "Natural Agency: An Ecological Approach." Doctoral Dissertation, University of Toronto.

———. 2017. "Natural Agency: The Case of Bacterial Cognition." *Journal of the American Philosophical Association* 3 (1): 69–90. https://doi.org/10.1017/apa.2017.5.

Galison, Peter. 1994. "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision." *Crit. Inq.* 21 (1): 228–66. https://doi.org/10.1086/448747.

———. 1998. "The Americanization of Unity." *Daedalus* 127 (1): 45–71.

Gambarotto, Andrea. 2020. "Teleology, Life and Cognition: Reconsidering Jonas' Legacy for a Theory of the Organism." In *Natural Born Monads: On the Metaphysics of Organisms and Human Individuals*, edited by P Biasetti A. Altobrando. Berlin: De Gruyter.

Gambarotto, Andrea, and Matteo Mossio. 2022. "Enactivism and the Hegelian Stance on Intrinsic Purposiveness." *Phenomenology and the Cognitive Sciences*, July. https://doi.org/10.1007/s11097-022-09823-9.

Gambarotto, Andrea, and Auguste Nahas. 2022. "Teleology and the Organism: Kant's Controversial Legacy for Contemporary Biology." *Studies in History and Philosophy of Science* 93 (June):47–56. https://doi.org/10.1016/j.shpsa.2022.02.005.

———. 2023. "Nature and Agency: Towards a Post-Kantian Naturalism." *Topoi*, March. https://doi.org/10.1007/s11245-023-09882-w.

García-Valdecasas, Miguel. 2021. "On the Naturalisation of Teleology: Self-Organisation, Autopoiesis and Teleodynamics." *Adaptive Behavior*, February, 1–15. https://doi.org/10.1177/1059712321991890.

García-Valdecasas, Miguel, and Terrence W. Deacon. 2024. "Biological Functions Are Causes, Not Effects: A Critique of Selected Effects Theories." *Studies in History and Philosophy of Science* 103 (February):20–28. https://doi.org/10.1016/j.shpsa.2023.11.002.

Garson, Justin. 2016. *A Critical Overview of Biological Functions*. 1st ed. 2016. Philosophy of Science. Cham: Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-319-32020-5.

———. 2018. "How to Be a Function Pluralist." *The British Journal for the Philosophy of Science*.

———. 2024. "Functions and Populations: Sharpening the Generalized Selected Effects Theory of Function." *Philosophy of Science*, February, 1–29. https://doi.org/10.1017/psa.2024.3.

Gayon, Jean. 2015. "Enzymatic Cybernetics: An Unpublished Work by Jacques Monod." *Comptes Rendus Biologies* 338 (6): 398–405. https://doi.org/10.1016/j.crvi.2015.04.007.

Gelber, Jessica. 2015. "Aristotle on Essence and Habitat." In *Oxford Studies in Ancient Philosophy, Volume 48*, edited by Brad Inwood, 1st ed., 267–94. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198735540.003.0007.

———. 2021. "Teleological Perspectives in Aristotle's Biology." In *The Cambridge Companion to Aristotle's Biology*, 18. Cambridge University Press.

Georgios, Scholarios. 2017. "In Search of a Quote's Origin." Georgios Scholarios. 2017. https://georgiosscholarios.wordpress.com/tag/ernst-wilhelm-von-brucke/.

Gerard, Ralph W. 1951. "Some Problems Concerning Digital Notions in the Central Nervous System." In *Cybernetics: Circular Causal and Feedback Mechanisms in Biological and Social Systems. Transactions of the Seventh Conference, March 23–24, 1950*, edited by Heinz Von Foerster, Margaret Mead, and Hans L. Teuber, 11–58. New York: Josiah Macy, Jr. Foundation.

Geroulanos, Stefanos, and Todd Meyers. 2018. *The Human Body in the Age of Catastrophe: Brittleness, Integration, Science, and the Great War*. University of Chicago Press.

Gibson, James J. 2014. *The Ecological Approach to Visual Perception: Classic Edition*. New York: Psychology Press. https://doi.org/10.4324/9781315740218.

Gijsbers, Victor. 2018. "Reconciling Contrastive and Non-Contrastive Explanation." *Erkenntnis* 83 (6): 1213–27. https://doi.org/10.1007/s10670-017-9937-8.

Ginsborg, Hannah. 2001. "Kant on Understanding Organisms as Natural Purposes." In *Kant and the Sciences*, edited by Eric Watkins, 231–58. Oxford University Press. https://philpapers.org/rec/GINKOU.

———. 2004. "Two Kinds of Mechanical Inexplicability in Kant and Aristotle." *Journal of the History of Philosophy* 42 (1): 33–65.

Glanville, Ranulph. 2002. "Second Order Cybernetics." *Kybernetes* 3:59–85.

Glennan, Stuart. 2017. *The New Mechanical Philosophy*. Oxford, New York: Oxford University Press.

Godfrey-Smith, Peter. 1993. "Functions: Consensus Without Unity." *Pacific Philosophical Quarterly* 74 (3): 196–208. https://doi.org/10.1111/j.1468-0114.1993.tb00358.x.

Goldstein, Kurt. (1939) 1995. *The Organism: A Holistic Approach to Biology Derived from Pathological Data in Man*. Zone Books, New York.

Gotthelf, Allan, and James G. Lennox, eds. 1987. *Philosophical Issues in Aristotle's Biology*. Cambridge; New York: Cambridge University Press.

Grene, Marjorie. 1972. "Aristotle and Modern Biology." *Journal of the History of Ideas* 33 (3): 395–424. https://doi.org/10.2307/2709043.

Griffiths, Paul E. 1996. "Darwinism, Process Structuralism, and Natural Kinds." *Philosophy of Science* 63 (September):S1–9. https://doi.org/10.1086/289930.

———. 1999. "Squaring the Circle: Natural Kinds with Historical Essences." https://direct.mit.edu/books/edited-volume/chapter-pdf/2292823/9780262286350_cah.pdf.

Guyer, Paul. 2001. "Organisms and the Unity of Science." In *Kant and the Sciences*, edited by Eric Watkins, 259–81. Oxford University Press. https://philpapers.org/rec/GUYOAT-2.

Haeckel, Ernst Heinrich Philipp August. 1889. *Natürliche schöpfungsgeschichte : Gemeinverständliche wissenschaftliche vorträge über die entwickelungslehre im allgemeinen und diejenige von Darwin, Goethe und Lamarck im besonderen*. Berlin : George Reimer. http://archive.org/details/natrlichesch00haec.

Hagen, Joel Bartholemew. 2021. *Life out of Balance: Homeostasis and Adaptation in a Darwinian World*. Nexus: New Histories of Science, Technology, the Environment, Agriculture & Medicine. Tuscaloosa: The University of Alabama Press.

Haldane, John S., Edward Stuart Russell, and Leslie Mackenzie. 1923. "IV.—Symposium: The Relations between Biology and Psychology." *Aristotelian Society Supplementary Volume* 3 (1): 56–94. https://doi.org/10.1093/aristoteliansupp/3.1.56.

Haraway, Donna Jeanne. 1976. *Crystals, Fabrics, and Fields: Metaphors of Organicism in Twentieth-Century Developmental Biology*. New Haven and London, Yale University Press.

Harrington, Anne. 1996. *Reenchanted Science: Holism in German Culture from Wilhelm II to Hitler*. Princeton, N.J: Princeton University Press.

Hart, John. (1958) 1980. "Preface." In *On the Mode of Existence of Technical Objects*, translated by Ninian Mellamphy, i–xxiii. University of Western Ontario, Ontario. https://monoskop.org/images/2/20/Simondon_Gilbert_On_the_Mode_of_Existence_of_Technical_Objects_Part_I_alt.pdf.

