Goal Attributions in Biology: Objective Fact, Anthropomorphic Bias, or Valuable Heuristic?

Samir Okasha

Affiliation: Department of Philosophy, University of Bristol, United Kingdom

Abbreviated title: Teleonomy, Agency, Unity of Purpose

Corresponding author: Samir Okasha, Department of Philosophy, Cotham House, Bristol BS6 6JL, Bristol, U.K. Tel.: + 44 7866670747. Email: Samir.Okasha@bristol.ac.uk

Word count: 9811

Acknowledgements: This paper is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 101018533).

Abstract

Goal-attributing statements – that attribute a goal or endpoint to an organismic activity or process – arise in three different biological contexts. The first context is the mid-20th century debate among biologists and philosophers over how to understand the apparent "goal-directedness" in the living world. The second context is the debate in cognitive ethology over whether non-human animals are capable of "goal-directed behavior", that is, behavior that results from having a mental representation of a goal state. The third is the practice common in evolutionary biology of treating evolved traits, including behaviors, as means by which an organism furthers its overall goal of survival and reproduction (or maximization of fitness). In each of these contexts, a similar philosophical issue arises: are the goal-attributing statements literally true? And if not, do they represent an anthropomorphic bias that should be expunged from science, or a valuable heuristic?

Goal attributions in biology:
objective fact, anthropomorphic bias, or valuable heuristic?

1. Introduction

The label "teleonomy" was introduced into biology by Pittendrigh (1958) and subsequently taken up by Williams (1966), Mayr (1974) and others – though they did not all define it in the same way. The original point of the teleonomy concept was to demarcate a notion of goal-directedness that was both objective and scientifically important, and to sharply distinguish it from discredited forms of teleology, such as the idea that the evolutionary process itself unfolds in accordance with a plan. As Corning (2022) explains, the teleonomy/teleology distinction was meant to capture the idea that living organisms exhibit a kind of "internal purposiveness" in both their ontogenetic development and their daily activities, that has evolved by natural selection. So, while Darwinism forces us to reject the idea that the process of evolution is in any sense goal-directed, this is quite compatible with recognizing that the products of that process – well-adapted organisms and their parts – exhibit goal-directed behavior. In short, we should not throw out the valid forms of teleology with the invalid.

The notion that organisms exhibit an internal purposiveness was endorsed by many mid-twentieth century biologists. Thus, for example, Monod (1973) wrote: "one of the most fundamental characteristics common to all living thing [is] that of being endowed with a project or purpose" (9). In a similar vein, Waddington (1957) wrote that "most of the activities of a living organism are of such a kind that they tend to produce a certain characteristic end result'', a phenomenon he referred to as "directiveness" (2). A more explicit statement came from Mayr (1988): "goal-directed behavior . . . is extremely widespread in the organic world; for instance, most activity connected with migration, food-getting, courtship, ontogeny, and all phases of reproduction is characterized by such goal orientation" (1988: 45). These quotations convey a reasonably clear sense of what was meant by teleonomy.

The concept of goal-directedness receives relatively little attention in contemporary philosophy of biology, by sharp contrast with the concept of function which continues to be a focal point.[1] It was not always thus. In the philosophical literature of the mid- to late- Twentieth Century, function and goal-directedness enjoyed equal airtime and were often discussed together (Sommerhoff, 1950; Braithwaite, 1953; Beckner, 1969; Nagel, 1977a,b; Woodfield, 1976; Wright, 1976). What explains this change? To some extent it may reflect the changing scientific climate, in particular the decline in prominence of cybernetics. Garson (2008: 539) suggests that goal-directedness fell from favour in philosophy because the prominent attempts to analyse it suffered from "conceptual shortcomings." In addition, the philosophical literature's re-orientation away from goal-directedness and towards function likely stemmed in part from the conviction that while "function talk" is ubiquitous in day-to-day biological practice, "goal talk" is neither. So naturalistically-inclined philosophers, who often conceive their task as (in part) to analyse the meaning of scientific terms, were led to focus on function rather than goal-directedness.

My aim in this paper is to reconsider goals and goal-directedness from (what I hope is) a new angle. I examine three biological contexts in which goal attributions arise, all of which are of philosophical interest. (By a "goal attribution" I mean a statement of the form "the goal of *x* is *y*", where *x* stands for an organismic behavior or process and *y* stands for an endpoint. In some cases, *x* may stand for a whole organism.) The first context is the classic mid-twentieth century debate, alluded to above, about how to understand the apparent purposiveness of organismic activities and processes (including development). The second context is the ongoing debate about whether non-human animals are capable of goal-directed behavior, meaning (roughly) behavior that results from having a mental representation of a goal. This is part of the broader question of animal intentionality, long a source of controversy among ethologists and philosophers. The third context is the practice found in evolutionary biology of treating evolved organisms *as if* they were rational agents pursuing a goal – as when we explain why female rats kill their malformed offspring by saying that they know the offspring won't survive and don't want to waste resources on them. This mode of explanation is commonly used by evolutionists studying the adaptive significance of organismic traits.

These three contexts are separate, involving goal attributions of different sorts, and raise distinct philosophical issues. However, a common theme runs through them all, which is whether the goal attributions in question state objective facts about the world, or not? Many authors have assumed that the answer is yes, but this assumption is not inevitable. An alternative is that goals are "projected" onto the biological world by humans. One version of this projectivist view holds that goal attributions reflect an anthropomorphic tendency and so have no place in science. Another version holds that goal attributions are of heuristic value even if they are not literally true. The main aim of this paper is to try to adjudicate between these views, in each of three contexts described above.

The structure of the paper is as follows. Section 2 offers a re-examination of the traditional debate on teleonomy and goal-directedness, focusing on the work of Mayr and E. Nagel. Section 3 examines the ethological debate over whether animal behavior is ever goal-directed in the sense of stemming from belief-like and desire-like mental representations. Section 4 discusses the practice of treating evolved organisms as if they were rational agents pursuing a goal, in the context of evolutionary explanation. Section 5 draws together the pieces and concludes.

2. The Traditional Debate Over Goal-directedness: A Retrospective

I start by briefly outlining Mayr's views on goal-directedness, juxtaposing them with the views of Nagel. Though divergent, Mayr's and Nagel's positions share certain presuppositions, and their respective analyses touch on most of the important points in the traditional debate.

2.1 Mayr's Views

Mayr (1992) sets out his mature views in an essay that builds on ideas he first developed in the 1970s. His aim is to distinguish between various categories of biological phenomena to which the label "teleological" has been applied. One such category is what Mayr calls "teleonomic processes", which he describes as "goal-directed" and "goal-oriented" (Mayr, 1992:52). In addition to processes, Mayr also applies the adjective teleonomic to activities and behaviors. Though Mayr does not try to define "process", "activity" or "behavior", his examples show that the category of the teleonomic as he conceives it is broad. It includes ontogenetic processes such as gastrulation, physiological processes such as thermoregulation, and whole-organism behaviors such as migration and mating displays. Quite often, Mayr uses "process" as a catch-all term that subsumes activities and behaviors, a policy I will follow here in the interest of brevity.

