

The Process Dynamics of Normative Function

Wayne D. Christensen Mark H. Bickhard

“The hand separated from the body is not a true hand.”

Aristotle, *Politics* I.2.1253a20-21.¹

Introduction

In this paper we outline a theory of normative functionality aimed at understanding the nature of adaptive systems as globally structured, integrated systems. More specifically, the account is concerned with understanding the process relations constitutive of such systems. The explanatory agenda of this approach includes the following questions:

- How does the system work? That is, what allows an adaptive system to ‘hang together’ as a complex interconnected system?
- What possibilities are open to the system under which it retains or improves its integrity as an adaptive system?
- Where does the system break down?

The central plank of our account is a theory of autonomous (or ‘self-governed’) systems that are composed of webs of interdependent processes whose collective activity is self-generating. In other words, autonomous systems possess a process organization that, in interaction with the environment, performs work to guide energy into the processes of the system itself. A paradigm autonomous system is a living cell, which transforms energy and materials from the environment through metabolic pathways into the forms of chemical energy and organic molecules required by the processes that sustain the cell. Examples of the processes by which cells help generate the conditions under which they survive include enzymatic catalysis, membrane regulation of energy and material flows in and out of the cell, and chemotactic amoebic motion to find a food source such as glucose. As such, autonomy is a property of widespread biological significance; living organisms in general are autonomous systems, as are reproductive lineages, species, and some kinds of biological communities. Autonomy also has wider relevance, however, since socio-cultural systems such as businesses, cities, geo-political regions, and nations can also be autonomous.

The account of normative function that we propose focuses on autonomy as a basic normative dynamical constraint on the functional organization of adaptive systems. Specifically, adaptive systems have to be functionally organized in a way that generates the conditions under which the system is autonomous. On our account that is what it is to be adaptively viable. Normative functional organization itself is, in turn, analyzed in terms of the interdependencies of the processes that constitute the system. Each of the processes that form part of the system requires outputs from other processes in the system to function, and in turn contributes to the requirements of other processes. These process

¹ Cited in Depew and Weber 1995, p.35.

interdependencies constitute norms on adaptive functioning, since if the requirements of a particular system process aren't met it will cease to produce the outputs required by other processes, potentially resulting in propagating dysfunction that may reduce or destroy the viability of the system.

This account is quite different to the orthodox etiological theory of normative function, and bears a complex relationship to it. In certain respects we are simply focusing on a different issue: etiological theory takes as its task the problem of explaining what it is for a part of a system to *have* a function, whereas we focus on what it is to be an adaptive system, and on the nature of *servicing* a function relative to such a system. In this paper we will not specifically address the issue of what it is for a part to 'have' a function. Our account does not threaten at least some versions of etiological function theory, and can be seen as a complementary form of analysis.

However, since it is generally assumed that the primary issue for theories of normative function *just is* to explain 'having' a function, there is a deeper clash of agendas between most contemporary etilogists and us concerning what should and should not form part of normative function theory. Our view is that the agenda we describe above is not only an important part of understanding normative functionality, it is conceptually fundamental for any naturalist theory of normative functions. Further functional issues, including understanding what it is for a function to be 'owned' by a part, are derivative on these issues. However in this paper we will concentrate on presenting the basic structure of the positive account rather than argue for this general thesis or develop a detailed critique of the dominant etiological position. With respect to the latter, we will confine ourselves to deflecting some arguments which have been taken to show that what we are trying to do is impossible.

1 Some reasons why what we are trying to do is impossible

Systems-oriented approaches to function, usually associated with the analysis of Cummins (1984), are commonly regarded as being unable to explain normative function. Indeed, they are sometimes dismissed as being largely or entirely arbitrary in their conditions of application, and hence without any significant theoretical interest for biology or philosophy. In some ways this negative view is rather surprising, since so much of biology is concerned with understanding systems of various kinds, such as organisms, reproductive populations, communities, ecologies, and so on. We will now examine some of the most common arguments and assumptions used to justify the marginalization of systems approaches, and argue that the conclusions are overstated and do not count against the kind of theory we are proposing. These arguments are as follows:

- (i) The agenda for normative function theory is concerned with understanding design, which system function theories cannot explain.
- (ii) Causal systems analysis is observer relative and hence arbitrary.
- (iii) Biological analysis starts with classification by proper function.
- (iv) Selection is the only possible naturalist explanation of biological norms, which are concerned with performance in accord with design specifications.

(i) *The agenda for function theory is understanding design.*

The etiological theory of functions focuses on identifying processes by which components come to be included in systems because of the role that they play, and on associating functions with those parts. Thus, on this theory the normative function of a part is specified in terms of the reasons why it was incorporated in the system; intuitively, its 'design specifications'. Following Wright (1973) these have come to be regarded as the canonical issues for function theory. Nevertheless, it is important to recognize that Wright's agenda is a hypothesis about the best way to analyze normative function, not a definition of it. 'Wild-type' function concepts are more complex than Wright's analysis suggests, and some scientific disciplines such as cellular physiology frequently use concepts that fit the functional agenda we outlined above more closely than they do Wright's account.

