Chapter 2
Goals and Functions

Abstract Contemporary philosophical debates about biological function started in the early 1970s, and they originated from earlier, related, debates about the nature of goal directed systems. These discussions were rooted in scientific advances in the 1920s and 1930s pertaining to cybernetic machines and homeostatic systems, which appear to be purposeful or goal-directed despite not having any conscious intentions. By the 1950s, there were two major philosophical traditions for analyzing goal directedness, the behavioristic and the mechanistic. According to the behavioristic approach, favored by theorists like Gerd Sommerhoff and Richard Braithwaite, a goal directed system is one that exhibits plasticity and persistence in its outward behavior. According to the mechanistic tradition, favored by Ernest Nagel and Norbert Wiener, a goal directed system must be governed by the right sort of mechanism (such as negative feedback). Both of those traditions faced severe philosophical criticism in the 1960s and 1970s. I begin this chapter by sketching the historical background of the earlier debates about goal directedness. I then present the behavioristic analyses of Sommerhoff and Braithwaite, and enumerate several serious criticisms. I discuss mechanistic approaches, namely those of Nagel and the cyberneticists, and their critics.

Keywords Goal directedness · Naturalized teleology · Cybernetics · Homeostatic systems · Negative feedback

2.1 The Theory of Goal Directed Systems

A beaver builds a dam in order to surround itself with water. A frog orients its head to catch a buzzing fly. An anaerobic bacterium swims toward geomagnetic north to move away from oxygen-rich water. Some plants track the movement of the sun across the sky (heliotropic behavior). One of the most distinctive characteristics of living things, and perhaps even life's defining characteristic, is its apparent goal directedness. We explain why living things act the way they do in terms of certain states that those behaviors tend to bring about.

J. Garson, *A Critical Overview of Biological Functions*,
Philosophy of Science, DOI 10.1007/978-3-319-32020-5_2

Traditionally, this sort of explanation has been called a "teleological" explanation. A teleological explanation is one that purports to explain the existence of an entity (such as an organism, trait, or behavior), in terms of some effect the entity tends to bring about. Why do we have teeth? Because teeth help us chew food. Why is the cat crouching by the mouse hole? Because crouching by a mouse hole is a good way to catch mice. Terms like "purpose," "end," and "goal," are signals for teleological explanations. Organisms, their parts, and their behaviors, seem particularly apt for such teleological explanations. At the beginning of Sect. 3.1, I explain more carefully what teleological explanations are.

What do functions have to do with goals? Goals are different from functions. Something can have a goal without having a function. People and other organisms have goals, but they do not have functions. When we consider something as a self-contained entity—an uncontained container—we typically do not attribute a function to it, but we do attribute goals to it. By the same token, something can have a function without having any goals. An artifact such as a paperweight has a function, which is derived from the intention of the person who made it, or perhaps by the way it is currently used. But it does not have a goal.

If goals are different from functions, why talk about goals here? This is a volume on functions, not goals. There are three good reasons to talk about goals. First, although they are very different, there is an intimate relationship between them. Goals and functions both have a role in teleological explanations. To explain an organism's behavior in terms of a goal is to explain it in terms of some state it tends to promote. Similarly, according to one long tradition of thinking about functions (see Sect. 3.1), to attribute a function to an item is to explain that item in terms of some effect it creates. Goal-explanations and function-explanations are two species of the same genus, like apples and oranges are two kinds of fruit.

But the relationship is potentially tighter than that. This brings us to a second reason for talking about goals here. Some theorists have tried to *define* functions in terms of goals. On the surface, this is a tempting idea. When we say that the heart has the function of circulating blood, what more are we saying than that blood circulation promotes one of the organism's goals, namely to survive? And when we say that a paperweight has the function of holding down papers, what more are we saying than that holding down papers promotes the goals of the people who make and use it? Philosophers who developed thoughts along these lines include Nagel 1953, 1961, 1977; Wimsatt 1972; Boorse 1976; Adams 1979; Schaffner 1993. Recent theorists who have explored the notion of goal-directedness and its importance for functions include McShea (2012, 2013); Trestman (2012); Piccinini (2015, Chap. 6); Maley and Piccinini (Submitted for publication). Although these authors disagree about what exactly goals are, they all agree with the idea that functions are contributions to goals. (Interestingly, and as I will note in some detail in Sect. 4.1, Boorse's view is often described as a fitness-contribution view, but this is a mistake. Boorse accepted the basic account of goal-directedness provided by Nagel and Schaffner, where goals are defined in cybernetic terms. He just thought that, *in the context of physiology and biomedicine*, goals are simply tantamount to fitness-contributions.)

A third reason for talking about goals here is that the theory of goal directed systems provided an important historical backdrop for modern discussions of biological function. The notion of goal directedness represented the first sustained attempt by philosophers and scientists to make sense of teleological explanations in natural terms. That is, it represented an attempt to understand how scientists could legitimately use and apply terms like “goal,” “purpose,” and “end” to biological systems (and even machines) without assuming that those systems were created by God for special purposes, or that those animals, plants, and machines, had thoughts and intentions of their very own.