Teleonomic processes have two key features, according to Mayr. Firstly, they have an "endpoint, goal, or terminus", the attainment of which leads the process to stop (Mayr, 1992:52). This endpoint might be a developmental structure, physiological state, or behavioral outcome, Mayr tells us. Thus, the goal of meiosis is to produce haploid germ cells; the goal of a bacterium's movement is to reach an area of higher oxygen; and the goal of the salmon's homing behavior is to arrive at its natal stream. Secondly, the process is guided by an evolved *program* that encodes a set of instructions for the unfolding of the process, and which thus encodes the goal. The significance of this for Mayr is that it dispels any spectre of backwards causality (that haunts other forms of teleology), since the program exists before the teleonomic process begins to unfold. The attribute of being program-guided thus shows how to reconcile goal-directedness with ordinary causality.

By an evolved program Mayr is primarily thinking of a genetic program, in the sense of a set of instructions written in DNA. He alludes to the well-known idea that natural selection, acting over phylogenetic time, has led to the accumulation of information in a species' genome, which then guides ontogenetic development and gives rise to species-specific activities and behaviors. However, Mayr emphasizes that adaptive behavior in higher animals is influenced by learning as well as genes; he thus distinguishes between a "closed program", which contains genetically hard-wired instructions, and an "open program" which can be modified by acquired information. Most complex animal behaviors are the result of an open program, he claims.[2]

Mayr contrasts his account of goal-directedness with a rival analysis, inspired by cybernetics, that defines it in terms of negative feedback, as for example in Rosenbleuth et al. (1943). Mayr allows that negative feedback often plays a role in explaining how the endpoint of a teleonomic process is reached despite perturbations, but denies that is the essence of such a process; rather it is something that improves its precision. The essential feature, he argues, is rather that a mechanism exists that initiates the goal-directed process, and this occurs because of the program-directed nature of the process.

While most organismic behavior counts as teleonomic for Mayr, he reserves the label "purposive behavior" for behavior that involves a deliberate attempt to attain a goal – that is, where the organism has a mental representation of the goal and performs the action in order to attain it. Following the lead of cognitive ethologists, Mayr suggests that such purposive behavior is not confined to humans but is common among mammals and birds too. Though Mayr puts purposive behavior in a separate category from teleonomic behavior, he describes them both as goal-directed; so, they are species of a single genus.

Mayr puts "adapted features" in a separate category again. These are structural and morphological attributes of organisms, including whole organs, that have evolved by natural selection. Mayr emphasizes that such features "do not involve movements", hence in his view are not appropriately described as teleological. However, Mayr does see a link between adapted features and teleonomy, for such features *perform* teleonomic processes and are thus "executive organs" of such processes (Mayr, 1992: 58) (So, circulation of the blood is a teleonomic process of which the heart is an executive organ.) Though he does not put it this way, in effect Mayr's point is that "static" adaptive features are not in the right ontological category to count as goal-directed, for they are not activities, behaviors or processes. This seems right: it makes good sense to enquire what a stag's antlers are *for* but no sense to ask what the antlers' *goal* is.

Mayr's analysis of goal-directedness is interesting, but from a modern perspective it contains questionable aspects, in particular its reliance on the notion of genetic program; this is discussed below.

2.2 Interlude: function and goal

Mayr's point about adapted features shows that a biological item's having a goal, in the sense that he is trying to capture, cannot be equated with having a function (in any of the usual senses of that term). A biological trait of any sort is a candidate for having a function; but only a sub-class of traits are candidates for having goals. Indeed, this observation was a commonplace among mid-twentieth century writers on teleology, who emphasized that goal-directedness and functionality were distinct aspects of teleology that should not be confused (Beckner, 1969; Wright, 1973; Woodfield, 1976; Nagel, 1977a). But what exactly is their relation?

One natural suggestion is that function and goal will coincide for any biological item that has both. That is, if we pick something that *is* in the right ontological category to have a goal, such as an organismic behavior, then the behavior's function, if it has one, will in general be its goal and vice-versa. This sounds plausible, for observe that the pair of statements "the function of the bird's dance is to attract mates" and "the goal of the bird's dance is to attract mates" seem practically synonymous.

However, an argument of Wright (1973) suggests that this coincidence does not always hold. Wright gives the example of freshwater plankton who diurnally vary their distance below the surface. He writes: "the goal of this behavior is to keep light intensity in their environment relatively constant. This can be determined by experimenting with artificial light sources. The function of this behavior, on the other hand, is keeping constant the oxygen supply, which normally varies with sunlight intensity" (Wright, 1973: 140). The moral, Wright says, is that the function of an organismic behavior need not be the behavior's goal but can instead be "some natural concomitant or consequence of the…goal" (ibid.).

Wright is partly correct here (though the biological details are not quite as he says). Light intensity is the (main) cue to which the plankton's daily vertical movement responds, but the selective advantage that the behavior confers – hence the function s*ensu* the standard "selected effect" theory of function – is likely some correlate of light intensity, such as predator avoidance.[3] So, it is indeed possible that a behavior's function may be a "natural concomitant", i.e., correlate, of its goal. However, the other possibility that Wright mentions – that the function be a downstream consequence of the goal – is less clear-cut. For consider a causal sequence of the form $b \rightarrow g \rightarrow c$, where *b* is the behavior, *g* the goal and *c* the consequence. If *c* confers a selective advantage, i.e., enhances present and/or past fitness, and so is a candidate for being *b*'s function, then the same must be true of *g* too. It is a familiar point that singling out an effect as "the" function of a trait involves a partly conventional decision about how to chop up a complex causal chain. (The same is likely true of "the" goal.) Thus, in causal sequences of this sort, it will always be possible to identify goal and function so long as the behavior in question does have a goal.

One final difference between function and goal deserves mention. Where a goal is attributed to an organismic behavior ,(though not to an internal process), it is generally possible to re-express this in terms of the organism's goal. Thus, instead of saying that the goal of a swallow's migration is to reach warmer climes, we can equivalently say that the swallow's goal (in flying south) is to reach warmer climes. Functional attributions, by contrast, cannot always be so re-expressed.

2.3 Nagel's tripartite taxonomy

It is useful to contrast Mayr's view of goal-directedness with alternative approaches that were prevalent at the time. In a well-known paper, Nagel (1977a) laid out three views of goal-directedness that he calls "the *intentional* view, the *program* view and the *system-property* view". Nagel criticises the first two views and defends the third. The intentional view posits a close link between goal-directedness and intentionality. Proponents of this view argue that the paradigm goal-directed phenomenon is the conscious pursuit of a goal by an intelligent agent who has a mental representation of the goal they wish to achieve. Nagel recognizes that the intentional view is attractive because it avoids the threat of reverse causality: the cause of a goal-directed action is not the goal itself but rather the agent's intention to bring the goal about. However, he regards the intentional view as ill-suited as a general analysis of goal-directedness, for it cannot make sense of goal attributions "in connection with organisms…which are incapable of having intentions and beliefs, in connection with subsystems of organisms; or in connection with inanimate systems" (Nagel, 1977a: 265).