For example, protein domains are classified in a variety of ways, including by primary sequence, tertiary conformation, 'functional action' (what the domain binds to), homology and evolutionary function. Each kind of classification is relevant to particular kinds of relations between proteins and wider cellular processes, and, significantly, none of the other categories is derived from or parasitic upon evolutionary function. The overall focus of cellular physiologists is on understanding cells as integrated systems, and consequently they are concerned with all of the attributes of proteins that affect wider cellular organization, not just the aspects of protein structure that can be distinctively attributed to selection. Similarly, functional morphologists study 'functional integration', by which is meant "the interconnectedness of structures and their CR [causal role] functions" (Amundson and Lauder 1998, p.252). For example, it might be discovered that the interconnections of the components of jaw of a particular species are such that modification to a single muscle in the jaw has deleterious effects on mouth opening (ibid., pp.252-3).

Here is an everyday example of normative function in our sense. The Jones family says that they like the house that they live in. They don't mean that it does exactly what the architect intended that it should. They mean that it is a good house to live in; it is 'highly functional'. The Jones's might go on to list features of the house that make it nice to live in. Some of these features may have been intended by the architect to do what they do, such as a north facing aspect that provides sun all year round. Other features are unintended, such as the fact that there are no bedrooms along the western wall, which faced onto an empty lot when the house was built and now faces onto a noisy road. What matters to the Jones's when they call it a good house for them is not how those features got there, but how they work now as an integrated 'livable' package under the circumstances.

Similarly, when a business analyst analyzes the viability and investment value of a business, success and failure, and the detailed systemic interrelationships relevant to this, are determined with respect to how well the business functions as an overall entity. Design issues, such as the management strategy of the business, are an important *part* of this analysis, but they are not the whole of it. What fundamentally matters to a business

analyst is how effectively the business functions as an integrated viable system, not whether it is doing exactly what its founders or current managers intend it should be doing. Some companies are successful through dumb luck and despite indifferent management.

One of the key criticisms leveled against systems theories of function is that they don't explain design, so they can't explain functional norms. However, this conclusion doesn't follow if there is more to normative functionality than design. The examples just given provide evidence that this is so, and specifically suggest that there are important normative functional issues associated with understanding how adaptive systems function as integrated systems under actual interaction conditions. Consequently, the problem is not that systems approaches to function don't address the right agenda, it is that the agenda for normative function theory should be broader.

(ii) Systems analysis is observer relative.

Another criticism leveled against systems-oriented approaches is that what is to be counted as a system can only be defined relative to the observer's interests, so there is no principled basis for performing such functional analysis. This idea lies behind the examples of trivial function attribution – such as that a function of mice is to explode in space (Millikan 1989), or that a function of the heart is to contribute to body mass (Sober 1993) – intended to show that functions defined by such causal systems analysis are essentially arbitrary. But such whimsical function attributions do not show that there are no principled causal ways to define systems, just that the author has failed to imagine one. It is true that Cummins' account of causal functional analysis doesn't provide any principles for application other than the informational value of the analysis, but this doesn't rule out the possibility that there may be other principled criteria. Moreover the claim amounts to asserting that there are no naturally individuated systems, which is a strikingly bold and counterintuitive hypothesis. In other words, if causal systems analysis is entirely interest relative, it follows that all ways of partitioning the world into causal groupings are ad hoc. This is a substantive thesis about the physical world, and whilst it often isn't obvious where to draw the boundaries for causal systems, neither is it obvious that solar systems, plate tectonics, weather systems, ecosystems, organisms, businesses, etc. are all just ad hoc groupings of causal phenomena.

On closer examination there are a variety of causal properties that can serve as a basis for principled specifications of system identity that are not observer-dependent. The concept of cohesion developed by Collier (1988) provides one promising approach to developing a principled causal analysis of system identity. This account was originally developed in the context of understanding species cohesion: i.e., the problem of understanding how species are to be individuated, and the processes that maintain this identity. On Collier's analysis cohesive systems in general are individuated by causal interactions that produce emergent unified dynamical behavior. Thus, the molecular bonds in the crystal lattice of a rock cause the rock as a whole to behave as a unified system under a large range of interactions. If you apply a moderate force to a rock sitting on the ground, e.g. by kicking it, the rock as a whole will move whilst the ground remains where it is. The rock is

coupled to the ground by gravity and friction, but these forces are weaker than the molecular bonds of the rock, so the rock behaves as an integrated system when kicked, whereas the rock plus ground doesn't. Which is lucky for you. If you kick a pile of sand, on the other hand, the sand doesn't subsequently behave as an integrated system in the way that the rock does. Instead, grains of sand fly off in all directions. This is because the causal interactions between the grains of sand do not form bonds that are strong enough for the pile to behave as an integrated system under the perturbation of being kicked.

The cohesion of any physical system is limited to within a range of conditions. Thus, if you hit a rock with a hammer and break it then you disrupt its cohesion, and it ceases to behave as an integrated system. This doesn't make cohesion an arbitrary property however, since the range of conditions under which a system is cohesive can be physically specified (e.g. a rock will be cohesive under a particular force and temperature range), and it is a determinate matter whether a particular system is cohesive with respect to its current interactions. Moreover the property of cohesion generates further causal properties: for instance, it is the cohesion of a kite that allows it to fly, because the integrity of the kite structure acts to sum the forces of the small interactions of air particles against it, generating a net lift force.² If cohesion isn't a causally significant property it is difficult to know what is. So understanding system identities is a far from casual, observer-relative issue.