I will begin with the historical context, and then dive into the conceptual part. This is not meant to be a detailed historical exposition. Instead, it is an attempt to illuminate in very broad strokes the social and scientific context that motivated this thinking. Historically, philosophical discussion of goal directedness originated from two different traditions in early twentieth century science (though I suppose that historians can trace it back much further). The first tradition was organismic biology (see, e.g., Gilbert and Sarkar 2000 for discussion of this tradition). There were several biologists in the first half of the century that rejected the vitalism of people like Hans Driesch, but still insisted that there was some fundamental difference between living systems and purely physical systems. What is distinctive about living systems, in their view, is that they are goal directed and purposeful (Rignano 1931; Russell 1945; Bertalanffy 1950). So, they asked, what makes those systems purposeful?

The second tradition was cybernetics, which flourished during World War II and was devoted to the study and manufacture of “servomechanisms,” or life-like machines (Wiener 1948). A heat-seeking missile is an example. It pursues its object with a deadly accuracy that we are tempted to describe as “purposeful.” There was an important paper published in *Philosophy of Science* in 1943 that attempted to justify the idea that such machines are inherently purposeful. This was written by the American mathematician Norbert Wiener, the Mexican physiologist Arturo Rosenblueth (who had earlier worked with Walter Cannon), and the American engineer Julian Bigelow (Rosenblueth et al. 1943). They argued that the sciences required a concept of purposiveness that could apply to machines as well as organisms, and that did not appeal to conscious intentions or design. A heat-seeking missile, they argued, has the goal or purpose of striking its target. Moreover, this goal is intrinsic to it, in the sense that having that goal does not logically depend on the existence of human intentions. Philosophers sympathetic to this tradition quickly extended this notion of intrinsic purposiveness to describe certain biological systems as well, such as homeostatic mechanisms, e.g., the system in mammals that regulates the water content of the blood (e.g., Nagel 1953).

So, what do we mean when we say that a system is goal directed? Let’s approach this, first, from the standpoint of conceptual analysis of ordinary language. (Perhaps we will choose to depart from the confines of conceptual analysis later.) One natural idea is that *purposeful human behavior* is the best starting point for thinking about goals (Taylor 1950; Woodfield 1976). To say that a person has a goal is usually just to say that the person has a desire or an intention. To say I have the goal of visiting

France is just to say that is what I intend to do. When we say that animals have goals we are also attributing intentions to them. Perhaps their intentions are not as sophisticated as ours, and perhaps they are not expressed in the same sort of linguistic or syntactical structures as ours, but they are intentions nonetheless.

This sort of analysis, which takes as its starting point conscious human intentions, has an obvious limitation. It would not apply to machines such as homing torpedoes. Nor would it apply to the behavior of creatures that do not have minds, such as plants or insects. But we often want to say that their behavior is goal directed, too. The proponent of the idea that goals are conscious intentions will insist that, when we say that plants are goal directed, or that machines are goal directed, we are speaking metaphorically. We are engaged in a bit of anthropomorphism. We observe that their behavior is *in some respects* like conscious intentional behavior, and so we metaphorically extend the notion of goals to describe them (Woodfield 1976, 194; Nissen 1993, 48).

I do not see any easy way to refute a philosopher who insists that all talk of goal directedness presupposes conscious intentions. But does that mean that there is nothing more to be said about the notion of goal directedness? Not at all. For notice that the proponent of this analysis (goals as conscious intentions) is willing to concede that the behavior of some organisms and machines is similar *in some respects* to human purposeful behavior, and that it is *by virtue of* these similarities that we are inclined to call those machines or organisms "purposeful." So, *in what respect*, precisely, is the behavior of machines and insects similar to conscious human behavior? Even if we believe that goals are *conceptually* equivalent to intentions, we can still discover some interesting things about the behaviors that are distinctive of goal directed systems.

Many philosophers have pursued this way of thinking. There are two lines of thought here. The first line of thought is that goal directed systems can be identified by a distinctive pattern of behavior. In other words, the idea is that, just by looking at a system from the outside, in complete ignorance of its inner mechanisms, we can judge that the system is goal directed. The core idea here is that a goal directed system is one that has the right kind of behavioral flexibility (often dubbed "persistence and plasticity") in attaining or maintaining the presumed goals. I will call this the "behavioristic" approach to goal directedness.

The second line of thought is that, when we say that a system is goal directed, we are not *merely* saying something about its distinctive patterns of behavior. Rather (or in addition to that) we are saying something about the mechanism that gives rise to that behavior. I will refer to these as "mechanistic" analyses of goal directedness. One version of this idea, which goes back to the cyberneticists, is that goal directed behavior is behavior that is governed by a special sort of mechanism called a "negative feedback system." Things get somewhat confusing here because, although cyberneticists like Wiener *claimed* to offer a purely behaviorist analysis of goal directedness, they clearly were not. They were tacitly, if not explicitly, committed to the claim that what makes a system goal directed is that it has the right sort of inner mechanism (see Wimsatt 1971 for discussion). Moreover, even theorists such as Sommerhoff, who were trying to articulate a purely behavioristic theory of goal directedness, were forced to make certain assumptions about underlying mechanisms. In Sect. 2.3, I will describe this mechanistic approach and identify its shortcomings.

2.2 From Intentions to Behavior

For the time being, let's set aside the idea that having a goal is simply tantamount to having conscious intentions. Let's suppose there is a literal sense in which animals, plants, and machines, can exhibit goal directedness even if they do not all have intentions. One natural idea here is that goal directedness has something to do with their characteristic forms of behavior. A popular slogan in the literature is that goal directed systems exhibit persistence and plasticity (e.g., Nagel 1977, 272).