Nagel briefly considers the position of Woodfield (1976), an adherent of the intentional view, who argues that while our "core concept" of goal-directedness involves wanting to achieve a goal, we also have a "broader concept" that applies to any process produced by an inner state that is suitably analogous to wanting. Nagel argues that Woodfield's "analogical extension" story cannot make sense of what he regards as *bona fide* goal-directed processes in biology, such as a tadpole developing into a frog. He argues that there is no useful sense in which "the inner state representing the goal…which is perhaps a complex subsystem of genetic materials in the tadpole…resemble[s] (or is analogous to) a desire or a belief" (Nagel, 1977a: 266). Nagel seems on strong ground here: morphogenesis is not much like the deliberate pursuit of a goal. However, the broader implication of Woodfield's view – that goal attributions in biology involve an element of anthropomorphism – may still be defensible.

The program view is essentially the position of Mayr. Nagel opposes Mayr in part because he regards the notion of an "open program" as obscure, in part because he thinks Mayr's analysis misclassifies clear-cut cases, and in part because he doubts that it reflects how we actually make judgements of goal-directedness. Nagel argues that some programmed organismic processes, such as the knee reflex, intuitively do not count as goal-directed, so the program view is too liberal. He also argues that Mayr's attempt to exclude "automatic" processes (that Mayr calls "teleomatic") from the domain of the goal-directed does not work, since if applied consistently it would exclude phenomena that Mayr wishes to include, such as the workings of a clock. But Nagel's main objection to the program view is that we cannot observe the program that controls an organismic process, so our judgment about whether it is goal-directed, and, if so. what the goal is, must derive from another source. Since Nagel regards such judgements as reliable, he is led to seek a purely behavioral definition of goal-directedness.

From a modern perspective, this last criticism is uncompelling since it conflates the metaphysical question of what *makes* something an *x* with the epistemological question of how we can *tell* whether something is an *x*. The latter is at best at a defeasible guide to the former. Nagel may well be right that the usual way for us humans, given our cognitive limitations, to determine whether a process is goal-directed is to observe the process unfold, rather than to inspect the program (if any) that controls it. However, this is compatible with Mayr's claim that *what it is to be* a goal-directed process is for that process to be controlled by an internal program. Mayr's view may be untenable for other reasons, but Nagel's criticism should not convince.

The systems view, Nagel tells us, says that "being goal-directed is a property of a system, in virtue of the organization of its parts" (Nagel, 1977a: 273). The systems view is essentially the cybernetical view of Rosenbleuth et al. (1943), and Sommerhoff (1950). The latter, Nagel says, sought a general account of what makes a system goal-directed "irrespective of whether the goal is pursued by purposive human agents, by living systems incapable of having intentions, or by inanimate systems", an aim of which he approves (Nagel 1977a: 271-2).

Nagel says that the systems view correctly identifies two of the essential properties of goal-directedness: *plasticity* and *permanence*. Plasticity means the system can reach the goal via different pathways, and from multiple starting points. Persistence means that that the system maintains its goal-directed behavior despite external perturbations, thanks to internal compensatory adjustments that keep it on course; this is closely related to the negative feedback idea. Nagel illustrates plasticity and permanence with the example of the homeostatic regulation of the water content of blood in humans, a textbook example of a goal-directed process (the goal being to keep the water content at 90%). Following Sommerhoff (1950), Nagel adds to permanence and plasticity a somewhat obscure requirement that he calls "orthogonality". The point of this requirement is to exclude processes that are an upshot of simple physical law, such as a marble rolling down the side of a bowl coming to rest at is base, which Nagel thinks should not count as goal directed.

2.4 A natural kind?

Looking back at Mayr's and Nagel's discussions, one point stands out. Both authors take for granted that, aside from a few problem cases, they know which phenomena count as goal-directed and which not – these are the "data" against which their analyses are to be tested. That is, they assume that goal-directedness is an objective scientific property that defines a natural kind, not something projected onto nature by humans. These commitments are not idiosyncrasies of Mayr and Nagel, rather they are widely shared by participants in the traditional debate. But are they justified?

There are three reasons for doubt. Firstly, the terms "goal-directed" and "goal-directedness" are not technical terms in biology nor widely used in the general biological literature (though they are common in neuroscience and psychology).[4] There is a contrast here with the term "function" which *is* widely used in biology. Now this does not prove that goal-directedness is not a scientifically important phenomenon; as Nagel himself notes, not all goal-attributing statements necessarily use the word "goal" (Nagel, 1977a: 263). However, it does lead one to wonder whether Mayr and Nagel were right to take the reality and ubiquity of goal-directedness as their starting point.

Secondly and relatedly, the class of things that are supposed to count as goal-directed is rather heterogenous. The class includes: the entire process of development from embryo to adult; development sub-processes such as cleavage and gastrulation; physiological processes such as thermoregulation and tissue repair; organismic activities such as foraging and migration; conscious human behaviors such as writing an article; and the operation of man-made artifacts such as engines and thermometers. (Perhaps it is unsurprising that science has no single term covering all these.) The supposedly unifying feature is that in each case there is an "endpoint" which is reliably reached despite perturbations. No doubt there is some truth to this. Certainly, these processes involve striking regularities that cry out for explanation. However, the explanation of how a tadpole develops into a frog, how an injured salamander regenerates its limbs, and how a salmon reaches its natal stream are all quite different. Understanding how one of these processes works tells us nothing about how the others work. It is not obvious that we have a natural kind here, rather than a class of processes at most superficially alike.

Thirdly, the idea that goal-directedness is something we project onto nature is consonant with findings in experimental psychology. Many studies suggest that humans have an inbuilt cognitive bias that leads us, from a young age, to see intention and purpose where there is none, to anthropomorphise, and to favour teleological over mechanistic descriptions of the world (Barrett, 2011). (This bias may have evolved because it conferred a survival advantage.) If this bias is real and universal, it is conceivable that it may influence scientists and philosophers too.

This suggests that there may be a purely psychological explanation for why the processes that theorists regard as goal-directed have been co-classified. Perhaps all these processes elicit a certain psychological reaction in us. They strike us as akin to what we would see if the system were consciously aiming at a goal (in the case of whole-organism activities and behaviors) or had been designed by an agent with a goal in mind (in the case of internal processes). On this view, what unites a firefly's mating display, a tadpole's metamorphosis, and an immune system's producing T cells is not any similarity of actual behavior or underlying causal mechanism; it is rather that in each case the system behaves *as if* it were goal-directed in the sense of having, or stemming from, a conscious mental representation of a goal. And if that is so, then goal-directedness is something that we project into nature rather than discover in it.

This is somewhat similar to what Nagel calls the intentional view; but there is a difference. The intentional view as Nagel characterizes it is based on the idea that "real" goal-directedness requires intentional causation, so other cases belong only if their etiology is analogous or isomorphic. What I am suggesting is something different (though compatible), namely that the disparate class of processes that theorists have treated as goal-directed may be unified only in that they all provoke a certain psychological reaction in us, that leads us to anthropomorphically assimilate the endpoint of the process to the intentional object of an agent. This is not a claim about isomorphism between goal-directedness processes and the conscious pursuit of goals, but rather about the unity (or lack of it) among the processes that have been pre-theoretically classed as goal-directed.

2.5 Upshot

Where does this leave us? There seem to me to be two options. Either something *like* Mayr's program view can be made to work, or we should jettison the idea that goal-directed processes form a natural kind in favour of the projectivist alternative sketched above.