We are not claiming that the concept of cohesion alone is sufficient for specifying the kinds of system identities relevant to normative function, however. So we aren't saying that rocks have normative functions. But there are a variety of kinds of cohesion, and this raises the possibility that there is a specific kind of cohesion that forms the basis of the identity of biological systems. If so, characterizing this form of cohesion could serve as a framework for principled systems analysis of biological functionality. The theory of far-from-equilibrium systems provides a useful starting point for finding a biologically relevant form of cohesion, since it is a highly distinctive causal property of biological organisms that they are far-from-equilibrium. When they stop being far-from-equilibrium they stop being alive. The account of autonomy that we outline below identifies autonomy with far-from-equilibrium systems of a particular kind, namely ones that are self-maintaining. On our account this is the appropriate principled basis for analyzing the functional organization of biological systems as integrated systems.

In this context it is worth examining Millikan's reasons for rejecting the idea that the concept of a 'life system' is well defined, other than in terms of proper functions. She says:

A life system maintains itself. Moreover it often does so in an environment that changes in important respects both over the individual organism's lifetime and from generation to generation. *[Therefore, t]he question what counts as part of the "species-maintaining" Cummins system, as opposed to helpful but accidental interference from outside, has no determinate answer.* (Millikan 1999, p.196, italics added.)

² This example is from Collier (1988).

The only way to place nonarbitrary limits on what counts as part of a biological system is to bound it with adaptations, with proper functions. Only Cummins functions that either are adaptations or support adaptations are sensibly treated as belonging to biological systems. (Ibid., p.197.)

There are several points to make about these claims. Firstly, whether there *is* a self-maintaining life system in a given case is a question that has a coherent physical answer. If it didn't then evolutionary biology would be impossible, because selection acts *on* physical systems. There must, therefore, be some physically well-defined way of characterizing the system. Furthermore, the fact that the conditions of self-maintenance vary over an organism's lifetime and across generations (due to environmental variability amongst other things) does nothing to show that being a living system is not a physically well-defined property.³ To suppose otherwise you have to assume that systems can only be defined as static structures. But the dynamical nature of living systems is itself a good reason for assuming that this isn't so. Moreover, developments in the 'complexity' sciences, such as nonlinear dynamics and far-from-equilibrium thermodynamics, are providing conceptual tools for characterizing systems that don't depend on assuming that systems are static structures. A particularly relevant concept here is that of 'nonstationary systems': systems whose dynamical form changes over time. Living systems may be nonstationary, but nonetheless perfectly good natural systems for all that.

Secondly, although the highly interactive and flexible nature of living systems means that specifying what counts as *part* of such a system is not a simple matter, it can't be concluded on that basis that it is entirely arbitrary. All living systems depend on a complex array of factors for viability, each of which must be specified for a full account of the system's characteristics. However there are good reasons for distinguishing some factors as distinctively part of the system and others as external to it. Thus, a plant depends on both its leaves and the sun for photosynthesis, but you wouldn't want to say that the sun is part of the plant. Our proposal for addressing this issue hinges on the concept of process interdependency described in the introduction. Specifically, there is an important distinction to be drawn between processes that contribute to the system without depending on the system for their occurrence, and those that both contribute to the system and also depend on it. Thus, sunshine plays a role in plant photosynthesis, but the nuclear fusion processes that generate sunshine don't depend on the plant. In contrast, the photochemical processes in the plant's leaves that use light to construct useful organic products *do* depend on the plant to occur. They are both generated by the plant and in turn contribute to it, so they are part of the plant by our criterion. Note that the criterion does not require morphological attachedness.⁴

³ Clearly the range of conditions in which we expect that a real system can sustain itself is influenced by history: a polar bear is expected to survive cold, but not a tropical butterfly. But this does not mean that viability itself is not a well defined except by reference to selection history. Polar bears also survive in sub-tropical Australian zoos.

⁴ This criterion is similar to the category of 'collectively replicated developmental resources' distinguished by Griffiths and Gray (1994).

To take a less extreme contrast, using this criterion we can distinguish between birds' nests and rock ledges used by birds for nesting. A nest constructed by the bird is part of the distributed self-maintaining bird life cycle in the sense that it both contributes to the lifecycle and is itself generated by it. On the other hand, rock ledges used by birds for nesting contribute to the birds' lifecycle but aren't *part* of the lifecycle, in the sense that they aren't generated by it. Important adaptive consequences follow from this distinction. Birds that rely on the provision of ledges by the environment for nesting have an ecological limitation which birds that construct their own nests don't. Putting it roughly, the more an adaptive system actively constructs the conditions for its own viability, the less dependent it is on specific and perhaps idiosyncratic ecological circumstances. Which is not to say that being a generalist is always the best adaptive strategy.

Now, we emphasize again that to understand a 'life system' *all* of the important contributing factors have to be included, and in this broad sense the sun and rock ledges are part of life systems. To a significant degree Millikan overemphasizes the issue of 'part' status – of the ontology of being a 'part' that 'has a function' – at the expense of understanding factors that are ('merely' or 'accidentally') useful, as though being useful was a trivial issue, and as though the distinction was itself exhaustive.⁵ Nevertheless, it is incorrect to claim that the question as to what forms a part of a self-maintaining life system is without sense. It is possible to characterize the interconnected processes of the life system that are collectively self-generating. There is arguably a distinctive sense in which these processes are 'part' of living systems. This account by no means solves all the problems with individuating systems that are inherently distributed and self-organizing, but it does make the case that a blunt assertion that all such distinctions are arbitrary is poorly justified.