Consider a heat-seeking missile closing in on a moving target. In what respect is it life-like? First, in the ideal case, if the target moves, the missile moves, too. The missile adjusts its trajectory to match the behavior of the target. If something interferes, such as a strong gust of wind, then (again, in the ideal case) the missile gets back on track. It exhibits *persistence*: it tends toward the goal even in the face of obstacles. This is what makes a heat-seeking missile different from, say, a projectile launched from a catapult. If one launches a projectile from a catapult and the target moves, the projectile does not change its course. Persistence also characterizes much of human purposeful behavior.

The second, plasticity, is the idea that one and the same end can be achieved through a variety of different starting points. The missile clearly exhibits plasticity in this sense. There is wide range of points from which the missile can be launched, and it will still result in the same outcome. In short, plasticity has more to do with the starting point of the goal directed behavior. Persistence has more to do with the behavior as it is being carried out.

When we attribute plasticity and persistence to an item, it seems that we are merely describing its outward behavior, independent of the mechanism by which it achieves this remarkable effect. It may be the case, as a matter of contingent empirical fact, that all systems that exhibit plasticity and persistence have certain inner mechanisms in common, such as negative feedback mechanisms. But arguably, when we judge that a system exhibits plasticity and persistence, we can remain neutral about the sort of inner mechanisms that cause the behavior. (Trestman 2012, 208 emphasizes this point in developing an approach to goal directedness.)

The first important attempt to give a purely behavioral analysis of goal directedness is due to the biologist Sommerhoff (1950; also see his 1969 and 1974). (One might object that Rosenblueth et al. 1943 have precedence here, but for reasons that I will shortly describe, their view is a version of the mechanistic approach.) Sommerhoff was very liberal in his use of neologisms, and he used fairly dense mathematical formalisms, so his work is not easy to read, but the basic ideas are fairly simple. Any goal directed system involves three features: the system itself (*S*), the target of the system, which is some object in its environment (*E*), and a goal (*G*). Consider a shooter who is aiming a rifle at a target. As the target moves around, the shooter repositions the rifle to match the target. Here, the shooter is the goal directed system (*S*). The target (*E*) is the thing that the shooter is aiming at. It is a physical object that moves somewhat independently of *S*. Finally, *S* has the goal *G* of hitting *E*. *G* is the outcome that explains *S*'s behavior. The same three variables can describe a chick pecking at some grain. The chick is the goal directed system, *S*. The grain is the relevant object in the environment, *E*. The goal, *G*, is that the chick eats the grain.

How would we model such a system? The most obvious way is by the use of two variables, each of which can take on different numeric values. We will call them V_S and V_E. V_S is a variable that describes some property of *S*. In the shooter example, we will use V_S to designate the angle of the rifle. Second, there is a variable V_E that represents some property of *E*. In the shooter example, we can let V_E represent the angle of the target. Finally, let's assume that the shooter's goal is simply to keep her sights on the target, rather than to destroy it. So we can describe the goal of the system, *G*, in terms of a certain relationship between V_S and V_E, namely one of equivalence. In other words, the shooter's goal is *that* $V_S = V_E$.

I still have not stated what it means to say that a system is goal directed, that is, what it means to say, "*S* has *G* with respect to *E*." Sommerhoff's main insight here is that, when we say that a system is goal directed, we are not merely describing the actual behavior of the system, here and now. We are making a counterfactual claim. We are describing how the system variable *would* change if the environmental variable were to change in various ways. Specifically, we are saying that, if V_E changed in various ways (and within a given range), V_S would change as well, *in such a way that G is satisfied* (Sommerhoff 1950, 54). We can summarize his rather abstract discussion as follows:

> Where *S* is a system, *E* is an object in *S*'s environment, *Vs* is a variable that represents *S*, V_E is a variable that represents *E*, and *G* is some relationship between *Vs* and V_E, "*S* is goal directed with respect to *E* and *G*" means:

(i) There exists a range of different values of V_S and V_E such that, if V_E were to change, V_S would also change in such a way that *G* is satisfied.

(I will add a second condition momentarily.) Much more could be added in order to extend Sommerhoff's analysis to more complex cases. Sommerhoff's examples typically embody what Canfield (1966, 5) calls the "target" schema rather than the "furnace" schema. In the target schema (such as a heat-seeking missile) a system is trying to attain some objective, rather than to maintain some inner state (such as temperature homeostasis). It is impossible to do justice to the full complexity of Sommerhoff's analysis, but I think (i) represents his core insight.

Most of the examples that Sommerhoff uses (the rifle example, the pecking example) involve a system adapting its behavior to a target that behaves in a way that is more or less independent of *S*. In other words, the relationship between *S* and *E* is asymmetrical: if *S* is goal directed with respect to *E*, *E* is typically not goal directed with respect to S (at least not with respect to the same goal). V_S depends on V_E in an asymmetrical matter. I will include this condition explicitly into the definition (though Sommerhoff (1950, 61) also mentions it):

(ii) V_S is asymmetrically dependent on V_E.