I argue this for two reasons. Firstly, attempts to make the systems view work have arguably failed (Garson, 2018). Despite much ingenuity, no version of the systems view has provided a definition of goal-directedness that can simultaneously rule out pseudo-cases such as the marble rolling down the bowl; can uniquely identify "the" goal towards which a process is directed; can explain how a process can be goal-directed and yet the goal not be achieved or the goal object not exist (as when a salmon's natal stream has dried up). (This last problem undermines the attempt to define goal-directedness in terms of feedback alone). The basic problem is simply that the goal towards which a process is directed, if any, is heavily underdetermined by the process's actual behavior. Appeal to hypothetical behavior may perhaps help; but philosophical experience with similar situations shows that it is preferable to simply abandon the behaviorist pretence and appeal directly to internal factors, as Mayr does.

Secondly, seeking unity among goal-directed processes at the level of observable behavior, rather than internal factors, sits uneasily with the inclusion of purposive human behaviors in the class of the goal-directed, which most theorists have agreed with. Features such as persistence and plasticity, on which systems theorists focus, hardly characterize all human behaviors. Some humans behave in a rather erratic way, fleetingly pursuing a goal and then doing something else (so not persisting); when a given action fails to achieve a goal, humans sometimes choose an alternative means to the goal (thus exhibiting plasticity), but sometimes they change goal or give up entirely. It seems that the real reason for including purposive behaviors in the class of the goal-directed is that they derive from a conscious mental representation of the goal, not that the behaviors themselves are particularly similar to other standard examples of goal-directed processes such as thermoregulation, for example.

If Mayr's program view were right, then goal-directed processes would indeed form a natural kind, despite their apparent heterogeneity. However, the program view sits uneasily with modern biological knowledge and is at best a metaphor. Certainly, genes play a crucial causal role in all of the processes – ontogenetic, physiological and behavioral – that Mayr regards as goal-directed, but the idea that the genome is a program controlling these processes is highly doubtful (Newman, 2022). Developmental genetics teaches us that that the genome is a highly *reactive* entity, not a fixed repository of instructions, and that gene expression often depends crucially on environmental triggers and conditions (Gilbert, 2003; Fox-Keller, 2014). Nor does Mayr's attempt to liberalise his view by allowing that programs can be "open" help much. For the program is still supposed to initiate the goal-directed process, which is hard to square with the context-sensitivity of gene action. In short, the program view rests on an undermotivated computer science analogy and an *a priori* privileging of genetic over environmental causes.

However, something in the spirit of Mayr's view may be salvageable even if we jettison the notion of evolved program. We may be able to retain the idea that the common feature of goal-directed processes is that they arise (in part) from a system having an inner state that *represents* the goal-state (or endpoint); and that this plays an essential role in explaining how such processes works. Now what exactly this means is a difficult question. But we can say the following. Representation does not mean conscious mental representation (though it includes this as a special case). It is a commonplace of contemporary cognitive science that organisms and their subsystems contain "sub-personal representations" of both internal and external states, which could in principle include the goal-states of teleonomic processes. Philosophers have made considerable progress with articulating this notion of representation and with showing how representing-involving explanations work (Shea, 2018).

There is some reason to think that inner representations can help understand goal-directedness. One way for a system to produce goal-directed behavior is to have an inner state representing the goal, to compare this inner state with its actual state, then to suitably alter the latter. This is how some (though not all) goal-directed animal behaviors work. Now whether this inner representation story covers all cases of putative goal-directedness in biology, including developmental and physiological processes, is not clear. For the story to work in full generality, the inner states doing the representing would presumably have to include genomic states (as per Nagel's suggestion that an inner state of the tadpole's genome represents the adult frog's form). Now the idea that genes "represent" phenotypic outcomes has been defended before; but it is more controversial than the idea that state of a cognitive system represent the world.[5] So, if the defender of the reality of goal-directedness wishes to go the inner representation route, some work would be needed to show that it picks out the desired class of phenomena.

Be that as it may, the inner representation story seems preferable to Mayr's program view for two reasons. Firstly, the notion of representation is arguably in better standing than that of internal program and is needed in some of areas of biology anyway. Secondly, it more easily explains why purposive human behaviors count as goal-directed, given that conscious mental representations are a sub-type of representation. By contrast, Mayr cannot easily explain this.

To conclude: the assumption that goal-directedness is an objective feature of the world is not obviously true. The assumption seems defensible only if goal-directed processes share a commonality at the level of internal mechanism, rather than observable behavior. The most promising such candidate is the existence of an inner representation of the goal-state that is causally implicated in producing the process. Should such a commonality turn out not to exist, or to pick out the "wrong" class of processes, we should conclude that the disparate phenomena traditionally treated as goal-directed do not form a natural kind.

3. Goal Attributions in Cognitive Ethology

A quite different debate about goal-directedness occurs in the fields of ethology and comparative cognition. At issue is the correct explanation of certain complex animal behaviors. The starting point is the apparent contrast between the instinctive behaviors that are common throughout the living world and the more sophisticated behaviors found in some vertebrate taxa. Thus contrast an insect flying towards light with a rat navigating its way out of an intricate maze. While the former is simply a hardwired instinct, the latter seems much more of a cognitive achievement, requiring memory, learning and inference. Impressed by this intuitive contrast, some researchers propose that certain non-human animals are capable of goal-directed actions that stem from internal mental representations, such as belief-like and desire-like states, pointing to intriguing experimental findings that seem to show this (Ristau, 1991; Dickinson, 2001; Clayton, Emery & Dickinson, 2006). These researchers thus posit continuity between the intentional behavior of humans and non-human animals, rejecting the suggestion that this is anthropomorphic.

Importantly, the concept of goal-directed behavior at work in this debate is not the same as that at work in the debate examined above. For note that many of the instinctive behaviors (such as insect phototaxis) that are supposed to *contrast* with goal-directed behavior in the ethological debate would be classified as goal-directed by Mayr, Nagel and others in the first debate. Rather, goal-directed behavior in the ethological discussion corresponds to Mayr's category of purposive behavior, that is, behavior that requires explanation in intentional-psychological terms, or equivalently that stems from belief-like and desire-like mental representations.

Whether non-humans are capable of goal-directed behavior in this intentional sense, and if so how widespread it is, is controversial. This reflects disagreement both about how to interpret the empirical data and how to define the relevant concepts. At one extreme is the view that only humans exhibit true goal-directed actions since non-humans lack the necessary cognitive requirements. Thus, for example, Davidson (1982) has argued, on essentially a priori grounds, that language is a prerequisite for having beliefs, desires and other intentional mental states; and, thus, that the behavior of non-human animals cannot stem from their being in such states. This view is not popular among scientists of animal behavior, though the position of Kennedy (1992), who regards all attribution of conscious mental states to non-humans as rooted in anthropomorphism and unwarranted by the data, comes close.

At the other extreme are those researchers who regard goal-directed behavior, or its cognitive preconditions, as pervasive in the living world, even extending beyond the animal kingdom. Thus, Trewavas (2014) argues that plants are intelligent, insisting that he is speaking literally; while Bray (2009) suggests that even single cells can be credited with "knowledge" of their environment. What underpins such arguments is the idea that wherever organisms exhibit adaptive plasticity and/or learning, hence can vary their behavior in response to the environment, it is legitimate to attribute to them (a rudimentary form of) cognition. Such arguments should not be dismissed out of hand, though whether the cognitive states in question should really be thought of as mental representations is debatable. We should also note that much organismic behavior is only plastic within narrow bounds; ingenious experimental interventions can often make apparently intelligent behavior seem rather dumb.