(iii) Biological analysis starts with proper functions.

Another reason given for assigning systems approaches to function a secondary role is the claim that they are parasitic on proper function analysis. Thus, Millikan argues that analysis of proper functions comes 'logically first' in the life sciences (1993, p.150). Much of the argument given for this conclusion revolves around pointing out that scientists don't just focus on arbitrary chunks of matter, they pay attention to wider systemic relations. However this doesn't support the claim that in so doing scientists are always overtly or implicitly analyzing proper functions, only that they tend to pay attention to meso and macro causal patterns. Millikan presumes here the previous argument, namely, that any purely causal investigation must be entirely arbitrary in scope, but we have just given some reasons for thinking this isn't so. Moreover, even with respect to the analysis of functional relations the point is ambiguous as between the kinds of functional relations we are characterizing and proper functions. That is, scientists might be doing integrated systems analysis, and in fact we think they often are. So the conclusion that *proper* function analysis comes first begs the question.

⁵ The nature of what it is to be useful is one of the primary theoretical issues our account addresses.

But there is a deeper and completely fatal problem with Millikan's argument, which is that proper function analysis depends on causal analysis of ecological relations. As Stotz and Griffiths (in press) point out, whilst it is frequently assumed that any meaningful analysis of an organism's relations with its environment must classify the activities of the organism in terms of adaptive functions, this actually gets the explanatory relations back-to-front. A theory of the adaptive origins of a trait is generated by analyzing of the effect of the trait on the organism's ability to survive and reproduce in the ancestral environment in which it appeared. But this is a purely causal analysis of functionality, since by hypothesis the function the trait is performing has not been subject to selection at this point. In other words, evolutionary functions are ascribed on the basis of causal analyses of ancestral ecological relations. Consequently, there *must* be some way of performing causal analysis of function without prior knowledge of evolutionary function, because if there weren't then explanation of evolutionary function couldn't get started.

(iv) Selection etiology is the only possible naturalist explanation of norms.

A factor that lends artificial weight to the arguments against system function theories is the presumption of Wright's agenda for function theory. In particular, Wright's thesis that normative function equates with *having* a function. A characteristic argument that system functional analysis doesn't explain norms is that, although in a given case an item X without an evolutionary function might be doing some useful Y, X doesn't *have* the function of doing Y. So doing Y isn't normative for X. Therefore, there is no normative function.⁶

But this argument doesn't show the absence of functional norms or substantive functional issues. As we discussed above, there is a strong case to be made that understanding integrated adaptive systems is an important functional issue in its own right, one that doesn't collapse to design. And this raises what to us is a glaring hole in the argument, namely, that it ignores the issue of what it is for Y to be useful. Our approach differentiates between serving a function and having a function, and we claim that the former is also normative.⁷ In other words, we claim that Y bears a normative *relation* (specifically, usefulness) to the *system* S (not X). We are *not* thereby claiming that in virtue of serving a function X *has* a function. It is a further issue as to what constitutes having a function.

However, one kind of response to this is to claim that functional usefulness isn't in fact a normative relation. In ordinary language 'useful' is a value term, so the response is perhaps counterintuitive. But the apparent normativity might be an artifact of using a commonsense term to describe a relationship that is better translated into a scientific statement such as 'X contributes to fitness'. As such, the statement might be no more

⁶ E.g. Millikan's (1993, chapter 2) critique of the Bigelow and Pargetter (1987) propensity theory of functions uses this argument.

⁷ In fact, we believe it is a more fundamental normative relation. A trait has to perform a useful task before it can be selected for it.

inherently normative than the statement ‘friction in clouds contributes to the electrical discharge of lightning’.

Thus, a possible line of argument concedes (almost) all of the previous points we have made concerning treating the problem of understanding integrated adaptive systems as a substantive functional issue, whilst still insisting that functional *norms* are uniquely associated with design. In other words, normative biological function is specifically a matter of traits doing what they are supposed to do. It may be *important* to understand how adaptive systems work as integrated wholes, and to know what counts as a contribution to such a system, but these are different issues to the problem of understanding normative functions.

To address this kind of criticism it is necessary to examine more closely what it is to be a norm. Fundamentally, a norm is an evaluative standard or principle. That is, it must be possible that there be departures from the norm, and whether there is departure or not must have value from some perspective. The etiological explanation of norms identifies selection history as an evaluative standard: a trait conforms, or fails to conform, to its selection norm according as it does or does not do what the ancestor trait did that led to it being selected. This has been considered a notable achievement because it identifies a norm with a causally based condition, in the context of a widespread philosophical presumption that norms cannot be explained causally. Lurking in the background of the normative function theory literature are longstanding philosophical attitudes concerning the non-causal nature of norms shaped in part by debates concerning the fact/value distinction and the problems of psychologism.

Etiological proponents further claim that selection history is the only possible causal standard or benchmark for norms. But the argument for this hinges largely on the claim that all other forms of functional attribution are arbitrary, and as we have seen this claim is poorly justified. Indeed, the intuition that causal relations are essentially normatively arbitrary is exactly the anti-naturalist assumption that makes norms appear so problematic. Furthermore, if selection history can count as a norm it is arguable that functional usefulness can be considered normative on the same kind of grounds. Usefulness on our account concerns the relationship between a structure or process and the overall functional organization of an autonomous system. That is, a structure or process is useful if it contributes to the other processes that constitute the system. This definition of usefulness satisfies the criteria for being a norm outlined above. Departures from being useful are possible, since an item X can be either useful or not to an autonomous system S. And there is a perspective with respect to which usefulness has value, namely, the autonomy of the system.