Here is a problem that arises for Sommerhoff's analysis, and it is one that affects it in its very core. I call it the problem of overbreadth. A purely behavioral criterion of goal directedness cannot distinguish goal directed systems from those that intuitively lack it. Consider a marble rolling to the bottom of a glass bowl, or a stretched rubber band snapping back to its original configuration. From a purely behavioral standpoint, either system exhibits plasticity (it can reach the same end point from a variety of starting points) and persistence (it can adjust its trajectory in the face of obstacles). But they are not goal directed. To call such systems "goal directed" would seem to trivialize the very notion.

I think the only way to avoid this problem is to give up a purely behavioral analysis. In other words, when we say that a system is goal directed, we are saying something about the mechanism that gives rise to the behavior, rather than (or in addition to) the behavior itself. A marble rolling down the side of a bowl is not goal directed because the mechanism that governs the marble's behavior does not have the right sort of complexity. The marble is too internally homogenous. Rather, when we say of, for example, a missile, that it is goal directed, we are implying that there are a number of different internal components (inside of the missile) that function somewhat independently of one another, and that they have to work together in just the right way in order to achieve the outcome (though all of this needs to be spelled out). Sommerhoff discusses this idea in some detail. He says that in order for a system to be goal directed, its governing components have to be "epistemically independent" of one another. I will return to this condition in the next section because it represents a kind of mechanistic orientation.

A somewhat different analysis of goal directedness comes from Braithwaite (1953). As I read him, Braithwaite emphasizes the idea that one and the same goal can be achieved in many different environments. That is, when we say that a system is goal directed, what we are saying is that the system tends to achieve the goal in large number of different circumstances. A post office in New York City bears the following inscription: "Neither snow nor rain nor heat nor gloom of night stays these couriers from the swift completion of their appointed rounds." For Braithwaite, this is the essence of goal directedness: "…the essential feature, as I see it, about plasticity of behavior, is that the goal can be attained under a variety of circumstances, not that it can be attained by a variety of means" (331–332).

Suppose there is a system S in some internal state i. Let S be a rat and i be the state of hunger. Suppose we put a bit of food in the room with it. Then S has the goal G, namely, the goal of eating the food. What do we mean when we say that S in i has goal G (the hungry rat has the goal of eating food)? We mean that there are a large number of different environmental conditions under which the rat will obtain its goal. Braithwaite's formulation, like Sommerhoff's, is fairly technical. To be precise, Braithwaite says that a system is goal directed with respect to G when the cardinality ("variancy") of the set of field conditions that, together with *i*, will result in a *G*-achieving causal chain, is greater than one (330–331). One nice feature of Brathwaite's analysis is that goal directedness is a property that comes in degrees. One might think that the larger the number of field conditions under which the system can achieve *G*, the greater the degree of goal directedness it possesses.

Braithwaite's analysis faces a number of problems. First, there is a problem of grain (Woodfield 1976, 34–35 makes this point). The number of field conditions under which a system can achieve its goal partly depends on how finely we discriminate between those conditions. Suppose that I am eating a sandwich in my office. Suppose that the precise configuration of air molecules in my office at this very moment represents one field condition, and its precise configuration two seconds from now represents another. Then my eating the sandwich is goal directed *merely* because I manage to eat my sandwich under a vast number of different field conditions.

I do not think this is an insoluble problem but this is the sort of thing that Braithwaite's analysis will have to register. Presumably we would want to say that two field conditions are different only insofar as they have different effects on the organism, such that the organism must adjust its behavior somewhat to achieve the same end. But then the analysis seems to have much more to do with the inherent adaptability of the system, rather than the number of environmental conditions. In other words, it seems to me that Braithwaite's analysis has the wrong focus, because it defines goal directedness by looking outward to the number of environmental conditions, rather than inward to the internal configuration of the system.

Second, Braithwaite's analysis is susceptible to what Scheffler (1959) called the "problem of multiple goals." Here is one version of the problem. Suppose there are many field conditions that, together with the internal state of the system, result in a certain event *G* (the presumptive goal). Suppose that every causal chain that results in *G* has a further consequence, namely *H* (for example, every time the rat eats a piece of cheese, it defecates). Then Braithwaite's analysis would force us to say that *H* is also a goal. That seems counterintuitive.

Third, Braithwaite's analysis shares, with Sommerhoff's, the problem of overbreadth, though unlike Sommerhoff, I am not aware that Braithwaite made any attempt to resolve it. The problem is that his simple schema can apply to ersatz goal directedness, such as rolling marbles and swinging pendulums, as well as genuine cases. Even if these behavioral or environmental analyses represent necessary conditions on goal directedness, they are not sufficient. That suggests that we have to supplement our behavioral analysis with some sort of mechanistic analysis. We must turn our focus inward, to the nature of the system that generates the behavior, rather than the behavior itself or its environment.

2.3 From Behavior to Mechanisms

We have seen in the last section how purely behavioral analyses of goal directedness stumble, and they stumble because of the problem of overbreadth. Theorists like Andrew Woodfield (1976) use the failure of behavioral analyses to support their own mentalistic theories of goal directedness. As I noted in Sect. 2.1, Woodfield's view is that, as a piece of conceptual analysis, goals are simply conscious intentions (also see Nissen 1993). To attribute a goal to a system is to attribute an intention to it. In his view, plasticity and persistence are, at best, *reliable indicators* of the presence of conscious intentions, but they are fallible indicators, in the way that crying is a reliable, though fallible, indicator of sadness. In Woodfield's view, purely naturalistic approaches to goal directedness inevitably go wrong because they mistake a reliable indicator of goal directedness (plasticity and persistence) for goal directedness itself. As I noted above, I do not think there is any easy way to refute this point. The best way to respond to Woodfield's point is to come up with a good naturalistic analysis of goal directedness that avoids the problem of overbreadth.