In between these extremes, one finds a spectrum of positions that allow that some complex behavior of animals with nervous systems is goal-directed in the intentional sense. Thus, for example, Ristau (1991) has studied the behavior of piping plovers that feign a wing injury when a predator approaches in order to lead them away from its young. The plover's broken wing display is highly sensitive to the predator's position, location and movement. Ristau argues that the best explanation of the plover's behavior is that it wants to lead the intruder away from its young; only this accounts for the precision and timing of its actions. Similarly, Clayton, Emery & Dickinson (2006) study the food-caching behavior of scrub jays. The jays not only store and retrieve food, but also use strategies to reduce the chance that their food is pilfered, such as delaying caching if other birds are watching, and choosing locations that are concealed from others' view. Clayton et al., insist that the jays' behavior should be explained by attributing to them beliefs, desires and memories, arguing that alternative non-intentional explanations fail. Other behaviors that have been thought to require intentional explanation include navigation, tool use, and future planning.

These middle positions prompt the question of whether precise behavioral criteria for goal-directedness can be laid down. This question is addressed by Dickinson (2001), who specifies two criteria for an action to count as goal directed as opposed to "habitual". His *goal* *criterion* says that a goal-directed action must be "sensitive to whether or not [its] outcome is currently a goal for the animal" (Dickinson, 2001:80). Dickinson illustrates this with the example of a rat pressing a lever to receive an outcome which has previously been devalued by conditioning. If the rat presses the level anyway, despite the devaluing of the reward, its behavior is purely habitual, Dickinson argues. His "instrumental criterion" says that the animal's action should be sensitive to the causal relationship between action and reward. Thus, if a food reward that is usually contingent on one action is suddenly made contingent on a different action by the experimenter, the animal must be capable of learning this for its action to count as goal directed. In effect, this is to say that goal-directed action requires that an animal has "causal knowledge", or more precisely a mental representation of the causal dependence of outcome on action.

Dickinson's criteria are quite strict (though he argues that rats meet them), making goal-directed action fairly uncommon. His instrumental criterion, in particular, seems overly demanding, as a number of authors have argued (Carruthers, 2004; Allen & Bekoff, 1995). Though it is quite plausible that learning of some sort is a requirement for having intentional states, and thus for goal-directed action, requiring that an animal be able to learn the pattern of causal contingency of reward on action seems too strong. The capacity for such causal learning is plausibly needed for complex means-end reasoning of the form "if I were to do *x*, I would get *y*", but this goes beyond merely acting from belief-like and desire-like mental representations; thus, the instrumental criterion is too restrictive. A more plausible, though admittedly vaguer, criterion is that animal behavior counts as goal-directed when it is sufficiently flexible, intelligent-seeming and complex that no non-intentional explanation is feasible. Experimental intervention is generally necessary to probe this.

Two further positions on the issue merit brief mention. The first is Dennett's idea that there is no sharp distinction between genuine and "as if" intentionality anyway. On Dennett's view, it is a mistake to ask whether a particular behavior, human or non-human, really stems from belief-like and desire-like states; the only question is whether it is heuristically useful to study the behavior from the "intentional stance". Dennett (1983) argues, perhaps somewhat optimistically, that adopting the intentional stance is often helpful for cognitive ethologists, as it leads to interesting hypotheses that can be tested. The second is the position of the prominent 20th century ethologist D. McFarland (1989a,b) who argues, radically, that not even human behavior is genuinely goal-directed; our belief in the goal-directedness of our own and others' actions is a delusion that has been programmed into us by natural selection. McFarland's reason for thinking this appears to be that he does not believe in mental representation at all, as he does not see a way of squaring our folk-psychological talk of beliefs and desires with the underlying facts of neuroscience. In philosophy, a similar position has been advocated by P. Churchland (1981) under the label "eliminative materialism".

This debate raises complex issues, philosophical and scientific, that I cannot hope to resolve here, so I will confine myself to a number of points. Firstly, *if* we accept that much human behavior is caused by inner mental representations, there seems every reason to allow that the same may be true of some non-humans, both on grounds of evolutionary continuity and known neurophysiological similarity (Glimcher, 2003). A priori arguments to the contrary should carry little weight. So, there is no reason to believe that all attributions of goal-directed behavior to non-humans are the result of anthropomorphic projection.

Secondly, the accusation of anthropomorphism is probably justified in some cases. For it is well-established that complex adaptive behavior may arise from mechanisms that bear no resemblance to belief-like and desire-like inner representations. Barrett (2011) argues that simple internal mechanisms can often produce complex behavior by taking advantage of environmental regularities. Barrett gives the example of predatory *Portia* spiders, which show a remarkable ability to detour around obstacles while hunting for prey. The spiders' behavior conveys the impression of advance planning, for they need to let the prey out of their sight to get around an obstacle, and they appear to carefully scan the terrain before starting a detour. However, experiments reveal that no planning is going on; the spider is using a simple rule-of-thumb, based on the presence or absence of horizontal lines in its field of vision, to determine which direction to move in at each moment; this leads it to trace out the most direct route towards its prey. So, if someone were to explain the spider's behavior in belief-desire terms this *would* be anthropomorphic; and, moreover, it would detract from rather than conduce towards a correct scientific understanding of the behavior.

Thirdly, it seems unrealistic to hope for fully explicit behavioral or experimental criteria for when attributions of goal-directedness are justified and/or of heuristic value. Some key variables are clear, including the degree of behavioral flexibility, the capacity for novelty, and the ability to learn. But the link between mental representations and observed behavior is too indirect to expect fully explicit criteria; an element of judgement will always be needed.

Fourthly, Dennett's idea that no hard-and-fast line separates "real" from "as if" intentionality, though out of fashion in contemporary philosophy, may well be true, and fits with the fact that most biological attributes come in degrees. But this is still compatible with some explanations of behavior being clearly anthropomorphic. That a distinction is not hard-and-fast, so admits of borderline cases, is compatible with the existence of clearcut cases on either side. Finally, even if Dennett is wrong and there is an objective distinction between behavior that is genuinely intentional and behavior that to all intents and purposes appears as if it were, this distinction is (by construction) impossible to operationalize. Empirical work on goal-directedness in animals is thus insensitive to the distinction and cannot resolve the question of whether it exists. This latter question is inherently a philosophical one.

4. Goal Attributions in Evolutionary Biology

The final context in which goal attributions arise is evolutionary biology. A well-known project in evolutionary biology seeks to explain an organism's evolved traits in terms of their adaptive significance (or function). Such explanations quite often treat the organism as if it were an agent with a goal. The organism's overall goal is often said to be survival and reproduction (or maximizing its fitness); to achieve this goal it needs to pursue intermediate goals such as finding food, attracting mates and raising its young, to which its various evolved traits, including its behaviors, make distinct contributions. Thus, for example, Roff (1992) writes: "the primary goal of any organism is to reproduce...the first "decision" it must make...is when to start reproducing" (2); West and Gardner (2013) describe maximizing its inclusive fitness as the "objective" (goal) of an organism's social behavior; while Grafen (2007) describes an evolved organism's phenotype as an "instrument" which it wields "in pursuit of a maximand", i.e., the quantity that the organism is trying to maximize.