At this point the crucial issue is whether autonomy is a suitable perspective for determining value relations. Why isn't it just a brute physical condition like lightning discharge? Our preliminary answer to this is that autonomy has a distinctive functional significance because of its relevance to understanding adaptive systems. Autonomous systems are a highly distinctive class of physical systems of key significance for biology. Moreover there are substantial dynamical and biological issues associated with

contribution to autonomy that we will discuss in the next section. So if functional relations can be normative at all then functional usefulness to an autonomous system should be considered normative. In this respect, note that the same question can be posed to selectionist etiology. In a rerun of Moore's naturalist fallacy argument, we can always ask of selection etiology, but why is conformity to *that* good? And there is no absolute sense in which it is good. Conformity with past selection is only 'good' with respect to a particular perspective, specifically, past selection. And selection history is only *important* because of its effect on adaptiveness. Thus, describing functional usefulness as normative is justifiable on the same type of grounds as claiming that selection etiology is normative. This is just a preliminary argument because it is important to know more about why particular natural conditions count as normatively relevant perspectives and others don't. We won't try to provide a general answer, but in the next section we will try to mount a case for considering autonomy as a normatively relevant perspective.

Autonomy⁸

Surveying the standard arguments against systems theories of normative function has allowed us to draw out some of the key issues that motivate our account. Specifically, there are promising ways of developing principled systems analyses of adaptive systems based on the underlying concept of cohesion, and such analyses should provide important insights into the process interdependencies involved in adaptively successful (normative) functioning. Moreover, as Stotz and Griffiths point out, there *must* be some coherent way of performing this kind of functional analysis because evolutionary proper function analysis depends on it.

Before we elaborate the details of our account of autonomy it is worth commenting on some of the distinctive features of the theoretical strategy we are pursuing. The general approach involves developing highly process-oriented concepts, theories and models, avoiding wherever possible the use of structurally defined and contingent physical entities as theoretical primitives.⁹ For example, whilst there are important roles for molecular gene concepts, it is better not to tie general evolution theory to a molecular gene concept because this prevents evolution theory from understanding processes of heritability that do not involve molecular genes. Dawkins' replicator concept seeks to generalize away from molecular genes by recursively defining genes as entities that play a role in a certain kind of evolutionary process, namely replication. Developmental Systems Theory (DST) adopts a different tack, rejecting the idea that evolution theory

⁸ The development of these ideas by the authors initially occurred independently, and we have subsequently combined our efforts to produce a unified approach. One of the legacies of this history is some difference in terminology. Bickhard developed a concept of self-maintenant systems (see Bickhard 1993), whilst Christensen developed a concept of autonomous systems that was subsequently elaborated in collaborative work with C.A. Hooker and J.D. Collier (see Christensen, Collier and Hooker 2000). Rosen's (1985) account of metabolic-repair systems, Ulanowicz's (1986) ecological concept of autonomy, and Collier's (1988) concept of cohesion influenced the formulation of Christensen's account of autonomy. There are minor differences between self-maintenance and autonomy that don't affect the discussion here; the choice of autonomy as the covering term in this paper is purely a matter of convenience.

⁹ Griesemer (forthcoming) discusses the merits of a process approach in biology at some length.

should require an essential link between genes and heritability at all (e.g. Griffiths and Gray 1994). Instead, the concept of *developmental processes* is substituted to encompass the diverse interactions that can be involved in generating traits that are heritable. From the DST perspective the trouble with the replicator gene concept is that it does not go far enough as a process generalization of evolution theory because it unnecessarily ties heritability to a particular kind of structurally defined entity. What matters to evolution is heritability per se, and any kind of developmental process involved in the construction of heritable traits should be part of evolutionary explanation.

Our purpose here is not to advocate DST in particular, but to illustrate the strategy and implications of shifting to process oriented theories on concepts in biology.¹⁰ It is possible that genes could be successfully defended as distinctive evolutionary entities by associating them with particular attributes of evolutionary processes, such as evolvability (Sterelny 2001). However it is also important from a wider biological perspective that a plurality of processes be recognized. The problem with focussing too narrowly on a particular process, such as evolutionary adaptation, is that contributing factors such as development tend to be idealized as static structures rather than treated as complex variables with distinctive characteristics. This in turn tends to oversimplify the interrelations amongst various contributing processes. Evolution of development research, for instance, opens up a large new theoretical terrain by explicitly recognizing a variety of processes and modeling their interactions in generating biological patterns. Thus, Raff (1996) argues that in sea urchins embryological, habitat, nutritional, population genetics and patterns of development are all interdependent. Population genetics shouldn't be privileged as the 'controlling' process, because, for example, ecological context affects the evolution of development, which in turn affects the population genetics of a species.

This kind of process pluralism makes a subtle but fundamentally important theoretical difference, because unlike most work in theoretical and philosophical biology we are not *defining* our concepts and models in terms of the evolutionary process. Specifically, we are not defining autonomous systems as evolutionary units, nor are we defining functions with respect to the evolutionary process, as Bigelow and Pargetter (1987) do by defining functions as contributions to fitness. Instead, we are treating the integrity and functional integration of biological systems as a distinctive process issue that is interrelated with that of evolution, not derived from it. The reasons for this are exactly the same as the reasons for not defining development recursively in terms of evolution. Development has its own distinctive process characteristics that contribute to evolution, and the same is true of the processes of functional integration in biological systems. The theory of autonomy can help to *explain* evolutionary units and evolutionary functions, but it is not recursively defined in those terms.