Hatfield, Gary. 2003. "Behaviourism and Psychology." In *The Cambridge History of Philosophy 1870–1945*, edited by Thomas Baldwin, 640–48. Cambridge: Cambridge University Press.

Hazelwood, Caleb. 2023. "An Emerging Dilemma for Reciprocal Causation." *Philosophy of Science*, October, 1–43. https://doi.org/10.1017/psa.2023.124.

Heft, Harry. 2001. *Ecological Psychology in Context: James Gibson, Roger Barker, and the Legacy of William James's Radical Empiricism*. Psychology Press.

Heims, Steve J. 1991. *The Cybernetics Group*. Cambridge, Mass: MIT Press.

Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15 (2): 135–75. https://doi.org/10.1086/286983.

Hempel, Carl Gustav. 1966. *Aspects of Scientific Explanation, and Other Essays In the Philosophy of Science*. Free Press.

Hesse, Mary. 1988. “The Cognitive Claims of Metaphor.” *The Journal of Speculative Philosophy* 2 (1): 1–16.

Heylighen, Francis. 2022. “The Meaning and Origin of Goal- Directedness: A Dynamical Systems Perspective.” *Biological Journal of the Linnean Society*, 32.

High, Richard P, and William R Woodward. 1980. “William James and Gordon Allport: Parallels in Their Maturing Conceptions of Self and Personality.” *Psychology: Theoretical-Historical Perspectives*, 57–79.

Hobhouse, Leonard T. 1913. *Development and Purpose*.

Hofstadter, Albert. 1941. “Objective Teleology.” *The Journal of Philosophy* 38 (2): 29–39. https://doi.org/10.2307/2018389.

Holt, Edwin B. 1915a. “Response and Cognition I: The Specific-Response Relation.” *The Journal of Philosophy, Psychology and Scientific Methods* 12 (14): 365–73. https://doi.org/10.2307/2013669.

———. 1915b. *The Freudian Wish and Its Place in Ethics*. New York : H. Holt and company. http://archive.org/details/freudianwishitsp00holtiala.

———. 1920. “Professor Henderson’s ‘Fitness’ and the Locus of Concepts.” *The Journal of Philosophy, Psychology and Scientific Methods* 17 (14): 365–81. https://doi.org/10.2307/2939963.

Holt, Edwin B., Walter T. Marvin, W. P. Montague, Ralph Barton Perry, Walter B. Pitkin, and Edward Gleason Spaulding. 1910. “The Program and First Platform of Six Realists.” *The Journal of Philosophy, Psychology and Scientific Methods* 7 (15): 393. https://doi.org/10.2307/2010710.

Holt, Edwin B., Walter Taylor Marvin, and William Pepperell Montague. 1912. *The New Realism : Cooperative Studies in Philosophy*. New York : Macmillan. http://archive.org/details/thenewrealismcoo00unknuoft.

Hu, Jingbo. 2024. “The Disappearing Agent and the Phenomenology of Agency.” *Erkenntnis*, July. https://doi.org/10.1007/s10670-024-00839-y.

Hull, Clark L. 1943. *Principles of Behavior An Introduction to Behavior Theory*. Appleton-Century Crofts, Inc. New York.

Hull, David L. 1974. *Philosophy of Biological Science*. Englewood Cliffs, N.J., Prentice-Hall. http://archive.org/details/philosophyofbiol0000hull.

———. 1994. “Ernst Mayr’s Influence on the History and Philosophy of Biology: A Personal Memoir.” *Biology and Philosophy* 9 (3): 375–86. https://doi.org/10.1007/BF00857944.

Hulswit, Menno. 1996. “Teleology: A Peircean Critique of Ernst Mayr’s Theory.” *Transactions of the Charles S. Peirce Society* 32 (2): 182–214.

Huneman, Philippe. 2014. “Purposiveness, Necessity, and Contingency.” In *Kant’s Theory of Biology*, edited by Ina Goy and Eric Watkins. Berlin, Boston: DE GRUYTER. https://doi.org/10.1515/9783110225792.185.

———. 2018. “Outlines of a Theory of Structural Explanations.” *Philosophical Studies* 175 (3): 665–702. https://doi.org/10.1007/s11098-017-0887-4.

Illetterati, Luca. 2014. “Teleological Judgment: Between Technique and Nature.” In *Kant’s Theory of Biology*, edited by Ina Goy and Eric Watkins. Berlin, Boston: DE GRUYTER. https://doi.org/10.1515/9783110225792.81.

Jacob, François. (1970) 2022. *The Logic of Life: A History of Heredity*. Princeton: Princeton University Press. https://doi.org/10.1515/9780691238999.

Jacob, François, and Jacques Monod. 1961. "Genetic Regulatory Mechanisms in the Synthesis of Proteins." *Journal of Molecular Biology* 3 (3): 318–56. https://doi.org/10.1016/S0022-2836(61)80072-7.

Jaeger, Johannes, Anna Riedl, Alex Djedovic, John Vervaeke, and Denis M. Walsh. 2023. "Naturalizing Relevance Realization: Why Agency and Cognition Are Fundamentally Not Computational," December. https://doi.org/10.31219/osf.io/pr42k.

James, William. 1878. "Remarks on Spencer's Definition of Mind as Correspondence." *The Journal of Speculative Philosophy* 12 (1): 1–18.

———. (1890) 1983. *The Principles of Psychology*. The Works of William James. Cambridge, Mass: Harvard University Press.

Johnson, Christopher. 2015. "'French' Cybernetics." *French Studies* 69 (1): 60–78. https://doi.org/10.1093/fs/knu229.

Johnston, John. 2008. *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI*. Cambridge, MA: MIT Press.

Jonas, Hans. 1951. "Comments on General System Theory: A New Approach to Unity of Science." *Human Biology* 23 (4): 328–35.

———. 1953. "A Critique of Cybernetics." *Social Research* 20 (2): 172–92.

———. 1968. "Biological Foundations of Individuality." *Int. Philos. Q.* 8 (2): 231–51. https://doi.org/10.5840/ipq19688218.

———. (1966) 2001. *The Phenomenon of Life: Toward a Philosophical Biology*. Northwestern University Press.

Jones-Imhotep, Edward. 2020. "The Ghost Factories: Histories of Automata and Artificial Life." *History and Technology* 36 (1): 3–29. https://doi.org/10.1080/07341512.2020.1757972.

Kant, Immanuel. (1790) 2000. *Critique of the Power of Judgment*. Edited by Paul Guyer. Translated by Paul Guyer and Eric Matthews. Cambridge University Press.

Kauffman, Stuart A. 2000. *Investigations*. Oxford University Press.

Kay, Lily E. 2000. *Who Wrote the Book of Life?: A History of the Genetic Code*. Stanford University Press.

Kee, Hayden. 2018. "Phenomenology and Naturalism in Autopoietic and Radical Enactivism: Exploring Sense-Making and Continuity from the Top Down." *Synthese*, June. https://doi.org/10.1007/s11229-018-1851-3.

Keller, Evelyn Fox. 2003. *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines*. 2. print. Cambridge, Mass.: Harvard Univ. Press.

———. 2008. "Organisms, Machines, and Thunderstorms: A History of Self-Organization, Part One." *Historical Studies in the Natural Sciences* 38 (1): 45–75. https://doi.org/10.1525/hsns.2008.38.1.45.

Kennett, Jeanette, and Steve Matthews. 2003. "The Unity and Disunity of Agency." *Philosophy, Psychiatry, & Psychology* 10 (4): 305–12. https://doi.org/10.1353/ppp.2004.0020.

Klein, Alexander. 2018. "The Curious Case of the Decapitated Frog: On Experiment and Philosophy." *British Journal for the History of Philosophy* 26 (5): 890–917. https://doi.org/10.1080/09608788.2017.1378866.

