1 Introduction

You, dear reader, are a goal-directed system. Arguably, the snake plants and thermostats in my home are goal-directed systems too. On a much larger scale, the water and rock cycles that you learnt about in science classes at school might well be goal-directed systems. If these are all goal-directed systems, it is appropriate to ascribe functions to their parts. For example, the function of a thermostat's bimetallic strip is to gauge temperature.

By describing these examples in this way, we appear to be ascribing *teleological* properties to such systems. The term 'teleology' is derived from 'telos', the Greek word for 'end' or 'goal'. Teleology as a discipline is thus the study of ends and goals, as well as the related concepts of functions, aims, and purposes. What all these phenomena have in common is that they involve means–end relations of some sort. For instance, a goal-directed system acts in order to bring about a certain end (its goal); something with a function acts in order to contribute to the end (or goal) of a system; and something with a purpose typically acts in such a way as to achieve that purpose. It is also plausible that these teleological concepts are interrelated in various ways. For example, according to a theory explored in Section 2, the concepts of function and goal are intimately connected: something performs a function precisely when it contributes towards a goal of a system.

Teleology is not only of interest to metaphysicians, and I hope this contribution to the Elements series will encourage readers to engage with a range of theoretical themes across philosophy, science, and technology. Moreover, my aim is not only to explain teleology as a discipline, but also to provide my novel take on how teleology arises in the natural world. It is important to understand teleology because teleological concepts are employed by many explanations that are offered both in everyday contexts[1] as well as in the special sciences. Biological examples are often used in discussions of teleology, and I will often employ such examples in this Element. However, talk of functions is also common in (for instance) biochemistry, medicine, psychology, the social sciences and technology, as well as the emerging area of artificial intelligence (AI). Functional explanations are widespread in many of these areas, and many scientists regard such explanations as indispensable (Nagel 1979: 276). One hallmark of many functional explanations is that they are forward-looking: they explain the presence, character, or activity of an item by reference to some possible future outcome that the item tends to bring about. A mundane example

[1] Recent research in experimental philosophy suggests that our 'folk' intuitions about the world are thoroughly teleological (Kertész and Kodaj 2023; Rose and Schaffer 2017; Rose, Schaffer, and Tobia 2020).

of such an explanation is that a knife is sharp because the knife's function is to cut; but many functional explanations are more surprising than this and reflect important scientific discoveries. On the front page of the module guide for a metaphysics course on which I was once a teaching assistant, there was a picture of a donkey curling up its top lip. Amusingly, it appeared that the donkey was laughing or talking – a reference, I suspect, to David Lewis's famous example of donkeys talking in distant possible worlds. I always thought the donkey's pose was accidental, but later I learnt that the function of this lip movement in donkeys is to expose the vomeronasal organ, which has the further function of sensing pheromones in the air.[2] Importantly, because these functions contribute to the survival of donkeys in various respects, those functions also help to explain why these types of lip and organ *are there in the first place*. Thus, many philosophers of science (e.g., Garson 2017, 2019a, 2019b; Mitchell 1993; Neander 1991) have emphasized that many, or even all, biological functions have an important historical or 'backward-looking' explanatory role. We shall explore the details in Section 2, where we discuss the selectionist, causal role and goal-contribution conceptions of function. For current purposes, the important point is just that discoveries of various functional explanations are often startling, illuminating, and significant.

Now, the term 'function' does not always occur explicitly in the teleological explanations offered in biology and other sciences, as when we say, 'bees huddle together for the sake of keeping warm', or 'the bee wiggles its body in order to show other bees where to find pollen'. But locutions such as 'in order to', 'for the sake of', or 'for the purpose of' are nevertheless teleological insofar as they express a means–end relation, just like the concepts of goal, aim, and purpose. This is not to say, however, that all these terms are interchangeable: for instance, I take it that when putting forward various functional explanations, scientists are not implying that all goal-directed systems and subsystems are purposeful agents.

In the course of the twentieth century, functional explanations became more prominent than ever. In the 1960s, for example, biologists addressed questions about the functions of the thymus in vertebrates, and these investigations, among others, led to important work in immunology. Details aside, it turns out that thymus-derived 'T' cells play crucial roles in the production of lymphocytes, which in turn control cellular immunity (Miller 1961, 1971). This research, along with work on the functions of 'B' (bursa) cells, contributed to significant medical research programmes on cancer, autoimmune diseases, and organ transplant rejection (Schaffner 1993: 84).

[2] I am grateful to Andrea Komkov for explaining this example to me.