¹⁰ One of us (Bickhard) has argued at length elsewhere that quantum physics and considerations of general naturalism require a fundamental process metaphysics in which process is ontologically prior to structure. Whether or not one follows the issues all the way through to physics and general naturalism, however, there are strong empirical reasons for treating process as prior to structure in biology.

For a process approach, organisms are another type of contingent, structurally specified entity begging for process explanation. It is important to know how organisms and other kinds of adaptive systems 'hang together', and our theory of autonomy aims at developing a process explanation for this. On our account a system is autonomous if it interactively generates the conditions required for its existence. It is 'self-governed' in the sense that it shapes the conditions involved in maintaining its integrity as a system. In this respect, autonomy as we are using the word is more closely related to the political meaning of self-governance than to the more general sense of independence often used in philosophy and science. In dynamics a system is called autonomous if it has no interactions with external factors. However, autonomous systems in our sense are far-from-equilibrium systems that *must* interact with the environment. What makes them autonomous is that they exert a degree of influence on the conditions under which they exist. Thus, Switzerland is an autonomous nation, notwithstanding the fact that it interacts extensively with other nations. This concept of autonomy also has connections with the chemical concept of autocatalysis, which concerns chemical reactions that synthesize their own catalysts and thereby tend to be self-enhancing. Indeed, one way of viewing the concept of autonomy as we are using it is as a generalization of the concept of autocatalysis beyond strictly chemical relations.

There are two key aspects to autonomy: firstly, autonomous systems are cohesive in the sense that they interact with the environment as a causally integrated whole, secondly, the conditions required for the cohesion of the system are, at least partly, generated by the system itself through the capacity to perform work. Rocks are cohesive but they are not autonomous systems in our sense because the cohesion conditions (a stable molecular lattice) are sustained after the processes that generated the cohesion conditions of the rock (such as magma cooling) have ceased. Consequently, if a rock is damaged it won't perform work to reform itself. Moreover, if a rock is split at an arbitrary point the result is two smaller rocks with exactly the same type of cohesion conditions.

Now contrast a living cell, which is a paradigm autonomous system. Cells are far-from-equilibrium systems that require energy input from the environment, moreover they perform work to acquire this energy. In this respect the processes that generate the cohesion conditions of the cell are an ongoing part of the cell itself. The result is that cells *do* perform work to repair themselves. On the other hand, if the processes that perform the work are sufficiently disrupted the cell will die. This means that, in contrast with a rock, if a cell is split at an arbitrary point it usually won't result in two new cells. The processes that generate cellular cohesion stop and the cell collapses into its biochemical constituents. Autonomy involves positive feedback, but it is considerably more complex than simple feedback processes of the kind that plague public address (PA) systems. Positive feedback in a PA contributes to its own existence, but almost all of the work comes from the amplifier. In contrast, cells contain infrastructure that allows them to convert energy into useful forms and respond flexibly to perturbations. It is this infrastructure that makes them self-generating in a much more interesting way than PA feedback.

In addition to cells, other examples of autonomous systems include multicellular organisms, species, some ecosystems, and cities. We will discuss some of the complexities of understanding the diversity of autonomous systems and the interrelations amongst them below, but with respect to our current purposes the key point is that autonomous systems are the relevant class of systems for understanding normative functionality. Living organisms in general are autonomous systems, and autonomy is essentially a dynamical way of characterizing viability. Autonomy thus serves to provide the fundamental perspective for determining normativity: processes that contribute to the autonomy of the organism are useful to it.

As such, contributions to autonomy are the basic instances of *servicing a function*. This model of servicing a function is, first, dynamically significant: it makes a definite causal difference whether or not a particular far from equilibrium autonomous system continues to exist, or goes to equilibrium and thereby ceases to exist. Second, it is normative: such a contribution can succeed or fail in supporting the system, and this makes a distinct difference to the system, and to the world. The asymmetry of functional/dysfunctional is derived in this model from the fundamental physical asymmetry between far-from-equilibrium and equilibrium systems.

2 Autonomy and functional organization

Autonomous systems are composed of webs of interdependent processes. Modeling functional organization in our approach is a matter of characterizing these process interdependencies. To illustrate what this involves we shall consider a simplified case in which an autonomous system is composed of three processes: A, B and C. Because the system is autonomous, these three processes are involved in generating the conditions for the system's cohesion. Effectively this means that the conjoint operation of A, B and C is self-sustaining. Typically, A, B and C will be *organizationally interdependent* in the sense that each process requires output from the other two. This dependency will generally be tolerant within a range of variation, so a way of describing this interdependence is to say that for A to operate, B must be operating within a range of variation B_v and C must be operating within a range of variation C_v . Likewise for B and C. All of this, of course, will also depend on a set of environmental conditions $E_{(1...n)}$ with their own characteristic ranges.