———. 2020. "The Death of Consciousness? James's Case against Psychological Unobservables." *Journal of the History of Philosophy* 58 (2): 293–323. https://doi.org/10.1353/hph.2020.0040.

Kline, Ronald. 2015. *The Cybernetics Moment: Or Why We Call Our Age the Information Age*. New Studies in American Intellectual and Cultural History. Baltimore: Johns Hopkins University Press.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. OUP Oxford.

Koutroufinis, Spyridon A. 2016. "Modern Biological Neo-Teleologism vs Aristotle's Genuine Telos" 6:13.

Kremer, Richard. 2009. "Physiology." In *The Cambridge History of Science: Volume 6: The Modern Biological and Earth Sciences*, edited by John V. Pickstone and Peter J. Bowler, 6:342–66. The Cambridge History of Science. Cambridge: Cambridge University Press. https://doi.org/10.1017/CHOL9780521572019.019.

Krieger, Gerald J. 1998. "Transmogrifying Teleological Talk?" *History and Philosophy of the Life Sciences* 20 (1): 3–34.

Krikorian, Yervant H. 1943. "Life, Mechanism and Purpose." *Philosophy of Science* 10 (3): 184–90.

Krippendorff, Klaus. 2008. "Cybernetics's Reflexive Turns." *Cybernetics and Human Knowing* 15 (3–4): 173.

Kuklick, Bruce. 2001. *A History of Philosophy in America, 1720-2000*. Oxford ; New York: Clarendon Press.

Kuo, Zing-Yang. 1921. "Giving up Instincts in Psychology." *The Journal of Philosophy* 18 (24): 645–64. https://doi.org/10.2307/2939656.

———. 1922. "How Are Our Instincts Acquired?" *Psychological Review* 29 (5): 344–65. https://doi.org/10.1037/h0073689.

———. 1924. "A Psychology without Heredity." *Psychological Review* 31 (6): 427–48. https://doi.org/10.1037/h0072837.

———. 1928. "The Fundamental Error of the Concept of Purpose and the Trial and Error Fallacy." *Psychological Review* 35 (5): 414–33. https://doi.org/10.1037/h0071808.

Laland, Kevin N., Tobias Uller, Marcus W. Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, and John Odling-Smee. 2015. "The Extended Evolutionary Synthesis: Its Structure, Assumptions and Predictions." *Proceedings of the Royal Society B: Biological Sciences* 282 (1813): 20151019. https://doi.org/10.1098/rspb.2015.1019.

Lange, Marc. 2016. *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford Studies in Philosophy of Science. Oxford, New York: Oxford University Press.

LaPorte, Joseph. 2003. *Natural Kinds and Conceptual Change*. Cambridge University Press.

Lawrence, Christopher, and George Weisz. 1998. *Greater Than the Parts: Holism in Biomedicine, 1920-1950*. Oxford University Press.
Le Roux, Ronan. 2018. *Une Histoire de la Cybernétique en France (1948-1975)*. Histoire des techniques 14. Paris: Classiques Garnier. https://doi.org/10.15122/isbn.978-2-406-07295-9.
Lee, Jong Gwan, and Daniel W. McShea. 2020. "Operationalizing Goal Directedness: An Empirical Route to Advancing a Philosophical Discussion." *Philosophy, Theory, and Practice in Biology* 12 (20200929). https://doi.org/10.3998/ptpbio.16039257.0012.005.
Lennox, James G. 1993. "Darwin Was a Teleologist." *Biology & Philosophy* 8 (4): 409–21. https://doi.org/10.1007/BF00857687.
———. 2010a. "Bios and Explanatory Unity in Aristotle's Biology." In *Definition in Greek Philosophy*, edited by David Charles, 329–55. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199564453.003.0010.
———. 2010b. "Bios, Praxis and the Unity of Life." In *Was Ist "Leben"? Aristoteles' Anschauungen Zur Entstehung Und Funktionsweise von Leben*, edited by Sabine Föllinger, 239–57. Philosophie Der Antike. Stuttgart: Franz Steiner Verlag.
———. 2013. "Darwin and Teleology." In *The Cambridge Encyclopedia of Darwin and Evolutionary Thought*, edited by Michael Ruse, 1st ed., 152–57. Cambridge University Press. https://doi.org/10.1017/CBO9781139026895.019.
Lenoir, T. 1982. *The Strategy of Life: Teleology and Mechanics in Nineteenth Century German Biology*. Springer Science & Business Media.
Levins, Richard, and Richard C Lewontin. 1985. "3. The Organism as the Subject and Object of Evolution." In *The Dialectical Biologist*. Aaker Books.
Levis, Nicholas A., and David W. Pfennig. 2020. "Plasticity-Led Evolution: A Survey of Developmental Mechanisms and Empirical Tests." *Evolution & Development* 22 (1–2): 71–87. https://doi.org/10.1111/ede.12309.
Lewens, Tim. 2004. *Organisms and Artifacts: Design in Nature and Elsewhere*. Life and Mind. Cambridge, Mass: MIT Press.
Lewontin, Richard. 2000. "Foreword." In *The Ontogeny of Information: Developmental Systems and Evolution*, edited by Susan Oyama. Durham: Duke University Press.
Lillie, Ralph S. 1942. "Living Systems and Non-Living Systems." *Philosophy of Science* 9 (4): 307–22. https://doi.org/10.1086/286778.
———. 1943. "The Psychic Factor in Living Organisms." *Philosophy of Science* 10 (4): 262–70.
Lipton, Peter. 1990. "Contrastive Explanation." *Royal Institute of Philosophy Supplements* 27 (March):247–66. https://doi.org/10.1017/S1358246100005130.
Loeb, Jacques. 1918. *Forced Movements, Tropisms, and Animal Conduct*. Philadelphia and London, J.B. Lippincott Company. http://archive.org/details/forcedmovements02loebgoog.
Longo, Giuseppe, Maël Montévil, and Stuart Kauffman. 2012. "No Entailing Laws, but Enablement in the Evolution of the Biosphere." In *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation*,

edited by Terrence Soule, 1379–92. Philadelphia Pennsylvania USA: Association for Computing Machinery: New York, NY.
Love, Alan C. 2009. "Typology Reconfigured: From the Metaphysics of Essentialism to the Epistemology of Representation." *Acta Biotheoretica* 57 (1): 51–75. https://doi.org/10.1007/s10441-008-9059-4.
Macarthur, David. 2019. "Liberal Naturalism and the Scientific Image of the World." *Inquiry* 62 (5): 565–85. https://doi.org/10.1080/0020174X.2018.1484006.
Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25.
Malaterre, Christophe, Jean-François Chartier, and Davide Pulizzotto. 2019. "What Is This Thing Called Philosophy of Science? A Computational Topic-Modeling Perspective, 1934–2015." *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, September. https://doi.org/10.1086/704372.
Malone, John C. 2014. "Did John B. Watson Really 'Found' Behaviorism?" *The Behavior Analyst* 37 (1): 1–12. https://doi.org/10.1007/s40614-014-0004-3.
Maturana, Humberto R., and Francisco J. Varela. (1972) 1980. *Autopoiesis and Cognition: The Realization of the Living*. Boston Studies in the Philosophy and History of Science. Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-94-009-8947-4.
Mayr, Ernst. 1955. "Karl Jordan's Contribution to Current Concepts in Systematics and Evolution." *Transactions of the Royal Entomological Society of London* 107 (1–14): 45–66. https://doi.org/10.1111/j.1365-2311.1955.tb00462.x.
———. 1961. "Cause and Effect in Biology." *Science* 134 (3489): 1501–6. https://doi.org/10.1126/science.134.3489.1501.
———. 1974. "Teleological and Teleonomic, a New Analysis." In *Methodological and Historical Essays in the Natural and Social Sciences*, edited by Robert S Cohen and Marx W Wartofsky, 91–117. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-2128-9_6.
———. 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, Mass.: Harvard Univ. Pr.
———. 1988. "The Multiple Meanings of 'Teleological.'" In *Toward a New Philosophy of Biology: Observations of an Evolutionist*, 38–66. Harvard University Press.
McCulloch, Warren S. 1950. "Why the Mind Is in the Head?" *Dialectica* 4 (3): 192–205. https://doi.org/10.1111/j.1746-8361.1950.tb01019.x.
Mcdonough, Jeffrey K. 2009. "Leibniz on Natural Teleology and the Laws of Optics." *Philosophy and Phenomenological Research* 78 (3): 505–44. https://doi.org/10.1111/j.1933-1592.2009.00254.x.
McDonough, Jeffrey K., ed. 2020. *Teleology: A History*. *Teleology*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/oso/9780190845711.001.0001/oso-9780190845711.