Perhaps the best-known development in the modern life sciences is the discovery of DNA; and again, it is difficult to explain the biological characteristics of DNA without invoking teleological concepts. For one thing, the functions of DNA contribute to the survival of a species, and therefore explain why DNA is there in the first place. Moreover, it seems to be in the very nature of DNA to encode information about future outcomes for an organism, and thus DNA appears to be inherently end-directed. As Feser puts it:

> in characterizing the DNA of bears, we take it to be relevant to note that it causes them to be furry and grow to a large size, but not that it also thereby causes them to be good mascots for football teams. The genetic information in bear DNA inherently 'points to' or is 'directed at' the first outcome, but not the second. (2009: 47)[3]

It may come as a surprise to students of philosophy and natural science that teleological concepts are deployed in so many branches of science. In the undergraduate classroom we are often told that natural teleology and the associated idea of final causation were eliminated from science and philosophy by Descartes and others during the early-modern scientific revolution. Teleology is thus often regarded as a bygone relic of ancient Aristotelian science. And yet, as noted earlier, many who work in the special sciences take teleological language and functional explanations to be indispensable.[4] Alongside the rise of modern science, over the last half-century there has also been a flourishing *metaphysical* debate about the nature of functions, giving rise to a philosophical cottage industry. This debate considers functions from a wide variety of scientific domains, and not just biology. So where does all this leave the early-modern rejection of teleology?

Here is one answer: when recent philosophers have said that modern science can do without teleology, what they typically meant is that talk of functions and goals in science can, *in principle*, be reduced to talk involving only non-teleological mechanistic concepts. Ernest Nagel is one early influential advocate of this reductive mechanistic approach, and we shall examine his account of goal-directed systems in Section 3. If Nagel is right, then the world is merely teleological with a small 't': scientists are *perfectly entitled* to employ teleological language, but such language is, in principle, dispensable.

[3] This is one way of thinking about the teleological nature of DNA, at any rate. Some caution is needed with such examples because, as a referee has helpfully pointed out, the literature on genetic functions is quite varied. For some different perspectives, see e.g., Doolittle 2013, Germain et al. 2014, and Bellazzi 2022.

[4] McDonough (2020b) also argues that the alleged rejection of teleology by early modern scientists and philosophers has been exaggerated by many historians. According to some recent scholars, even Spinoza is committed to teleology in some form (Sangiacomo 2015).

One reason to be attracted to the Nagelian reductive programme is that teleology is traditionally associated with the Aristotelian doctrine of the four causes and its underlying metaphysics of substantial forms, and many of us are sceptical about various aspects of Aristotelianism.[5] However, a point sometimes overlooked is that we can accept that there are teleological properties in the world without accepting *all* aspects of Aristotle's metaphysics. Aristotle, for example, argued that the world is imbued with inherent causal powers, and according to many Aristotelian scholars all powers – and not just those of goal-directed systems – are thoroughly teleological, that is, end-directed. As Witt puts it, for Aristotle 'A dormant power is intrinsically dependent upon, and teleologically directed toward, activity, or actuality, and that is the character of its being: it exists potentially' (2008: 130; see also Witt 2003). Importantly, one can be a realist about such power properties without accepting other aspects of Aristotelian physics or metaphysics, such as the controversial doctrine of substantial forms.

Realism about powers has attracted significant support in recent philosophy of science. For instance, it has been argued that even in physics it is plausible to think that what scientists are often doing is uncovering the fundamental dispositional properties or 'powers' of entities (e.g., Bird 2007; Cartwright 1992, 1999, 2019; Cartwright and Pemberton 2013; Corry 2019; Ellis 2001; Ellis and Lierse 1994; Kistler 2006; Mumford 2006). For example, the property of mass is often defined in terms of its gravitational and inertial powers. And importantly for our purposes, power properties are arguably needed to make sense of end-directedness in nature—in a way that reductive analyses of teleological statements fail to do. We explore these issues in Section 4. Our ultimate aim is to cast realism about teleology is a new positive light and encourage new work on it. If our proposals are correct, then the powers metaphysics might well rehabilitate a worldview in which reality is teleological with a big 'T'.

Another traditional complaint about teleology is that it implies something like backwards causation. The worry is that if the current presence, character, or activity of an item is being explained by some future outcome, then that outcome must be reaching out from the future in such a way as to have a causal influence on current states of affairs.[6] While some philosophers do entertain the metaphysical possibility of backwards causation, few would regard it as an actual phenomenon, never mind a widespread one. However, the backwards causation worry arguably

[5] In brief, substantial forms are kind universals like *proton* or *donkey*, which are instantiated essentially by their members and help to explain and unify the attributes of individuals. For a detailed discussion of this idea, see Oderberg 2007, and for discussion of a serious problem, see Alvarado and Tugby 2021.