The most obvious idea for avoiding the problem of overbreadth is to take a more mechanistic approach. In other words, in order for a system to be goal directed, it is not enough that it behaves in a certain way. Rather, the behavior has to be caused in the right way. The behavior has to be caused by the right sort of inner mechanism. The reason we do not attribute goal directedness to a rolling marble is that its inner constitution is too simple. It does not have enough moving parts, or the right sort of organization between those moving parts. It is just too internally homogenous. On the other hand, a moth that instinctively heads for the light has the right sort of inner constitution to qualify as goal directed. It has a number of moving parts that have to work together in just the right way to make its phototropic behavior possible.

So, what exactly must this inner constitution be like in order for the system to qualify as goal directed? There are two ideas here. The first is that the system has to be governed by negative feedback. The second is that the components of the system must have the right sort of independence. Both approaches are problematic. I will begin with negative feedback.

In their important paper, Rosenbleuth et al. (1943) attempt to explicate a naturalistic notion of purposefulness that applies to machines as well as animals. Unfortunately, they are not entirely consistent in their terminology. Sometimes they use "purpose" very generally: "the term purposeful is meant to denote that the act or behavior may be interpreted as directed to the attainment of a goal" (18). Sometimes they say that purposeful behavior is just behavior controlled by negative feedback: "all purposeful behavior may be considered to require negative feed-back" (19). I think the most consistent way to interpret their view is that goal directed systems are, necessarily, governed by negative feedback mechanisms. So what is negative feedback? As it turns out, defining the notion of negative feedback raises a host of fresh problems.

A feedback system is a special kind of input-output system. It takes inputs from its environment and produces certain outputs as a result. (There has to be some objective way of demarcating what constitutes its input and what constitutes its output. If I kick a pile of rocks and cause them to scatter, we would not say that the force of my kick is the input and the scattered rocks are output. I will not attempt to define these notions.) More specifically, a feedback system takes some of the energy from its output and uses it as input, making a loop. A thermostat is a kind of feedback mechanism. It detects the temperature of the room (its input) and then uses that information to produce a certain output (e.g., switching on a furnace). When the temperature of the room increases (as a result of its behavior) it does something else (switches it off). So, at the very minimum, a feedback system has to be equipped with a sensor (that monitors some feature of the environment) and an effector (that produces a behavior). Specifically, the effector has to be able to produce a behavior that modifies the environment in a way that the sensor can detect. That is what creates the feedback loop.

There are two sorts of feedback loops, positive and negative. In a positive feedback loop, some of the energy of the output feeds back into the system and produces more of the same sort of behavior. It has an amplifying effect. In a negative feedback loop, some of the energy of the output feeds back into the system and produces the opposite sort of behavior. It has a dampening effect. Negative feedback devices are common in the natural world because they give rise to self-regulating behavior. For example, negative feedback is used to regulate the level of synaptic dopamine produced by mesolimbic dopamine neurons (see Kandel et al. 2013). When those neurons are activated, they release some dopamine into the synapse. The axon terminal of the neuron has a number of autoreceptors that monitor the amount of dopamine in the synapse. When the level gets too high (that is, when a large number of autoreceptors are bound by dopamine), the autoreceptor sends a signal to stop releasing dopamine. When the level gets too low, it sends a signal to release more. The dopamine neuron constantly monitors its own behavior in order to modify that behavior, and it uses negative feedback to do so.

Negative feedback systems can give rise to plasticity and persistence. Suppose we say that the goal of the dopamine neuron is to maintain a constant level of synaptic dopamine. The neuron exhibits persistence in maintaining this goal. Various perturbations cause the level to increase or decrease, and the neuron modifies its own behavior in such a way as to maintain a steady level. It also exhibits plasticity, as it can achieve this effect across a wide range of starting points. Heat-seeking missiles do the same sort of thing. The missile senses the direction of its target. It then uses this information to modify its own trajectory. This is achieved through an error-minimization computation. It calculates the difference (error) between its own trajectory and the trajectory of the target. It then adjusts its direction in such a way as to minimize that difference.

The reason that the rolling marble is not a goal directed system is that it does not have the right inner constitution. It does not have anything that would objectively correspond to a distinctive sensor and effector. Nor is there anything in its inner constitution that would instantiate an error-minimization computation. The marble is completely passive in relation to its environment. Its behavior is most effectively explained in terms of the external forces (gravity and friction) that control it.

There are several major problems with this feedback approach. I will discuss three of them. The first has to do with the definition of "feedback." Wimsatt (1971) argued that the term is poorly defined. For example, Rosenblueth et al. (1943, 19) define a feedback system simply as one in which "some of the output energy of the apparatus or machine is returned as input." But this general definition encounters trivial counterexamples. Wimsatt (1971, 251) claims that any "closed loop of material transport" constitutes a feedback system, for example, two hoses that cycle water around and around.