McDougall, William. 1911. *Body and Mind: A History and a Defense of Animism*. New York : MacMillan. http://archive.org/details/bodymindhistorya1911mcdo.
———. 1923. "Purposive or Mechanical Psychology?" *Psychological Review* 30 (4): 273–88. https://doi.org/10.1037/h0074955.
McGilchrist, Iain. 2014. "What Happened to the Soul?" RSA, 8 John Adam Street, London WC2N 6EZ.
McKenna, Kenneth Z., Richard Gawne, and H. Frederik Nijhout. 2022. "The Genetic Control Paradigm in Biology: What We Say, and What We Are Entitled to Mean." *Progress in Biophysics and Molecular Biology* 169–170 (March):89–93. https://doi.org/10.1016/j.pbiomolbio.2022.02.003.
McShea, Daniel W. 2012. "Upper-Directed Systems: A New Approach to Teleology in Biology." *Biology & Philosophy* 27 (5): 663–84. https://doi.org/10.1007/s10539-012-9326-2.
Mead, Margaret. 1968. "Cybernetics of Cybernetics." In *Purposive Systems: Proceedings of the First Annual Symposium of the American Society for Cybernetics*, edited by Heinz Von Foerster, John D. White, Larry J. Peterson, and John K. Russell, 1–11. New York: Spartan Books.
Mercer, Christia. 1993. "The Vitality and Importance of Early Modern Aristotelianism." In *The Rise Of Modern Philosophy: The Tension Between the New and Traditional Philosophies from Machiavelli to Leibniz*, edited by Tom Sorell, 33–67.
Merleau-Ponty, Maurice. 1964. *L'Oeil et l'Esprit*.
———. (1961) 2004. "Eye and Mind." In *Maurice Merleau-Ponty: Basic Writing*, edited by Thomas Baldwin, 291–234. London: Routledge.
Meyer, Russell, and Nick Brancazio. 2022. "Putting down the Revolt: Enactivism as a Philosophy of Nature." *Frontiers in Psychology* 13. https://doi.org/10.3389/fpsyg.2022.948733.
Milam, Erika Lorraine. 2010. "The Equally Wonderful Field: Ernst Mayr and Organismic Biology." *Historical Studies in the Natural Sciences* 40 (3): 279–317. https://doi.org/10.1525/hsns.2010.40.3.279.
Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. The MIT Press. https://doi.org/10.7551/mitpress/4124.001.0001.
———. 1989. "In Defense of Proper Functions." *Philosophy of Science* 56 (2): 288–302. https://doi.org/10.1086/289488.
Mills, Simon. 2014. "Gilbert Simondon- Causality, Ontogenesis & Technology." University of the West of England, Bristol.
Mindell, David A. 2002. *Between Human and Machine: Feedback, Control, and Computing Before Cybernetics*. JHU Press.
Mindell, David A, Jérome Segal, and Slava Gerovitch. 2003. "From Communications Engineering to Communications Science: Cybernetics and Information Theory in the United States, France, and the Soviet Union." In *Science and Ideology: A Comparative History*, edited by Mark Walker. London ; New York: Routledge.

Misak, Cheryl. 2013. *The American Pragmatists*. First edition. Oxford History of Philosophy. Oxford: Oxford University Press.

———. 2016. "The Subterranean Influence of Pragmatism on the Vienna Circle: Peirce, Ramsey, Wittgenstein." *Journal for the History of Analytical Philosophy* 4 (5). https://doi.org/10.15173/jhap.v4i5.2946.

Moczek, Armin P., Karen E. Sears, Angelika Stollewerk, Patricia J. Wittkopp, Pamela Diggle, Ian Dworkin, Cristina Ledon-Rettig, et al. 2015. "The Significance and Scope of Evolutionary Developmental Biology: A Vision for the 21st Century." *Evolution & Development* 17 (3): 198–219. https://doi.org/10.1111/ede.12125.

Montagnini, Leone. 2017. *Harmonies of Disorder- Norbert Wiener: A Mathematician-Philosopher of Our Time*. Springer Biographies. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-50657-9.

Montévil, Maël. 2019. "Possibility Spaces and the Notion of Novelty: From Music to Biology." *Synthese* 196 (11): 4555–81. https://doi.org/10.1007/s11229-017-1668-5.

Montévil, Maël, and Matteo Mossio. 2015. "Biological Organisation as Closure of Constraints." *Journal of Theoretical Biology* 372:179–91. https://doi.org/10.1016/j.jtbi.2015.02.029.

Moosavi, Parisa. 2019. "From Biological Functions to Natural Goodness." *Philosophers' Imprint* 19 (51): 1–20.

———. 2022. "Natural Goodness without Natural History." *Philosophy and Phenomenological Research* 104 (1): 78–100. https://doi.org/10.1111/phpr.12751.

Morange, Michel. (1998) 2020. *The Black Box of Biology: A History of the Molecular Revolution*. Translated by Matthew Cobb. Cambridge, Massachusetts: Harvard University Press.

Moreno, Alvaro, and Matteo Mossio. 2015. *Biological Autonomy: A Philosophical and Theoretical Enquiry*. Vol. 12. History, Philosophy and Theory of the Life Sciences. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9837-2.

Morison, Robert S. 1947. "Letter to Rosenblueth," January 17, 1947. Group 1.1 Series 224, Box 1, Folder 3. Rockefeller Archive Center.

Moss, Lenny. 2004. *What Genes Can't Do*. MIT press.

Moss, Lenny, and Stuart A. Newman. 2015. "The Grassblade Beyond Newton: The Pragmatizing of Kant for Evolutionary-Developmental Biology." *Lebenswelt. Aesthetics and Philosophy of Experience.*, 94–111. https://doi.org/10.13130/2240-9599/6686.

Mossio, Matteo, and Leonardo Bich. 2017. "What Makes Biological Organisation Teleological?" *Synthese* 194 (4): 1089–1114. https://doi.org/10.1007/s11229-014-0594-z.

Mossio, Matteo, and Cristian Saborido. 2016. "Functions, Organization and Etiology: A Reply to Artiga and Martinez." *Acta Biotheoretica* 64 (3): 263–75. https://doi.org/10.1007/s10441-016-9283-2.

Mossio, Matteo, Cristian Saborido, and Alvaro Moreno. 2009. "An Organizational Account of Biological Functions." *The British Journal for the Philosophy of Science* 60 (4): 813–41. https://doi.org/10.1093/bjps/axp036.

Nagel, Ernest. 1961. *The Structure of Science- Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World, Inc.

———. 1977. "Goal-Directed Processes in Biology." *Journal of Philosophy* 74 (5): 261–79. https://doi.org/10.2307/2025745.

———. 1979. "Teleology Revisited." In *Teleology Revisited and Other Essays in the Philosophy and History of Science*, 275-. The John Dewey Essays in Philosophy, no. 3. New York: Columbia University Press.