The idea of an organism as pursuing a goal towards which its evolved traits conduce may seem innocuous, at least on the assumption that those traits are broadly adaptive. But it becomes philosophically interesting when the traits in question are evolved behaviors. For then, the idea often assumes a particular form, in which the behavior's function is treated as if it were the organism’s goal, and the evolutionary explanation is re-cast in an intentional idiom. Why do swallows migrate? Because they want to escape the cold. Why do female rats kill their offspring? Because they know that the offspring will not survive and don't want to waste resources on them. Why do worker honeybees eat the eggs laid by fellow workers? Because they want the offspring of the queen to be reared instead. In this way, the language of instrumental rationality (as philosophers call it) is used to describe and theorize about evolved behavior: the organism is treated as an agent who acts for reasons, makes decisions, and pursues goals.

Note that this evolutionary use of intentional-psychological language arises in the context of giving ultimate explanations. As such, it raises quite different issues from those discussed in the ethological literature above, where the focus was on proximate explanation. A honey-bee’s nervous system is too simple, and its behavior too rigid, for a proximate explanation of its egg-eating behavior in intentional terms to be plausible. The bee does not really *want* anything; it is simply obeying its hormonal impulses. But this fact does not prevent us from construing the evolutionary explanation of its behavior, metaphorically, in terms of what the bee wants, i.e., the goal that it is trying to achieve. We just need to be clear that “goal” in this context refers to the behavior’s evolutionary function not its proximate cause.

Expressing evolutionary explanations in this psychological fashion is fairly common and has been explicitly defended by Dennett (1987) and Dawkins (1976). But one might reasonably wonder what its point is, given that, unless used carefully, it invites a confusion of ultimate with proximate explanation (Scott-Phillips, 2011). Moreover, given that evolutionary biology in any case needs the notion of function (in the sense of adaptive significance), what if anything is to be gained by treating a trait’s function as if were the organism’s goal and then introducing psychological descriptors? Does this add anything important?

I think that it does, for the following key reason. Functional talk applies to traits, but intentional-psychological talk applies to the whole organism. A particular trait has a function; but it is the whole organism, not its traits, that pursues a goal, or prefers one thing to another, or adjusts its behavior to achieve a certain end. Thus, when intentional language is used in an evolutionary context, the subject of the intentional attribution is the whole organism, not one of its traits. The meerkat's warning behavior has a function, but it is the meerkat that sees the danger and wants to warn its companions.

Why does this matter? Because it highlights an implicit theoretical commitment of the “organism as agent with goal” idea, namely that the organism exhibits what may be called a *unity of purpose*. This means that its different traits have evolved because of their contributions to a *single* overall goal: enhancing the organism's fitness (or perhaps its inclusive fitness). Where this unity does not obtain, the organism cannot be regarded as agent-like, and treating its behavior as a means by which it furthers its goal will impede, not facilitate, evolutionary understanding of its behavior. I develop this theme below.[6]

4.1 Unity of purpose

Consider unity of purpose in the human context first, where it is a fundamental aspect of human agency.[7] This unity has two components. Firstly, a person's goals should cohere with each another in the sense of being mutually reinforcing, or at least not clearly inconsistent; secondly, their actions should tend to further their goals, i.e., they should be instrumentally rational. Minor deviations from this unitary ideal are common, but if they are too many, or too great, it becomes impossible to treat the person as a unified agent, and to rationalize their actions in terms of their goals. Indeed, if a person is sufficiently disunified, psychological descriptors lose their grip entirely: we cannot say what they believe or are trying to achieve. In the biological case, an analogous unity of purpose is necessary in order to sensibly treat an evolved organism as akin to an agent with a goal and is presupposed when intentional-psychological idioms ("wants", "tries") are applied to the organism in an evolutionary context, in the manner described above. By contrast the functional idiom, since it applies on a trait-by-trait basis, involves no such presupposition.

To illustrate this point, let us consider three cases where the required unity-of-purpose partly breaks down, two actual and one hypothetical. All three involve within-organism conflict. In *Drosophila pseudoobscura*, males that carry a particular X-chromosome variant produce no Y-bearing sperm at all, as a result of "sperm killer" genes on the X chromosome which disrupt spermatogenesis. As a result, far fewer viable sperm are produced than in normal males. This trait – failure to produce Y-bearing sperm – evolved not because it benefits the organism, which it does not, but rather because it benefits the X-chromosome itself (and the genes on it). So, the fly exhibits a partial disunity of purpose. Most of its traits, e.g., its mating behavior, pull in the direction of maximizing its reproductive success, but the trait of producing no Y-bearing sperm pulls in a completely different direction. If a biologist studying fly spermatogenesis treats the fly as a fully unified agent, they will not understand what they see.

Examples of this sort could easily be multiplied, since a certain amount of intra-genomic conflict is found in many species (Burt & Trivers, 2006). For the most part, though, organisms have evolved mechanisms to suppress such internal conflict, and thus to ensure that all their constituent genes work for the common good. It is precisely because this suppression is so effective that we are usually able to treat organisms as agents pursuing goals, and to describe their evolved behavior using the language of instrumental rationality.

Our second example involves parasitic manipulation. Consider an ant infected by the liver fluke parasite *Dicrocoelium dendriticum*. This parasite induces a change in the ant's behavior, causing it to climb to the top of a blade of grass every evening and stay there, clamped to the tip with its mandible. This increases the chance that the ant will be ingested by a sheep, which is what the parasite needs to complete its life cycle. An infected ant thus exhibits a partial disunity-of-purpose. Most of its traits, such as its foraging behavior, further its goal of survival and reproduction; but its nightly ascent of a blade of grass detracts from that goal. It is as if the ant is simultaneously pursuing two incompatible goals. If a biologist tries to treat the ant as a fully unified agent, they will not understand what they see.

These two failures of biological unity-of-purpose are the analogue of a human agent whose goals conflict. Our third, hypothetical example is the analogue of a human agent performing an action that detracts from, rather than furthers, one of its goals. Imagine a mouse gene, expressed in females, that causes a female mouse to kill any pups in its litter which do not contain a copy of that gene. (This is not as far-fetched as it sounds.[8]) Such a gene could easily spread by natural selection (though genes at unlinked loci in the mouse genome would be selected to suppress it). If the gene does spread, then although an evolutionary explanation of the mouse's infanticidal behavior could be given, it could not be couched in terms of what the mouse "wants" or is "trying" to achieve. For the behavior detracts from, rather than furthers, the mouse's goal of leaving surviving offspring. The mouse thus lacks (the second component of) unity of purpose.

The general moral is this. To treat an evolved organism as agent-like requires that we can treat the organism's various traits as instruments for achieving sub-goals – finding food, keeping warm, producing gametes, mating – which contribute to a single overarching goal, namely enhancing the organism's fitness. Empirically, this requires that the genes coding for the traits have identical evolutionary interests, so that the traits evolve functions which are complementary rather than antagonistic. This in turn requires that intra-genomic conflict and parasitic manipulation are largely absent. For otherwise, then although each trait considered individually can be given a functional explanation, the traits cannot be treated as contributions to a single overarching goal.