[6] Gunnar Babcock tells me that this (misguided) objection probably originates in Spinoza's *Ethics* (2018/1677).

rests on a misunderstanding. For one thing, teleological explanations do not entail that the relevant future occurrence ever comes about. Clearly, a knife can have the function to cut even if it never has the occasion to cut anything. But if a cutting occurrence never comes to pass, there can be no question of backwards causation occurring. Arguably, then, function statements are at least partly about particular here-and-now capabilities of an item which may or may not be manifested. This idea will come out clearly in Section 4, where we discuss the idea that dispositions or powers are real here-and-now properties of things. This by no means resolves all the problems, of course, because the possibility of unmanifested functions leads to difficult questions about the nature of goal-directedness, which we raise in Sections 3 and 4. However, those difficulties are not difficulties concerning backwards causation.

When discussing theories of function, I shall assume that an adequate account should satisfy some interrelated desiderata. First and foremost, the theory should shed light on how true judgements about functions in science are grounded by aspects of the world around us (see Forber 2020: 260 for a similar thought). Call this the truthmaking desideratum ('TD'):

TD: An adequate theory of function should specify worldly truthmakers for true function claims.

Secondly, and relatedly, the theory should aim at extensional adequacy, so that it is at least roughly consistent with how teleological concepts are deployed in scientific practice and everyday discourse. Call this the extensional adequacy desideratum ('EAD').

EAD: An adequate theory of function should (at least roughly) preserve the extension of function concepts as they are ordinarily used.

Finally, I follow Garson (2016: Ch. 1.2) and Forber (2020: 261–262) in thinking that an adequate theory of functions should accommodate the normative and explanatory dimensions of function statements. That is, the theory should shed light on the various explanatory roles of function statements in science and also underwrite the normative distinctions between normal and accidental function, and between non-function and malfunction. Call these the explanatory desideratum ('ED') and the normative desideratum ('ND'), respectively.

ED: An adequate theory of function should accommodate, and shed light on, the explanatory roles of function statements in science.

ND: An adequate theory of function should accommodate, and shed light on, the distinctions between normal and accidental function, and between non-function and malfunction.

In the following sections, I discuss these desiderata further and make a case for thinking that a goal-contribution theory of functions (suitably understood) can satisfy them. Where further work is required, which will be inevitable at some points, this will be indicated.

I have had to make some difficult decisions about what to include and what to exclude in this Element. Teleology has a long history, going back at least to Plato and Aristotle, via (among others) Aquinas and the Scholastics in the medieval period, and major figures in early modern philosophy such as Hegel (2010/1816), Kant (2007/1790), and Schelling (2000/1800). Some historical theories of teleology are naturalistic, while others have been grounded in theism. Outside of metaphysics, teleology has also played a prominent role in disciplines including aesthetics, AI, cosmology, environmental philosophy, epistemology and perception, ethics, linguistics, metaethics, philosophy of action, philosophy of language, philosophy of mind, and philosophy of religion. This is not a long Element, and it would be unrealistic here to attempt even a brief survey of work on teleology in such a wide variety of intellectual periods, traditions, and disciplines. Moreover, excellent historical surveys have been provided elsewhere in the philosophical literature.[7] I have therefore confined this Element to work on natural teleology in recent metaphysics of science, which focuses mainly on the concepts of functions, goals, and powers. This approach is (moderately) naturalistic and therefore I shall, for example, have little to say about theistic accounts of teleology or teleological arguments for creationism, in which teleology is ultimately grounded extrinsically in some divine or supernatural power.[8] Nor shall I have much to say about alleged evaluative aspects of teleology. For some neo-Aristotelians, teleology is bound up with the evaluative notion of goodness, insofar as the good is that which fulfils a thing's natural end (for discussion, see e.g., Bedau 1992a; McLaughlin 2001: Ch. 9; Oderberg 2020; Page 2021; Sorabji 1964). The question of whether function statements entail a value judgement is, however, a difficult one that rests upon complex issues in value theory and metaethics. I have therefore not attempted to address the evaluative dimensions of teleology (or lack thereof) in this Element.

Here, then, is a roadmap of what follows: Section 2 offers an overview of the modern debate about functions, which first began to take shape in the 1970s.

[7] For further discussion of historical work on teleology in various intellectual periods, see for example, Feser 2014, 2019: Ch. 6, McDonough 2020a, and Ransome Johnson 2005.