Nahas, Auguste, and Carl Sachs. 2023. "What's at Stake in the Debate over Naturalizing Teleology? An Overlooked Metatheoretical Debate." *Synthese* 201 (4): 142. https://doi.org/10.1007/s11229-023-04147-w.

Nanney, David. 1957. "The Role of the Cytoplasm in Heredity." In *A Symposium on the Chemical Basis of Heredity*, edited by William D. McElroy and Bentley Glass, 134–63. Baltimore, Johns Hopkins Press. http://archive.org/details/chemicalbasisofh0000unse.

Nassar, Dalia. 2016. "Analogical Reflection as a Source for the Science of Life: Kant and the Possibility of the Biological Sciences." *Studies in History and Philosophy of Science Part A* 58 (August):57–66. https://doi.org/10.1016/j.shpsa.2016.03.008.

Neander, Karen. 2017. *A Mark of the Mental: In Defense of Informational Teleosemantics*. Life and Mind. Cambridge, MA: MIT Press.

Nicholson, Daniel J. 2019. "Is the Cell Really a Machine?" *Journal of Theoretical Biology* 477 (September):108–26. https://doi.org/10.1016/j.jtbi.2019.06.002.

Nicholson, Daniel J., and Richard Gawne. 2015. "Neither Logical Empiricism nor Vitalism, but Organicism: What the Philosophy of Biology Was." *History and Philosophy of the Life Sciences* 37 (4): 345–81. https://doi.org/10.1007/s40656-015-0085-7.

Novick, Rose. 2023. *Structure and Function*. 1st ed. Cambridge University Press. https://doi.org/10.1017/9781009028745.

Okasha, Samir. 2002. "Darwinian Metaphysics: Species and the Question of Essentialism." *Synthese* 131:191–213.

———. 2018. *Agents and Goals in Evolution*. First edition. Oxford: Oxford University Press.

Osler, Margaret J. 2001. "Whose Ends? Teleology in Early Modern Natural Philosophy." *Osiris* 16 (January):151–68. https://doi.org/10.1086/649343.

Oyama, Susan, ed. 2000. *The Ontogeny of Information: Developmental Systems and Evolution*. Durham: Duke University Press.

Parkovnick, Sam. 2000. "Contextualizing Floyd Allports's Social Psychology." *Journal of the History of the Behavioral Sciences* 36 (4): 429–41. https://doi.org/10.1002/1520-6696(200023)36:4<429::AID-JHBS8>3.0.CO;2-Q.

Pauly, Philip J. 1987. *Controlling Life: Jacques Loeb & the Engineering Ideal in Biology*. 1st edition. New York: Oxford University Press.

Pearce, Trevor. 2020. *Pragmatism's Evolution: Organism and Environment in American Philosophy*. Chicago ; London: The University of Chicago Press.

Peluffo, Alexandre E. 2015. "The 'Genetic Program': Behind the Genesis of an Influential Metaphor." *Genetics* 200 (3): 685–96. https://doi.org/10.1534/genetics.115.178418.

Pepper, Stephen C. 1923. "Misconceptions Regarding Behaviorism." *The Journal of Philosophy* 20 (9): 242–44. https://doi.org/10.2307/2939766.

Perlman, Robert L. 2000. "The Concept of the Organism in Physiology." *Theory in Biosciences* 119 (3): 174–86. https://doi.org/10.1078/1431-7613-00015.

Perry, Ralph Barton. 1909. "The Hiddenness of the Mind." *The Journal of Philosophy, Psychology and Scientific Methods* 6 (2): 29. https://doi.org/10.2307/2011606.

———. 1917a. "Purpose as Systematic Unity." *The Monist* 27 (3): 352–75.

———. 1917b. "Purpose as Tendency and Adaptation." *The Philosophical Review* 26 (5): 477–95. https://doi.org/10.2307/2178045.

———. 1918. "Docility and Purposiveness." *Psychological Review* 25 (1): 1–20. https://doi.org/10.1037/h0072690.

———. 1921. "A Behavioristic View of Purpose." *The Journal of Philosophy* 18 (4): 85–105. https://doi.org/10.2307/2939384.

Peters, John Durham, and Benjamin Peters. 2016. "Norbert Wiener as Pragmatist." *Empedocles: European Journal for the Philosophy of Communication* 7 (2): 157–72. https://doi.org/10.1386/ejpc.7.2.157_1.

Peterson, Erik L. 2016. *The Life Organic: The Theoretical Biology Club and the Roots of Epigenetics*. Pittsburgh, Pa: University of Pittsburgh Press.

Pfennig, David W., Matthew A. Wund, Emilie C. Snell-Rood, Tami Cruickshank, Carl D. Schlichting, and Armin P. Moczek. 2010. "Phenotypic Plasticity's Impacts on Diversification and Speciation." *Trends in Ecology & Evolution* 25 (8): 459–67.

Pias, Claus, ed. 2016. *Cybernetics - the Macy Conferences 1946-1953: The Complete Transactions*. First printing. Zürich Berlin: Diaphanes.

Pickering, Andrew. 2010. *The Cybernetic Brain: Sketches of Another Future*. Chicago ; London: University of Chicago Press.

Pittendrigh, Colin S. 1958. "Adaptation, Natural Selection, and Behavior." In *Behavior and Evolution*, edited by Anne Roe and George Gaylord Simpson, 390–416. Yale University Press New Haven, CT.

———. 1993. "Temporal Organization: Reflections of a Darwinian Clock-Watcher." *Annual Review of Physiology* 55 (1): 17–54.

Potochnik, Angela. 2017. *Idealization and the Aims of Science*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226507194.001.0001.

Quarfood, Marcel. 2006. "Kant on Biological Teleology: Towards a Two-Level Interpretation." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 37 (4): 735–47. https://doi.org/10.1016/j.shpsc.2006.09.007.

Quine, Willard Van Orman. 1948. *On What There Is*. Catholic University of America, Philosophy Education Society Washington, DC.

Ramsey, Grant, and Cristina Villegas. 2024. "Developmental Channeling and Evolutionary Dappling." *Philosophy of Science*, March, 1–18. https://doi.org/10.1017/psa.2024.5.

Reutlinger, Alexander. 2017a. "Does the Counterfactual Theory of Explanation Apply to Non-Causal Explanations in Metaphysics?" *European Journal for Philosophy of Science* 7 (2): 239–56. https://doi.org/10.1007/s13194-016-0155-z.

———. 2017b. "Explanation beyond Causation? New Directions in the Philosophy of Scientific Explanation." *Philosophy Compass* 12 (2): e12395. https://doi.org/10.1111/phc3.12395.

Reutlinger, Alexander, and Juha Saatsi, eds. 2018. *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*. Oxford, New York: Oxford University Press.

Rheinberger, Hans-Jörg, and Staffan Müller-Wille. 2017. *The Gene: From Genetics to Postgenomics*. Translated by Adam Bostanci. University of Chicago Press. https://doi.org/10.7208/chicago/9780226474786.001.0001.

Rice, Collin. 2015. "Moving Beyond Causes: Optimality Models and Scientific Explanation." *Noûs* 49 (3): 589–615. https://doi.org/10.1111/nous.12042.

———. 2021. *Leveraging Distortions: Explanation, Idealization, and Universality in Science*. MIT Press.

Richards, Robert J. 1987. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press.

Riskin, Jessica. 2016. *The Restless Clock: A History of the Centuries-Long Argument Over What Makes Living Things Tick*. Chicago: The University of Chicago Press.

———. 2020. "Biology's Mistress, a Brief History." *Interdisciplinary Science Reviews* 45 (3): 268–98. https://doi.org/10.1080/03080188.2020.1794388.