Of course, nothing prevents us from developing a much richer definition of feedback (as in Adams 1979). One could say that a negative feedback system must have some mechanism by which it receives signals emanating from the presumptive target, and it must use this information to modify its trajectory. Presumably, this will involve an error-minimization computation (that is, it must have an internal representation of the target, and of its current behavior, and it must be able to compute the difference between them and use this information to adjust its behavior). But the problem is that we are invoking several terms that stand in as much need of clarification as the term "goal directedness" itself, such as "information," "computation," and "representation." Short of an analysis of these crucial notions, it is hard to tell whether this constitutes an advance.

Second, even if we could avoid the definitional problems of negative feedback, it is not clear that having a negative feedback device (in some suitably rich sense) is either necessary or sufficient for goal directedness. Woodfield (1976) summarizes the problems well. He points out that many electronic devices, such as televisions and radios, utilize various electronic feedback mechanisms, but they are not thereby goal directed (189). So, being controlled, in part, by negative feedback is not sufficient for goal directedness. Nor is it necessary. Consider the frog that snaps at a passing fly. Once the tongue-snap is triggered, it is no longer controlled by negative feedback. If the fly moves, the frog cannot modify the trajectory of its tongue in the course of a single snap. But the tongue-snap itself, that is, that specific piece of behavior, seems goal directed (191). See Faber (1986, 80), however, who criticizes Woodfield's choice of examples. Faber points out that some of the feedback systems in electronic equipment, such as feedback oscillators, are not negative feedback devices, and that, even though a single tongue-snap is not governed by negative feedback, it is part of a system of behavior that is, so it still fits the general analysis.

Third, and finally, even if we have a satisfying and non-vacuous notion of negative feedback, we encounter the problem of the missing goal object (Scheffler 1959). The basic idea behind the negative feedback approach is that in order for a system to be goal directed it must utilize signals from the target to modify its behavior. Yet presumably, a system can be goal directed even if the target does not exist. People can have goals regarding imaginary objects, such as the fountain of youth. A cat can crouch by a hole with the goal of catching a mouse, even if there are no mice around. The cat is not utilizing signals from the target to guide its behavior. (Ehring 1984, 501 elaborates a related problem for Nagel's account.)

Let's set aside the negative feedback approach. A second approach emphasizes the idea that, in a goal directed system, the controlling variables are all independent of, or orthogonal to, one another, in a special sort of way (Sommerhoff 1950; Nagel 1953, 1961, 1977). Sommerhoff called this the "epistemic independence" condition. Ernest Nagel also developed this idea. I will first explain Nagel's own theory of goal directedness very briefly, and then I will return to this notion of independence.

Nagel proposed something like a mechanistic account of goal directedness. As I noted above, for Sommerhoff, the paradigmatic cases of goal directedness involve a system attempting to attain some end, such as a shooter aiming for a target or a bird pecking at grain. In contrast, Nagel adopted what Canfield (1966, 5) describes as a "furnace" schema. Nagel was mainly interested in homeostatic behavior, such as the system that maintains a relatively stable body temperature in the face of external changes, or the system that ensures that the water content of the blood stays around 90 % regardless of fluctuations in intake. Here, goal directedness is not so much about achieving a certain end as it is about maintaining a steady internal state. The environment is, if anything, a source of perturbation. I think of it as a mechanistic analysis because it focuses more on the way that we decompose a system into parts, and in the way that those parts interact with one another, than with the behavior of the system in relation to some external object.

Specifically, Nagel says that when we describe a system as being goal directed, we analyze that system into a set of components. Consider the system that regulates the water content of the blood and keeps it at around 90 %. There are three components that are most directly relevant: the kidneys, the muscle, and the blood. When the water content of the blood starts to drop, the muscles release more water into the blood. When the water level gets too high, the kidneys extract more water from the blood. Each component can be described by a certain variable that can take different numeric values. One variable V_M represents the rate at which the muscles release water into the blood. Another variable V_K represents the rate at which the kidneys extract water. A third, the goal variable V_G, represents the water content of the blood. The important point is that when the goal variable departs from the goal-state (when V_G is greater than or less than 90 %) the other variables change in such a way as to bring the system back into the goal state. So, goal directedness has to do with the way the components of the system interact so as to maintain what he called the "goal-state."

(Incidentally, one might think that, formally speaking, Sommerhoff's analysis is just a special case of Nagel's analysis. Sommerhoff examines the relationship between a system variable V_S and an environmental variable V_E. He says that a system is goal directed with respect to goal G when the following counterfactual is true: if V_E were to change in certain ways, then V_S would also change in certain ways such that G is attained or maintained. In contrast, Nagel examines the components within a system, the V_i. He says that a system is goal directed under the following conditions: if V_G changes, the other V_i change in such a way as to maintain G. Perhaps Sommerhoff's analysis reduces to Nagel's analysis if we re-label Sommerhoff's "V_E" with Nagel's "V_G," that is, if we treat Sommerhoff's V_S and V_E as two components of a larger system. I think the only problem with the idea is that, as I noted above, Sommerhoff emphasizes a certain asymmetry between V_S and V_E. V_E changes more or less independently of V_S; V_S just adapts itself to V_E. But on Nagel's view, there is an inherent symmetry between the variables. They affect each other's behavior equally. For example, if the water content of the blood (V_G) changes, the rate at which kidneys extract water (V_K) also changes. But the converse is also true. If V_K changes, then so does V_G. At any rate, I think there is some interesting logical relationship between the two analyses but it is not entirely trivial to flesh it out.)