The contrast between intentional attributions to organisms and functional attributions to traits is reminiscent of a traditional contrast in the philosophy of psychology, between personal and sub-personal attributions. Folk-psychological notions, such as believing and desiring, are personal-level: it is whole persons that occupy these states. By contrast, the computational processes described by cognitive psychology are sub-personal; they are carried out not by persons but by parts of their brains. Thus, the cerebral cortex processes visual information, but the person sees the approaching car and moves out of the way. Essentially, we have here a biological analogue of this distinction. It is the parts of an organism, i.e., its traits, that have Darwinian functions, but it is the whole organism that has aims, goals, and preferences. Moving from the former sort of attribution to the latter is only possible in so far as the organism exhibits a biological unity of purpose.

This point is a corollary of a widely accepted evolutionary principle, namely that internal conflict tends to undermine the integrity of a larger unit. Thus, multi-celled organisms have evolved numerous mechanisms for suppressing conflict among their constituent genes and cells, including fair meiosis, uniparental inheritance of organelles, and programmed cell death (Frank, 2003; Bourke, 2011). It is because these mechanisms usually work well that organisms are as cohesive and integrated as they are. This biological principle is an empirical one, but it has a conceptual counterpart. It is only because an organism’s constituent traits typically cohere with each other in this way that the organism can be treated as akin to an agent pursuing a goal.

Since the unity-of-purpose requirement is not always satisfied, this might be regarded as a limitation of expressing evolutionary explanations in terms of organisms pursuing goals. In a way this is so, but it also shows that this practice has a genuine rationale and is not idle metaphor. For most of the time, the requisite organismic unity does obtain, at least to a high degree of approximation. There is thus a real pattern in nature that is captured by treating the organism as if it were an agent pursuing a goal, which the functional idiom alone does not capture. Therein lies the heuristic value of psychologizing evolutionary explanations in this way, a practice that at first blush may seem unmotivated.

This point should be sharply distinguished from a suggestion of Trivers (2009), which postulates a causal link between intra-genomic conflict and internal conflict in the human psyche. Trivers focuses on genomic imprinting, in which a gene has different phenotypic effects depending on whether it is paternally or maternally inherited. This leads to intra-genomic conflict; for, if a gene in an organism is paternally inherited, then it has no genetic interest in the future reproduction of the organism's mother, while genes at other loci do. Trivers suggests that this will have psychological consequences: ``we literally have a paternal self and a maternal self, and they are often in conflict" (Trivers, 2009: 163). Haig (2006) argues similarly.

The Trivers/Haig hypothesis is interesting though speculative. I take no stand on the matter here. I do not claim a direct connection between biological unity of purpose and unity of purpose in human agents (or their absence). Rather, my point is that the attribution of goals and intentions to a subject only makes sense if the subject exhibits sufficient unity, i.e., is an undivided self, or close enough. So, to treat an evolved organism as akin to an agent trying to achieve a goal, for the purposes of evolutionary theorizing, requires that the organism's traits do not have mutually antagonistic functions; and empirically, this requires the absence of intra-genomic conflict or nearly enough. This is a claim about the presuppositions of a particular psychologically derived idiom that we apply, usually metaphorically, to evolved organisms; not a claim about the evolutionary roots of psychological unity or disunity in humans.

5. Conclusion

We have examined three different biological contexts in which goal attributions arise. The first is the mid-twentieth debate over goal-directedness among both biologists and philosophers, where the driving concern was to better understand teleology (or the appearance of it). The second is the ongoing debate among cognitive ethologists about the presence or otherwise of goal-directed behavior, in the intentional sense, in non-human animals. The third is the evolutionary biologist’s penchant for treating organisms as if they were rational agents pursuing a goal, in the context of seeking evolutionary explanations for their behavior.

The three contexts are distinct, but there are interrelations between them. The first and second are related since the sort of goal-directed behavior that the cognitive ethologists are concerned with (Mayr’s “purposive behavior”) was usually treated as a special case of goal-directedness by participants in the first debate. The second and third are connected since the intentional-psychological descriptors that are used in an as-if sense in the third, evolutionary context, are exactly those whose literal applicability is at stake in the second, ethological context. Finally, the first and third are related since one theme in the first debate was the need to distinguish something’s having a goal from its having a function; whereas in the third context, the mode of evolutionary explanation in question precisely involves treating a behavior’s function as if it were the organism’s goal.

Moreover, there is a thematic question that runs through all three contexts, captured in the sub-title of this paper: objective fact, anthropomorphic bias, or valuable heuristic? In the first context, the underlying issue is whether there really is an objective property of goal-directedness in the first place – does “goal-directed process” pick out a natural kind? We expressed scepticism on this score, because of the heterogeneity of the class and the unclarity regarding its membership; and we suggested that something like Mayr’s “program view” would need to be defensible if the objectivity of goal-directedness was to be sustained. In the second context, the issue is whether non-human animals are really capable of (intentional) goal-directed behavior, or whether this is just anthropormorphism. We tentatively suggested that goal-directed behavior is likely a reality in some non-humans, but that the anthropomorphism accusation is not without basis in other cases, since complex adaptive behavior can be produced by mechanisms that have no resemblance to belief-like and desire-like inner states. In the third context, the issue is whether psychologizing evolutionary explanations by treating behavior-functions as organismic goals has any real point to it. We argued that it does, since the unity of purpose that this practice presupposes reflects a real and biologically important fact about organisms, namely that their evolved traits (for the most part) are designed to achieve a single overall goal, namely maximization of fitness.

These answers are provisional; they not intended as the final word on how we should understand goal attributions in the three contexts. But I hope that our discussion has shown that goals and goal-directedness in biology are still topics worthy of philosophers' and biologists' attention.

References

Allen C. and Bekoff, M. 1995. “Cognitive ethology and the intentionality of animal behavior.” *Mind and Language* 10(4): 313-328.

Barrett. L. 2011. *Beyond the Brain*. Princeton NJ: Princeton University Press.

Beckner, M. 1969. “Function and teleology.” *Journal of the History of Biology* 2(1): 151-64.

Bourke, A.F.G. 2011. *Principles of Social Evolution*. Oxford: Oxford University Press.

Braithwaite, R.B. 1953. *Scientific Explanation*. Cambridge: Cambridge University Press.

Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bray, D. 2009. *Wetware: A Computer in Every Living Cell*. New Haven,, CT: Yale University Press.

Burt, A. and R. Trivers. 2006. *Genes in Conflict*. Cambridge, MA: Harvard University Press.

Carruthers, P. 2004. “On being simple minded.” *American Philosophical Quarterly* 41(3): 205-220.

Churchland, P. 1981. “Eliminative materialism and the propositional attitudes.” *Journal of Philosophy* 78(2): 67-90.

Clayton, N., Emery, E., and Dickinson, A. 2006. “The rationality of animal memory.” In *Rational Animals*, edited by S. Hurley and M. Nudds, 197-216. Oxford: Oxford University Press.

Corning, P A. 2022. “Teleonomy in evolution: The ghost in the machine”. In *Evolution on Purpose: Teleonomy in Living Systems*, edited by P. Corning and R. Vane-Wright. Cambridge: Cambridge University Press.

Davidson, D. 1982. “Rational Animals.” *Dialectica* 36: 318-27.

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

Dennett, D.C. 1983. “Intentional systems in cognitive ethology.” *Behavioral and Brain Sciences* 6: 343-355.

Dennett, D.C. 1987. *The Intentional Stance*. Cambridge MA: MIT Press.