Robin, Richard S. 1967. *Annotated Catalogue of the Papers of Charles S. Peirce*. Amherst, University of Massachusetts Press. http://archive.org/details/annotatedcatalog0000robi.

Roe, Anne, and George Gaylord Simpson, eds. 1958. *Behavior and Evolution*. Yale University Press New Haven, CT.

Rose, Anne C. 2016. "William McDougall, American Psychologist: A Reconsideration of Nature-Nurture Debates in the Interwar United States." *Journal of the History of the Behavioral Sciences* 52 (4): 325–48. https://doi.org/10.1002/jhbs.21811.

Rosenblueth, Arturo. 1947. "Letter to Morison," January 25, 1947. Group 1.1 Series 224, Box 1, Folder 3. Rockefeller Archive Center.

———. 1963. "A Critique of Homeostasis." In *Perspectives in Biology; a Collection of Papers Dedicated to Bernardo A. Houssay on the Occasion of His 75th Birthday*, edited by Carl Ferdinand Cori, 323–31. Amsterdam: Elsevier Pub. Co.

Rosenblueth, Arturo, and Norbert Wiener. 1945. "The Role of Models in Science." *Philosophy of Science* 12 (4): 316–21.

———. 1950. "Purposeful and Non-Purposeful Behavior." *Philosophy of Science* 17 (4): 318–26.

Rouse, Joseph. 2023. “The Radical Naturalism of Naturalistic Philosophy of Science.” *Topoi*, February. https://doi.org/10.1007/s11245-023-09885-7.
Royce, Josiah. 1914. “The Mechanical, the Historical and the Statistical.” *Science* 39 (1007): 551–66.
Ruesch, Jurgen, and Gregory Bateson. 1951. *Communication: The Social Matrix of Psychiatry*. New York: W. W. Norton & Company.
Rupik, Gregory. 2024. *Remapping Biology with Goethe, Schelling, and Herder: Romanticizing Evolution*. London: Routledge. https://doi.org/10.4324/9781003441809.
Ruse, Michael. 1971. “Functional Statements in Biology.” *Philosophy of Science* 38 (1): 87–95. https://doi.org/10.1086/288342.
———. 1981. *Is Science Sexist?: And Other Problems in the Biomedical Sciences*. 1981st edition. Dordrecht, Holland ; Boston : Hingham, MA: Springer.
———. 2003. *Darwin and Design: Does Evolution Have a Purpose?* Cambridge, Mass.: Harvard University Press. http://archive.org/details/darwindesigndoes0000ruse.
Russell, Bertrand. 1919. *Introduction to Mathematical Philosophy*. London; New York: George Allen & Unwin, LTD; The Macmillan Co.
Russell, Edward Stuart. 1934a. *The Behaviour of Animals: An Introduction to Its Study*. E. Arnold & Company.
———. 1934b. “The Study of Behaviour.” *Nature* 134 (3396): 835–39. https://doi.org/10.1038/134835a0.
———. 1945. *The Directiveness Of Organic Activities*. Cambridge University Press. http://archive.org/details/directivenessofo035049mbp.
Ruyer, Raymond. 1961. “La Cybernétique et La Finalité.” *Les Études Philosophiques* 16 (2): 165–76.
Sachs, Carl. 2023. “Naturalized Teleology: Cybernetics, Organization, Purpose.” *Topoi*, January. https://doi.org/10.1007/s11245-022-09851-9.
Salmon, Wesley C. 1998. *Causality and Explanation*. Oxford University Press.
Samelson, Franz. 1981. “Struggle for Scientific Authority: The Reception of Watson’s Behaviorism, 1913–1920.” *Journal of the History of the Behavioral Sciences* 17 (3): 399–425. https://doi.org/10.1002/1520-6696(198107)17:3<399::AID-JHBS2300170310>3.0.CO;2-2.
Schapiro, Edited by Tamar, Kyla Ebels-Duggan, and and Sharon Street, eds. 2022. *Normativity and Agency: Themes from the Philosophy of Christine M. Korsgaard*. Oxford, New York: Oxford University Press.
Scheffler, Israel. 1959. “Thoughts on Teleology.” *British Journal for the Philosophy of Science* 9 (36): 265–84.
Sellars, Wilfrid. 1963. “Philosophy and the Scientific Image of Man.” *Science, Perception and Reality* 2:35–78.
Shaw, Robert. 2011. “Ecological Realism as a Reaction to New Realism: Holt’s Legacy to Gibson.” In *A New Look at New Realism: The Psychology and Philosophy of EB Holt*, 157–90. Transaction Publishers.
Shea, Nicholas. 2018. *Representation in Cognitive Science*. Oxford, New York: Oxford University Press.

Short, T. L. 1981. "Peirce's Concept of Final Causation." *Transactions of the Charles S. Peirce Society* 17 (4): 369–82.

Simondon, Gilbert. 2005. *L'individuation à la lumière des notions de forme et d'information*. Krisis. Grenoble: Éditions Jérôme Millon.

———. (1958) 2012. *Du Mode d'Existence Des Objets Techniques*. Éditions Aubier-Montaigne.

———. (1958) 2017. *On the Mode of Existence of Technical Objects*. Translated by Cecile Malaspina and John Rogove. Minneapolis: Univocal Publishing, University of Minnesota Press.

———. 2020. *Individuation in Light of Notions of Form and Information*. Translated by Taylor Adkins. Posthumanities 57. Minneapolis: University of Minnesota Press.

Singer, Edgar A. 1911. "Mind as an Observable Object." *The Journal of Philosophy, Psychology and Scientific Methods* 8 (7): 180–86.

Skow, Bradford. 2014. "Are There Non-Causal Explanations (of Particular Events)?" *The British Journal for the Philosophy of Science* 65 (3): 445–67. https://doi.org/10.1093/bjps/axs047.

Smith, Bradon. 2009. "The Price of Metaphor Is Eternal Vigilance: Language Metaphors in Popular Genetics." *International Journal of the Humanities* 6 (10).

Smith, Charles. 2004. "Wallace's Unfinished Business: The 'Other Man' in Evolutionary Theory." *Complexity* 10 (November):25–32. https://doi.org/10.1002/cplx.v10:2.

Smith, David Livingstone. 2017. "Introduction: Biophilosophy." In *How Biology Shapes Philosophy*, edited by David Livingstone Smith, 1–8. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107295490.001.

Smith, Laurence D. 1982. "Purpose and Cognition: The Limits of Neorealist Influence on Tolman's Psychology." *Behaviorism* 10 (2): 151–63.

Smocovitis, Vassiliki Betty. 1994. "Disciplining Evolutionary Biology: Ernst Mayr and the Founding of the Society for the Study of Evolution and Evolution (1939-1950)." *Evolution* 48 (1): 1–8. https://doi.org/10.1111/j.1558-5646.1994.tb01287.x.

———. 1996. *Unifying Biology: The Evolutionary Synthesis and Evolutionary Biology*. Princeton, N.J: Princeton University Press.

Sober, Elliott. 1980. "Evolution, Population Thinking, and Essentialism." *Philosophy of Science* 47 (3): 350–83. https://doi.org/10.1086/288942.

———. 1983. "Equilibrium Explanation." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 43 (2): 201–10.

Sommerhoff, Gerd. 1950. *Analytical Biology*. Oxford University Press.

Sorabji, Richard. 1964. "Function." *The Philosophical Quarterly* 14 (57): 289–302.

Stein, Nathanael. 2011. "Causation and Explanation in Aristotle." *Philosophy Compass* 6 (10): 699–707. https://doi.org/10.1111/j.1747-9991.2011.00436.x.

Strevens, Michael. 2011. *Depth: An Account of Scientific Explanation*. Harvard University Press.

Stuart, Michael T., and Daniel Wilkenfeld. 2022. “Understanding Metaphorical Understanding (Literally).” *European Journal for Philosophy of Science* 12 (3): 49. https://doi.org/10.1007/s13194-022-00479-5.