To apply this form of analysis to a real biological example, consider the process interrelationships between heart activity, cellular metabolism, and motor activity (including breathing, eating and excreting). As everybody knows, the function of the heart is to pump blood, or more accurately to pump blood as a contribution to an ensemble of activities that result in blood circulation. The function that this serves, however, is to provide fluid transport for delivering nutrients to cells and removing metabolic end products. In this respect heart activity and cellular metabolism are interdependent processes. Without heart activity fluid transport stops, and with it cellular metabolism. And if cellular metabolism ceases then heart activity also ceases, and subsequently fluid transport. In addition to heartbeat, cellular activity also produces other motor action that contributes to interaction processes such as breathing, food acquisition,

eating and excreting. In turn these processes provide the resources required for cellular metabolism and expel waste products, thus contributing to the cellular processes that subservise them.

In our view these patterns of process interdependence in biological systems are the central issue for understanding normative function. Functions are essentially relations, and these process interdependency relations are what determine the nature of organisms as viable (cohesive) systems. Individual parts and processes serve normative functions within autonomous systems because of the way they satisfy the requirements of other processes within the system.

There is a range of additional functional issues that can be addressed using this approach. Here we will briefly discuss two: *resiliency* and *propagating dysfunction*. We can characterize resiliency in terms of the way the system responds to potentially disruptive perturbation. Assume that when process A approaches the lower bounds of activity required by B and C for some reason, B and C enter a brief phase of hyperactivity that tends to amplify A, pushing it back towards the activity range in which it adequately sustains B and C. In this situation the system exhibits a degree of resiliency, since a perturbation such as a local reduction in the availability of an environmentally sourced substrate of A can be compensated for. For example, if food resources become scarce a baboon troupe may increase the range and intensity of its foraging until new food resources are found.

Propagating dysfunction is the converse of resiliency. It may be the case that the burst of hyperactivity by B and C is insufficient to push A back within its range of required activity. Furthermore, the hyperactivity may have depleted reserves of substrates that B and C need. In this situation the system is likely to go into a spiral of decline as activity by B and C plummets, further reducing activity by A, in turn reducing activity by B and C. Thus, as the baboon troupe increases its foraging activity it increases its energy expenditure. If not enough additional food is acquired then the baboons will begin to weaken, *reducing* the amount of foraging they can engage in, and further reducing their energy intake.

In general, one of the most important aspects of autonomous systems analysis will be characterizing the patterns of resiliency and dysfunction dynamics to which the system is subject. In characterizing process interdependence we have for the sake of simplicity focussed on the situation where there is maximum interdependence, but patterns of interdependence will generally be complex and sometimes asymmetrical. In other words some processes will be more vital than other processes. Heart failure is lethal because it stops fluid transport and cellular metabolism, and *every* biological process depends on cellular metabolism. Mild vision impairment, on the other hand, has an immediate effect only on sensorimotor coordination. If the impairment isn't strong enough to compromise crucial sensorimotor processes such as food acquisition and predator detection, or if the animal can find ways to compensate, then there will be only limited flow-on effects to other processes.

In this respect it is significant to note that the concept of dysfunction employed by the etiological theory of functions makes no reference to the relative significance of the dysfunction to the organism. A trait is dysfunctional if it doesn't perform the task it was selected for, but this has no particular implications for the wider ramifications of the dysfunction for the organism. The dysfunction could range from being lethal to being a minor inconvenience. From a general conceptual point of view this should be troubling, because etiological theory claims to be the *only* explanation of dysfunction. Ipso facto it ought to explain every aspect of dysfunction. The fact that it doesn't explain the relative significance of dysfunction suggests at the very least that etiological theory is not satisfying the agenda it has set itself. From a more practical explanatory standpoint there are a number of reasons why understanding the relative significance of dysfunction is an important adaptive issue. It is important to understand the wider systemic implications of failure in order to understand whether and how the system can compensate. It is also important to know how the system can recognize failure as part of its compensatory abilities. These are surely important issues for understanding functional organization.

However it is also important to note that the concept of dysfunction suggested by autonomy is highly unusual, in the sense that the basic form of dysfunction is a process requirement that is not met. Thus, the fundamental dysfunction that occurs if the heart stops beating is that the fluid transport requirements for cellular metabolism are not met. If the process requirement can be met in some other way, such as with an artificial heart, then the autonomy-theoretic dysfunction goes away. Whether or not the heart is doing what it is supposed to be doing from an evolutionary design perspective is a further issue.

So our account of normative function does not furnish an intentional-design concept of proper function of the kind sought by proponents of the etiological account. As noted above, in this paper we will not address the issue of what it is for a part to *have* a function in this paper. In this area we think there is a role for an etiological account, although we also believe that the issue involves more factors than current etiological theories allow for.¹¹ Our criticisms have thus not been aimed at rejecting all forms of etiological analysis, but are rather concerned with using etiology as the fundamental basis for understanding normative function. Nevertheless the issue of having a function shouldn't be overemphasized. Treating it as the main question leads to focus on parts at the expense of understanding the interrelations between the parts and the rest of system. At the start we cited Aristotle's claim that "The hand separated from the body is not a true hand", and this has some genuine force on our account.

Notes on the diversity of autonomous systems

¹¹ Wimsatt (1972) provides a comprehensive and biologically informed philosophical analysis of selection based function that has been badly neglected. See also Wimsatt (1997). One suggestion for a non-selection factor in having function is degree of entrenchment. In other words, if an organism has a part that has outputs which supply one or more crucial process requirements then the functioning of those processes *presupposes* the presence of the part, and hence it 'has' that function. For an elaboration on such a notion of functional presupposition, see Bickhard (2000).