[8] This is not to say that the theories discussed here are incompatible with theology. Indeed, many theistic approaches are inspired by aspects of Aristotelianism and realism about powers. For example, see Oliver 2013, Page 2015, and Schmid 2011 for recent discussions of theistic teleology in the Thomistic tradition.

Here we discuss the selectionist, causal role, and goal-contribution approaches. We make a preliminary case for the goal-contribution theory and show that it is consistent with many recent theories of function in the philosophy of biology and other scientific domains. Although the goal-contribution theory accommodates functions that are not naturally selected, it can readily acknowledge the importance of selected functions in biology and the role they play in explaining the very existence of function bearers. Section 3 then focuses on the concepts of goals and goal-directedness. It critically examines the reductive cybernetic theory of goal-directed systems that became popular in the middle part of the twentieth century. Section 4 then develops a non-reductive realist theory of powers and end-directedness, with a view to shed light on many cases of goal-directedness.

Each of the Sections in this Element can to some extent be read in isolation. However, they are also related in important ways, and the ordering of the Sections is not accidental. The goal-contribution theory of function, discussed in Section 2, depends on the concepts of goals and goal-directedness. Hence, it is necessary to delve deeper into those concepts in Section 3. However, we shall see that some traditional attempts to analyse goal-directedness face serious problems. This leads us to Section 4, where we consider whether a realist metaphysics of powers can shed light on goal-directedness, and end-directedness more generally. If the powers theory can be shown to do this, then we will be able to bring to light an important new benefit of the powers metaphysics. In the course of the discussion, it is also our aim to open up new and fruitful avenues of debate within the metaphysics of teleology.

2 Functions

2.1 The Concept of Function

As explained in the introductory section, teleology, as I shall approach it, is the study of a cluster of interrelated concepts including those of function and goal. This section uses a discussion of the philosophy of functions as a springboard for the metaphysical investigation of goal-directedness that we begin in Section 3.

Talk of functions is pervasive. In everyday contexts we readily ascribe functions to the artefacts around us, such as a chair or toothbrush. Functions are also part of the explanatory practices of many branches of science, particularly in biology, medicine, and technology. Some argue that functions can even be ascribed in cases of physical, chemical, biochemical, and sociopolitical systems such the water and rock cycles, autocatalysis, Bénard cells, vitamin

B12, mineral species, and welfare systems.[9] If we are to take all this talk of functions at face value, we require an analysis of functions which explains how they arise in the natural world.

We start this section by offering a critical survey of two popular philosophical approaches to functions which are sometimes interpreted as trying to provide a *fully general* analysis of functions. These analyses received much attention in the 1970s and are still discussed today. We begin with the selectionist or 'etiological' approaches (e.g., Garson 2017, 2019a, 2019b; Mitchell 1993; Neander 1991; Wimsatt 1972; Wright 1973), before moving on to the causal role view (e.g., Cummins 1975). There is by now a vast literature on the philosophy of functions, with sophisticated versions of each approach. There are also hybrid theories which incorporate elements of both the selectionist and causal role views, such as the organizational theories of McLaughlin (2001) and Mossio et al. (2009). It is not our aim here to undertake a broad survey of all the various versions and their nuances;[10] we shall, however, try to go into enough detail to provide preliminary support for the idea that our theory of functions requires a sufficiently developed notion of goal-hood. In their basic forms, neither the selectionist nor the causal role analyses place much weight on the notion of goals. But this creates a problem on both sides for those who want their theory to provide a single, fully general account of functions across the board. On the one hand, selectionist analyses are arguably narrow because they apply mainly to the biological (and perhaps technological) domains, and it appears that in many domains function ascriptions are appropriate in cases lacking the appropriate causal history of selection (Section 2.2). This limitation motivates the causal role analysis, which accommodates functions that are not naturally selected. Unfortunately, however, the causal role theory arguably goes too far the other way and *over-generates* functions (Section 2.3).

In Section 2.4, we see how this dilemma can be avoided by employing the notion of goal-directedness. In order for some causal feature of an entity to count as a function, it must be one that contributes to the goal(s) of a system. We shall follow others in calling this the 'goal-contribution' theory of functions. Importantly, the concept of goal-directedness is not so narrow that it applies only to naturally selected functions and biological systems. On the contrary, as we shall see in Sections 3 and 4, influential analyses of goal-directedness (e.g., Babcock 2023) allow us to posit functions in wholly inorganic systems. What

[9] For a recent discussion of functions in biochemistry, see Bellazzi (forthcoming). On the alleged teleological features of mineral species, see Babcock 2023. For discussion of a classic debate on functional explanation in social science, see Schwartz 1993.

[10] More comprehensive critical appraisals of the various theories of function are provided in Nissen 1997 and Garson 2016.