Suárez, Javier, and Roger Deulofeu. 2019. “Equilibrium Explanation as Structural Non-Mechanistic Explanations: The Case of Long-Term Bacterial Persistence in Human Hosts.” *Teorema: Revista Internacional de Filosofía* 38 (3): 95–120.

Sultan, Sonia E., Armin P. Moczek, and Denis Walsh. 2021. “Bridging the Explanatory Gaps: What Can We Learn from a Biological Agency Perspective?” *BioEssays*, November, 2100185. https://doi.org/10.1002/bies.202100185.

Tamburrini, Guglielmo, and Edoardo Datteri. 2005. “Machine Experiments and Theoretical Modelling: From Cybernetic Methodology to Neuro-Robotics.” *Minds and Machines* 15 (3–4): 335–58. https://doi.org/10.1007/s11023-005-2924-x.

Taylor, Richard. 1950a. “Comments on a Mechanistic Conception of Purposefulness.” *Philos. Sci.* 17 (4): 310–17. https://doi.org/10.1086/287106.

———. 1950b. “Purposeful and Non-Purposeful Behavior: A Rejoinder.” *Philos. Sci.* 17 (4): 327–32. https://doi.org/10.1086/287108.

Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, Mass: Belknap Press of Harvard University Press.

Thompson, Nicholas S. 1987. “The Misappropriation of Teleonomy.” In *Perspectives in Ethology*, 259–74. Springer.

Thorndike, Edward L. 1898. “Animal Intelligence: An Experimental Study of the Associative Processes in Animals.” *The Psychological Review: Monograph Supplements* 2 (4): i–109. https://doi.org/10.1037/h0092987.

Tibaldeo, Roberto Franzini. 2008. “Beyond Ontological Reductionism, and Towards an Ontological Revolution A Comparison between Norbert Wiener’s Cybernetics, Ludwig von Bertalanffy’s General System Theory, and Hans Jonas.Pdf.” In . Institute for the Study of Nature.

———. 2012. “Hans Jonas’ ‘Gnosticism and Modern Nihilism’, and Ludwig von Bertalanffy.” *Philosophy & Social Criticism* 38 (3): 289–311. https://doi.org/10.1177/0191453710389451.

Toepfer, Georg. 2008. “Teleology in Natural Organized Systems and in Artefacts. Interdependence of Processes versus External Design.” In *Purposiveness: Teleology Between Nature and Mind*, edited by Luca Illetterati and Francesca Michelini, 163–82. De Gruyter. https://doi.org/10.1515/9783110321241.163.

———. 2011. “Kant’s Teleology, the Concept of the Organism, and the Context of Contemporary Biology.” In *Final Causes and Teleological Explanation*, edited by Dominik Perler and Stephan Schmid, 14:107–24. Logical Analysis and History of Philosophy. mentis. https://brill.com/view/book/edcoll/9783957439291/B9783957439291_s009.xml.

———. 2012. “Teleology and Its Constitutive Role for Biology as the Science of Organized Systems in Nature.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 113–19. https://doi.org/10.1016/j.shpsc.2011.05.010.

———. 2021. “Forms as Forces: The Causal Regime of Morphology in Biology.” *Perspectives on Science*, September, 627–42. https://doi.org/10.1162/posc_a_00387.

———. 2023. “‘Organization’: Its Conceptual History and Its Relationship to Other Fundamental Biological Concepts.” In *Organization in Biology*, edited by Matteo Mossio. History, Philosophy and Theory of the Life Sciences. Springer.

Tolman, Edward C. 1925. “Behaviorism and Purpose.” *The Journal of Philosophy* 22 (2): 36–41. https://doi.org/10.2307/2015202.

———. 1928. “Purposive Behavior.” *Psychological Review* 35:524–30. https://doi.org/10.1037/h0070770.

Tresch, John. 2012. *The Romantic Machine: Utopian Science and Technology after Napoleon*. Chicago; London: The University of Chicago Press.

Uller, Tobias, Lisandro Milocco, Jana Isanta-Navarro, Charlie K. Cornwallis, and Nathalie Feiner. 2024. “Twenty Years on from Developmental Plasticity and Evolution: Middle-Range Theories and How to Test Them.” *Journal of Experimental Biology* 227 (Suppl_1): jeb246375. https://doi.org/10.1242/jeb.246375.

Velleman, J. David. 1992. “What Happens When Someone Acts?” *Mind* 101 (403): 461–81. https://doi.org/10.1093/mind/101.403.461.

Villalobos, Mario, and Dave Ward. 2016. “Lived Experience and Cognitive Science: Reappraising Enactivism’s Jonasian Turn.” *Constructivist Foundations* 11 (2): 802–31.

Virenque, Louis. 2022. “From Surviving to Living: A Major Challenge for Autonomy.” Presented at the Outonomy Workshop, June 22nd-24th, University of the Basque Country, San Sebastian, Spain.

———. 2023. “Autonomous Perspective and Ecological Approach: Exploring the Conceptual Connections.” Presented at the ISHPSSB Biennial Meeting, July 9th-15th, University of Toronto, Toronto, Canada.

Virenque, Louis, and Matteo Mossio. 2024. “What Is Agency? A View from Autonomy Theory.” *Biological Theory* 19 (1): 11–15. https://doi.org/10.1007/s13752-023-00441-5.

Von Bertalanffy, Ludwig. 1950. “An Outline of General System Theory.” *Br. J. Philos. Sci.* 1:134–65. https://doi.org/10.1093/bjps/I.2.134.

———. 1951. “Towards a Physical Theory of Organic Teleology: Feedback and Dynamics.” *Human Biology* 23 (4): 346–61.

Von Foerster, Heins, Margaret Mead, and Hans L. Teuber. 1953. “A Note By The Editors.” In *Cybernetics. Circular Casual and Feedback Mechanisms in Biological and Social Systems, Transactions of the Ninth Conference, March 20 and 21, 1952.*, edited by Heinz Von Foerster, Margaret Mead, and Hans L. Teuber, ix–xx. New York: Josiah Macy, Jr. Foundation.

Waddington, C. H. 1968. “Comments by C.H. Waddington Reprinted from Science (1962).” In *The Origin of Life*, 1st ed., 55–56. Routledge. https://doi.org/10.4324/9781315133638-3.

Wallace, Alfred Russel, ed. (1858) 2009. “On the Tendency of Varieties to Depart Indefinitely from the Original Type.” In *Contributions to the Theory of Natural Selection: A Series of Essays*, 2nd ed., 26–44. Cambridge Library Collection - Darwin, Evolution and Genetics. Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/contributions-to-the-theory-of-natural-selection/on-the-tendency-of-varieties-to-depart-indefinitely-from-the-original-type/2FF1878297E79800FE15B167AE5BA192.

Wallace, Alfred Russell. 1866. “Wallace, A. R. to Darwin, C. R.,” July 2, 1866. DCP-LETT-5140. Darwin Correspondence Project, University of Cambridge. https://web.archive.org/web/20101125154407/http://www.darwinproject.ac.uk/entry-5140#back-mark-5140.f5.

Walsh, Denis M. 2006. “Evolutionary Essentialism.” *The British Journal for the Philosophy of Science* 57 (2): 425–48. https://doi.org/10.1093/bjps/axl001.

———. 2012a. “Mechanism and Purpose: A Case for Natural Teleology.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 173–81. https://doi.org/10.1016/j.shpsc.2011.05.016.

———. 2012b. “Situated Adaptationism.” In *The Environment: Philosophy, Science, and Ethics*, edited by William P. Kabasenche, Michael O’Rourke, and Matthew H. Slater. The MIT Press. https://doi.org/10.7551/mitpress/9780262017404.001.0001.