Let's return to the problem of overbreadth. Nagel recognized that his analysis of goal directedness suffered from the problem of overbreadth. Consider a pendulum at a state of rest. A gust of wind displaces it, and the pendulum swings back to rest. Is the pendulum goal directed with respect to the state of rest? After all, we can analyze the movement of the pendulum in terms of two variables, the force of displacement and the force of restoration, and we can describe the changing relationship between these variables in terms of the formal apparatus he recommends.

Nagel, borrowing from Sommerhoff's analysis, tries to avoid this problem by saying that the variables of a goal directed system must have the right sort of independence (Nagel 1953, 211). As Nagel (1977, 273) put it, there has to be an "orthogonality of variables." Specifically, at a given moment, the value of one controlling variable must not determine the value of any other controlling variable, *at that very moment*. At any point in time, each controlling variable must maintain some degree of freedom from the other controlling variables. Consider the system that regulates the water content of the blood. At any given moment, the rate at which the kidneys extract water is independent of the rate at which the muscles release water. We could test this claim by performing the right sort of intervention, for example, by artificially decreasing the rate at which the kidneys extract water. *Eventually* our intervention will affect the rate at which the muscles release water *but not at the very moment* we perform the intervention. By contrast, consider a pendulum swinging to a state of rest. The controlling variables include the force of displacement (away from the vertical position) and the force of restoration. Yet, at any given moment, these two forces must be equal and opposite, according to Newton's third law of motion. Thus, the pendulum's controlling variables lack the right sort of independence.

Other physical systems also lack the right sort of independence to qualify as goal directed. By Boyle's law, the pressure and volume of a given amount of gas, at a constant temperature, are inversely proportional. So, they are not epistemically independent. Nor are the voltage and current flow in an electrical circuit of constant resistance (by Ohm's law). So, the idea of epistemic independence seems to give us exactly what the other analyses lack. It gives us a plausible principle that we can use to differentiate real and fake goal directed systems.

Some commentators have attacked this maneuver by arguing that the very idea of epistemic independence is incoherent or that it fails to distinguish the different sorts of systems. Woodfield (1976, 67) criticizes it on the grounds that, in a genuinely goal directed process, the variables are never entirely independent of one another. Nissen (1980, 130) echoes this criticism. So long as the goal directed system is in functional condition, he argues, the variables will all be correlated with one another and hence derivable from one another. But I think this misses Nagel's point. Nagel's point was that, in a goal directed system, the variables are independently modifiable *at one and the same moment*. Again, go back to the system that regulates the water content of the blood. At any time, t, if I increase the water content of the blood at t, the behavior of the kidney will not change *at t*. It will change at some $t_1 > t$. But I cannot do this for the pressure and volume of a gas at constant temperature, or for the forces of displacement and restoration on a pendulum. The basic laws of nature do not allow one to modify independently the force of displacement and force of restoration of a pendulum. So it is not a goal directed system.

But problems remain. Presumably, even in a legitimate goal directed system, not *all* of the controlling variables will be independent. Humans have the goal of maintaining balance. The movement of fluid in the inner ear mediates this goal. The movements of the fluid are sensitive to the rotation of the head via basic physical laws (Newton's laws of motion) so at least some of the controlling variables that allow us to maintain balance are not independent. Moreover, at a certain level of description, even the atoms that control the movement of a rolling marble are somewhat independent of one another. Does that mean that, in a rolling marble, the controlling variables are epistemically independent? The problem here is that whether the set of controlling variables are independent of one another depends crucially on how we decompose the system. On some decompositions, they will be independent, and on others, they will not be. That suggests that whether or not a system is goal directed depends on how we choose to analyze it.

Interestingly, Nagel himself seemed to be perfectly happy with this outcome. In other words, he was willing to concede that whether a system is, or is not, goal directed partly depends on how we choose to analyze it (Nagel 1977, 275). According to this way of thinking, the notion of goal directedness is part of an *analytic strategy* for coming to grips with complicated systems. It is a useful construct that helps us to make sense of systems that are otherwise intractably complex, but it is not a mind-independent fact of nature. Imagine trying to describe and predict the complex movements of the frog's head, retina, and tongue, without describing the frog as having the goal of catching flies. It can scarcely be done.

The biologist McShea (2012) recently defended the value of goal directedness for studying complex systems. He himself develops a viewpoint that is similar in some ways to Braithwaite's in that it emphasizes the role of the environmental variables in bringing about the adaptive behavior. His view is that all of the traditional examples of goal directedness, such as homing torpedoes, fly-catching frogs, and even human intentions, exemplify the very same physical structure. In each case, there is a larger entity that constrains the behavior of a smaller entity within it, without determining it precisely. He uses an example of a bacterium moving up a concentration gradient. The larger system is the concentration gradient, and the smaller system is the bacterium. The concentration gradient influences the behavior of the bacterium in such a way that it appears goal directed. He believes that a similar dynamic plays out in the context of evolution by natural selection and even human intentional behavior.