Dickinson, A. 2011. “Goal-directed behavior and future planning in animals”. In *Animal Thinking: Contemporary Issues*, edited by R. Menzel and J. Fischer, 79-91, Cambridge, MA: MIT Press.

Fox-Keller, E. 2014. “From gene action to reactive genomes.” *Journal of Physiology* 592(11): 2423-2429.

Frank, S.A. 2003. “Repression of competition and the evolution of cooperation.” *Evolution* 57: 693-705.

Garson, J. 2008. “Function and teleology.” In *A Companion to the Philosophy of Biology,* edited by A. Plutynski and S. Sarkar, 525-549. Malden, MA: Blackwell.

Garson, J. 2018. “Nature and normativity: biology, teleology and meaning”. *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/reviews/ature-and-normativity-biology-teleology-and-meaning/

Gilbert, S.F. 2003. “The reactive genome”. In *Origination Of Organismal Form: Beyond The Gene In Developmental And Evolutionary Biology*, edited by G.B. Müller and S.A. Newman, 87-101, Cambridge, MA: MIT Press.

Glimcher, P. 1983. *Decisions, Uncertainty and the Brain*. Cambridge, MA: MIT Press.

Grafen, A. 2007. “The formal Darwinism project: a mid-term report.” *Journal of Evolutionary Biology* 20(4): 1243-54.

Haig, D. 2002. *Genomic Imprinting and Kinship*. New Brunswick, NJ: Rutgers University Press.

Haig, D. 2006. “Intrapersonal conflict”. In *Conflict*, edited by M.K. Jones, and A.C. Fabian, 8-22. Cambridge: Cambridge University Press.

Kant, I. 1790 (2000). *Critique of the Power of Judgment*. Edited by P. Guyer, translated by P. Guyer and E. Matthews, Cambridge: Cambridge University Press.

Kennedy, J. S. 1982. *The New Anthropomorphism*. Cambridge: Cambridge University Press.

Kennett, J. and S. Matthews. 2003. “The unity and disunity of agency.” *Philosophy, Psychiatry and Psychology* 10: 308-312.

Korsgaard, C.M. 1989. “Personal identity and the unity of agency.” *Philosophy and Public Affairs* 18: 103-131.

Maynard Smith, J. 2000. “The concept of information in biology.” *Philosophy of Science* 67 (2): 177-194.

Mayr, E. 1974. “Teleological and teleonomic. a new analysis.” *Boston Studies in the Philosophy of Science* 14: 91-117.

Mayr, E. 1988. *Toward a New Philosophy of Biology*. Cambridge, MA: Harvard University Press.

Mayr, E. 1992. “The idea of teleology.” *Journal of Historical Ideas* 53: 117-535.

McFarland, D.J. 1989a. *Problems of Animal Behavior*. London: Longmans.

McFarland, D.J. 1989b. “Goals, no-goals and own goals”. In *Goals, No-Goals and Own Goals,* edited by A. Montefiore and D. Noble, 39-57. London: Unwin Hyman.

McShea, D.W. 2012. “Upper-directed systems: a new approach to teleology in biology.” *Biology and Philosophy* 27: 663-684.

Monod, J. 1973. *Chance and Necessity*. New York, NJ: Vintage Books.

Nagel, E. 1977a. “Teleology revisited: goal-directed process in biology.” *Journal of Philosophy* 74(5): 261-279.

Nagel, E. 1977b. “Functional explanations in biology.” *Journal of Philosophy* 74(5): 280-301.

Newman, S. 2022. “Self-organization in embryonic development: myth and reality”. In *Self-Organization as a New Paradigm in Evolutionary Biology: From Theory to Applied Cases in the Tree of Life,* edited by A.D. Malassé, Springer.

Okasha, S. 2018. *Agents and Goals in Evolution*. Oxford: Oxford University Press.

Pittendrigh, C.S. 1958. “Adaptation, natural selection and behavior”. In *Behavior and Evolution*, edited by A. Rose and G.G. Simpson, 390-416. New Haven, CT: Yale University Press.

Ristau, C. 1991. “Aspects of the cognitive ethology of an injury-feigning bird, the piping plover. In Ristau, C., editor, Cognitive Ethology, edited by C. Ristau, 93-124. Hillsdale,, NJ: Lawrence Erlbaum Associates.

Roff, D.A. 1992. *Evolution of Life Histories*. New York, NY: Chapman and Hall.

Rosenblueth, A., N. Wiener, and J. Bigelow. 1943. “Behavior, purpose and teleology.” *Philosophy of Science* 10: 18-24.

Rovane, C. 1998. *The Bounds of Agency*. Princeton: Princeton University Press.

Scott-Phillips, T.C., T.E. Dickins and S.A. West. 2011. “Evolutionary theory and the ultimate/proximate distinction in the human behavioral sciences.” *Perspectives on Psychological Science* 6: 38-47.

Shea, N. 2013. “Inherited representations are read in development.” *British Journal for the Philosophy of Science* 64 (1):1-31.

Shea, N. 2018. *Representation in Cognitive Science*. Oxford: Oxford University Press.

Sommerhoff, G. 1950. *Analytical Biology*. Oxford: Oxford University Press.

Trewavas, A. 2015. *Plant Behavior and Intelligence*. Oxford: Oxford University Press.

Trivers, R. 2009. “Genetic conflict within the individual.” *Sonderdruck der Berliner-Brandenburgische Akademie der Wissenschschaften* 14, 149-199, Berlin: Akademie Verlag.

Waddington, C.H. 1957. *The Strategy of the Genes*. London: Ruskin House.

West, S.A. and A. Gardner. 2013. “Adaptation and inclusive fitness.” *Current Biology* 23: R577-R584.

Williams, G.C. 1966. *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.

Woodfield, A. 1976. *Teleology*. Cambridge: Cambridge University Press.

Wright, L. 1973. “Functions.” *Philosophical Review* 82: 139-68.

Wright, L. 1976. *Teleological Explanations: An Etiological Analysis of Goals and Functions*. Berkeley, CA: University of California Press.

Notes

[1] A notable exception is McShea (2012).

[2] Mayr also introduces the category of a "somatic program", but defines it, confusingly, in essentially the same way as an open program.

[3] There are various hypotheses about why vertical migration is advantageous, of which predator avoidance is the leading contender. Maintenance of the oxygen supply is unlikely to be the reason.

[4] A Web of Science search for articles published between 1900 and 2021 with the terms "goal-directed" or "goal-directedness" in the title, abstract or keywords produces 14,105 results, of which only 198 are in biology journals. Neuroscience and psychology journals account for the largest share.

[5] Shea (2013) defends the idea that genes contain "representational content". Maynard Smith (2000) defends the related idea that genes contain "semantic information" about phenotypes.

[6] A fuller exposition of this line of argument is given in Okasha (2018) ch.1

[7] Unity of purpose is discussed by Kennett and Matthews (1993) and Okasha (2018, ch. 1). It is closely related to the "rational unity" discussed by Rovane (1998) and the "unity of agency" described by Korgsaard (1989). Bratman (1987) emphasizes that agents are rationally required to have consistent intentions, and to exhibit means-end coherence.

[8] David Haig's kinship theory of genomic imprinting has uncovered phenomena of exactly this sort (Haig 2002).