———. 2014a. “Function and Teleology.” In *Evolutionary Biology: Conceptual, Ethical, and Religious Issues*, 193–216. Cambridge University Press.

———. 2014b. “The Affordance Landscape: The Spatial Metaphors of Evolution.” In *Entangled Life*, edited by Gillian Barker, Eric Desjardins, and Trevor Pearce, 4:213–36. History, Philosophy and Theory of the Life Sciences. Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-94-007-7067-6.

———. 2015. *Organisms, Agency, and Evolution*. Cambridge University Press.

———. 2017. “‘Chance Caught on the Wing’ Metaphysical Commitment or Methodological Artifact?” In *Challenging the Modern Synthesis: Adaptation, Development, and Inheritance*, edited by Denis M. Walsh and Philippe Huneman. Vol. 1. Oxford University Press. https://doi.org/10.1093/oso/9780199377176.003.0008.

———. 2018. “Objectcy and Agency: Towards a Methodological Vitalism.” In *Everything Flows*, edited by Daniel J Nicholson and John A Dupre, 416. Oxford, New York: Oxford University Press.

———. 2020. “Action, Program, Metaphor.” *Interdisciplinary Science Reviews* 45 (3): 344–59. https://doi.org/10.1080/03080188.2020.1795803.

Walsh, Denis M., and Gregory Rupik. 2023. “The Agential Perspective: Countermapping the Modern Synthesis.”

Warren, Howard C. 1916a. "A Study of Purpose I." *The Journal of Philosophy, Psychology and Scientific Methods* 13 (1): 5–26. https://doi.org/10.2307/2012418.

———. 1916b. "A Study of Purpose. II Purposive Activity in Organisms." *The Journal of Philosophy, Psychology and Scientific Methods* 13 (2): 29–49. https://doi.org/10.2307/2013020.

———. 1916c. "A Study of Purpose. III: The Role of Purpose in Nature." *The Journal of Philosophy, Psychology and Scientific Methods* 13 (3): 57–72. https://doi.org/10.2307/2013259.

Watson, John B. 1913. "Psychology as the Behaviorist Views It." *Psychological Review* 20 (2): 158.

Weber, Andreas, and Francisco J Varela. 2002. "Life after Kant: Natural Purposes and the Autopoietic Foundations of Biological Individuality." *Phenomenol. Cognitive Sci.* 1 (2): 97–125. https://doi.org/10.1023/A:1020368120174.

West-Eberhard, Mary Jane. 2003. *Developmental Plasticity and Evolution*. Oxford University Press.

———. 2005. "Developmental Plasticity and the Origin of Species Differences." *Proceedings of the National Academy of Sciences* 102 (suppl_1): 6543–49. https://doi.org/10.1073/pnas.0501844102.

White, Roger M., M. J. S. Hodge, and Gregory Radick. 2021. *Darwin's Argument by Analogy: From Artificial to Natural Selection*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108769518.

Wiener, Norbert. 1924a. "Dualism." In *The Encyclopedia Americana*, 9:367. New York; Chicago: The Encyclopedia Americana Corporation. http://archive.org/details/encyclopediaamer09unse_0.

———. 1924b. "Mechanism and Vitalism." In *The Encyclopedia Americana*, 18:527–28. New York; Chicago: The Encyclopedia Americana Corporation.

———. 1924c. "Metaphysics." In *The Encyclopedia Americana*, 18:707–10. New York; Chicago: The Encyclopedia Americana Corporation.

———. 1942. "Letter to J. B. S. Haldane," June 22, 1942. Box 4, Folder 62. Norbert Wiener papers, MIT Archive.

———. 1949. "Letter to Weaver," April 4, 1949. Group 1.1 Series 224, Box 1, Folder 4. Rockefeller Archive Center.

———. 1964. *I Am a Mathematician*. MIT Press.

———. 2018. *Norbert Wiener—A Life in Cybernetics: Ex-Prodigy: My Childhood and Youth and I Am a Mathematician: The Later Life of a Prodigy*. https://doi.org/10.7551/mitpress/11597.001.0001.

———. (1948) 2019. *Cybernetics or Control and Communication in the Animal and the Machine*. https://direct-mit-edu.myaccess.library.utoronto.ca/books/book/4581/Cybernetics-or-Control-and-Communication-in-the.

Wilkinson, John. 1961. "The Concept of Information and the Unity of Science." *Philosophy of Science* 28 (4): 406–13.

Williams, George Christopher. (1966) 2018. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Wimsatt, William C. 1971. "Some Problems with the Concept of 'Feedback.'" In *PSA 1970*, 241–56. Springer. https://link.springer.com/chapter/10.1007/978-94-010-3142-4_17.

———. 1972. "Teleology and the Logical Structure of Function Statements." *Studies in History and Philosophy of Science Part A* 3 (1): 1–80. https://doi.org/10.1016/0039-3681(72)90014-3.

Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.

Winning, Jason. 2023. "Open-Ended Control vs. Closed-Ended Control: Limits of Mechanistic Explanation." University of Pittsburgh: PhilSci Archive. http://philsci-archive.pitt.edu/20803/1/Winning_2021_Open-Ended_Control_vs_Closed-Ended_Control.pdf.

Winther, Rasmus Grønfeldt. 2020. *When Maps Become the World*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226674865.001.0001.

Wisdom, John Oulton. 1951. "The Hypothesis of Cybernetics." *Br. J. Philos. Sci.* 2 (5): 1–24.

Witt, Charlotte. 2015. "In Defense of the Craft Analogy: Artifacts and Natural Teleology." In *Aristotle's Physics: A Critical Guide*, edited by Mariska Leunissen, 107–21.

Witteveen, Joeri. 2015. "'A Temporary Oversimplification': Mayr, Simpson, Dobzhansky, and the Origins of the Typology/Population Dichotomy (Part 1 of 2)." *Studies in History and Philosophy of Biological and Biomedical Sciences* 54 (December):20–33. https://doi.org/10.1016/j.shpsc.2015.09.007.

Wolfe, Charles T. 2015. "Was Canguilhem a Biochauvinist? Goldstein, Canguilhem and the Project of Biophilosophy." In *Medicine and Society, New Perspectives in Continental Philosophy*, edited by Darian Meacham, 197–212. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9870-9_12.

———. 2023. "Varieties of Organicism: A Critical Analysis." In *Organization in Biology*, edited by Matteo Mossio, 41–58. History, Philosophy and Theory of the Life Sciences. Springer.

Wolin, Richard. 2015. *Heidegger's Children: Hannah Arendt, Karl Löwith, Hans Jonas, and Herbert Marcuse*. Princeton, [New Jersey]: Princeton University Press.

Wong, Andy, and Charles T. Wolfe. 2014. "The Return of Vitalism: Canguilhem, Bergson and the Project of Biophilosophy." In *The Care of Life: Transdisciplinary Perspectives in Bioethics and Biopolitics*, edited by Miguel de Beistegui, Giuseppe Bianco, and Marjorie Gracieuse. Rowman & Littlefield.

Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford university press.

Wright, Larry. 1968. "The Case Against Teleological Reductionism." *The British Journal for the Philosophy of Science* 19 (3): 211–23. https://doi.org/10.1093/bjps/19.3.211.

———. 1972. "Explanation and Teleology." *Philosophy of Science* 39 (2): 204–18.

———. 1973. "Function." *Philos. Rev.* 82:139–68.

Wright, Peter E, and H. Jane Dyson. 1999. "Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm." *Journal of Molecular Biology* 293 (2): 321–31. https://doi.org/10.1006/jmbi.1999.3110.

Zammito, John H. 2012. "The Lenoir Thesis Revisited: Blumenbach and Kant." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 120–32. https://doi.org/10.1016/j.shpsc.2011.05.011.