Now, McShea does not purport to offer a conceptual analysis of "goal directedness." Rather, he is simply trying to identify, empirically, a common structure that underlies most cases of goal-directedness, but which also underlies some behaviors that do not strike us as goal directed (such as a swinging pendulum). The important part, for my purposes, is that he is happy to concede that, if one were forced to distinguish precisely between systems that are goal directed and those that are not, one would have to do so on pragmatic grounds. As he puts it, in discussing the problem of overbreadth, "the present suggestion that goal directedness is a function of the perceived complexity of the system is likewise relative to our knowledge" (682). Trestman (2012, 215) also suggests that what counts as a goal directed system is relative to our perspective.

I have to admit that I find this pragmatic and, if you will, anti-realist line unsatisfying. The reason is simple. It seems to me that whether or not frog has the goal of catching flies does not depend on how much knowledge we happen to have of it, but is intrinsic to it. Perhaps a few hundred years from now, the frog's ability to catch flies will seem as simple to scientists, then, as swinging pendulums seem to us, now, and scientists will find no use for analyzing the system as goal directed. But I take it that frogs will still have goals. This is not just because the frog has conscious intentions, which I believe it does. It is because whatever (behavioral, psychological, or mechanistic) hallmarks we are picking out *now* that explain our judgment that the frog is goal directed will still be true of it hundreds of years from now when its behavior can be entirely explained at the level of sub-atomic physics and when the language of goals and purposes fails to yield any novel predictive benefits. But I appreciate that we are wandering into more general issues in the philosophy of science pertaining to realism and antirealism that cannot be resolved here.

References

Adams, F. R. (1979). A goal-state theory of function attributions. *Canadian Journal of Philosophy, 9*, 492–518.

Boorse, C. (1976). Wright on functions. *Philosophical Review, 85*, 70–86.

Braithwaite, R. B. (1953). *Scientific explanation*. Cambridge: Cambridge University Press.

Canfield, J. (1966). Introduction. In J. Canfield (Ed.), *Purpose in nature* (pp. 1–7). Englewood Cliffs, NJ: Prentice-Hall.

Ehring, D. (1984). The system-property theory of goal directed processes. *Philosophy of the Social Sciences, 14*, 497–504.

Faber, R. J. (1986). *Clockwork Garden: On the mechanistic reduction of living things*. Amherst: University of Massachusetts Press.

Gilbert, S. F., & Sarkar, S. (2000). Embracing complexity: Organicism for the 21st century. *Developmental Dynamics, 219*, 1–9.

Kandel, E. R., et al. (2013). *Principles of neural science* (5th ed.). New York: McGraw Hill.

McShea, D. W. (2012). Upper-directed systems: A new approach to teleology in biology. *Biology and Philosophy, 27*, 663–684.

McShea, D. W. (2013). Machine wanting. *Studies in History and Philosophy of Biology and Biomedical Sciences, 44*, 679–687.

Nagel, E. (1953). Teleological explanation and teleological systems. In S. Ratner (Ed.), *Vision and action* (pp. 537–558). New Brunswick, NJ: Rutgers University Press.

Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace and World.

Nagel, E. (1977). Teleology revisited: Goal directed processes in biology and functional explanation in biology. *Journal of Philosophy, 74*, 261–301.

Nissen, L. (1980). Nagel's self-regulation analysis of teleology. *Philosophical Forum, 12*, 128–138.

Nissen, L. (1993). Four ways of eliminating mind from teleology. *Studies in History and Philosophy of Science A, 24*, 27–48.

Piccinini, G. (2015). *Physical computation: A mechanistic account*. Oxford: Oxford University Press.

Rignano, E. (1931). The concept of purpose in biology. *Mind, 40*, 335–340.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of Science, 10*, 18–24.

Russell, E. S. (1945). *The directiveness of organic activities*. Cambridge: Cambridge University Press.

Schaffner, K. (1993). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press.

Scheffler, I. (1959). Thoughts on teleology. *British Journal for the Philosophy of Science, 9*, 265–284.

Sommerhoff, G. (1950). *Analytical biology*. London: Oxford University Press.

Sommerhoff, G. (1969). The abstract characteristics of living systems. In F. E. Emery (Ed.), *Systems thinking* (pp. 147–202). Middlesex: Penguin.

Sommerhoff, G. (1974). *The logic of the living brain*. New York: Wiley.

Taylor, R. (1950). Comments on a mechanistic conception of purposefulness. *Philosophy of Science, 17*, 310–317.

Trestman, M. A. (2012). Implicit and explicit goal directedness. *Erkenntnis, 77*, 207–236.

von Bertalanffy, L. (1950). An outline of general systems theory. *British Journal for the Philosophy of Science, 1*, 134–165.

Wiener, N. (1948). *Cybernetics*. Cambridge, MA: MIT Press.

Wimsatt, W. C. (1971). Some problems with the concept of feedback. *Boston Studies in the Philosophy of Science, 8*, 241–256.

Wimsatt, W. C. (1972). Teleology and the logical structure of function statements. *Studies in the History and Philosophy of Science, 3*, 1–80.

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

http://www.springer.com/978-3-319-32018-2

A Critical Overview of Biological Functions
Garson, J.
2016, VIII, 113 p., Softcover
ISBN: 978-3-319-32018-2