So far we have focussed on organisms as examples of autonomous systems. This has been useful because it allows us to clearly illustrate the central features of our account. However it is also important to examine some of the complexities of autonomy and some of the more marginal cases, because this helps to show how the principles of the theory can be applied and elaborated, and reveals more of their fine-grained structure.

In cases of highly autonomous systems the patterns of process interdependence are powerfully marked, although the couplings can be extremely complicated. As a result there will be relatively strong patterns of resiliency and dysfunction. Consequently, one kind of operational test for autonomy is to place an arbitrary slice through the system and see whether it collapses. If the system is autonomous the slice should cut through some of the process relations critical to the system's integrity. As we noted above, if you slice through a cell the result usually isn't two new cells. However, in the case of less strongly autonomous systems the couplings are weaker, and sometimes a relatively well directed slice may result in two new autonomous systems.

For these kinds of reasons it is important to examine some of the factors involved in the relative strength of autonomy, because these have implications for the diverse patterns of autonomy that can occur. As a first approximation, we can identify three dimensions to increases in autonomy: (i) increasing collective benefit, (ii) increasing dependency of component processes on collective activity, (iii) collectively imposed constraints on membership of the system.¹²

Free living unicellular organisms such as bacteria are perhaps the simplest kinds of autonomous systems to analyze. Their metabolic pathways show complex process interdependencies, the outcome of metabolic activity is the collective maintenance of those pathways, and there are relatively clear and strong constraints imposed by the organism on what is admitted to the system and what isn't. The cell membrane, for instance provides a highly selective barrier to ingress and egress, through the regulation of processes such as transmembrane protein channels, endocytosis and exocytosis.

Multicellular organisms introduce an added layer of complexity because the organism as a whole forms an autonomous supersystem composed of smaller, usually specialized, autonomous cells. One of the complexities multicellularity poses for autonomous systems analysis is understanding the trade-offs between the autonomy of the supersystem and the autonomy of its constituents. Multicellularity arises because groups of cells can achieve adaptive capacities beyond the ability of individual cells, through increasing size and articulated specialization. In very simple multicellular organisms, however, the collective benefits can be of a very simple aggregative kind, involving little or no specialisation. Seaweed, for example, consists of brown and red algae that live a multicellular existence, but show little cellular specialization into tissues and organs. Slime molds provide an

¹² We do not wish to give an unduly a priori account of the nature of these dimensions. There are likely to be many dimensions involved in autonomy, and their nature and precise formulation needs to be explored through detailed empirical investigation and further modeling. For this reason the analysis should be treated as a working approximation.

interesting intermediate example, in which slime mold cells spend part of their time as free living cells and part of their time in a multicellular phase.

When the degree of collective benefit is relatively limited the adaptive advantages of specialization for the cellular constituents are proportionately low, and conversely, because there is little specialization, the relative autonomy of the supersystem is weak. In these cases it may be relatively easy to divide the system into several new systems that are also autonomous in the same way. For example, it is comparatively easy to grow new plants from plant cuttings. On the other hand, when the collective benefits are very strong it typically requires a high level of specialization on the part of the system's constituents. The adaptive commitment of the cells may then hinge much more completely on the success of the supersystem as a whole, and it is likely to be much harder to divide the system and retain any autonomous units. The cells of higher animals can be kept alive in laboratory conditions, but they are quite incapable of sustaining a free-living mode. Moreover it is virtually impossible to propagate a new human by slicing off the finger of an old human, other than, perhaps, with extremely sophisticated cloning technology that is not yet available.

The same pattern of tradeoffs occurs in the evolution of sociality. Groups of conspecifics can be better able than individuals to control resources like food and habitat, and detect and defend against predators. The degree and nature of social specialization can vary quite markedly, however. In the simplest cases the collective benefits might be primarily aggregative, such as in a large bird colony, where many eyes provide greater chance of predator detection, and the movement of many individuals tends to confuse predators. There may be comparatively little specialized activity on the part of individual birds that promotes collective organization. Moreover, unlike cells, where there is complicated and sophisticated regulation of admission to the system, there may be little regulation of admission to a bird colony. This may be limited to driving birds of other species away. But admission of members of the species to the colony may be limited only by the availability of nesting sites. Consequently, most bird colonies are likely to be extremely marginal autonomous systems.

In other cases of social organization, however, there can be significantly greater specialization and interdependence. In this respect it is important to distinguish between specialization that *promotes* collective activity and specialization that allows individuals to cope with and utilize collective activity. The complex predator calls of vervet monkeys are clearly activities that primarily promote collective activity. On the other hand, adaptations to live in dominance hierarchies may primarily improve the ability the individual to function adaptively *within* the collective. Of course, as animals become increasingly specialized to live in dominance hierarchies, the size of groups that can be maintained without excessive conflict rises, thereby indirectly increasing collective benefit. It may also be the case that the dominant individual functions to coordinate collective activity in ways that are of direct collective benefit, such as where the leader of a baboon troupe determines the location of foraging activity. In humans the degree of cooperative social interaction very strongly dominates the adaptiveness of individuals. We are an exceptionally social species and we are correspondingly highly dependent on

social interaction for survival. Consequently, we are the only species that tends to form autonomous social supersystems with anything like the complexity and specialization of multicellularity.

There are further important issues that we cannot examine for reasons of space, including whether and how ecosystems and lineages can be autonomous. We finish the discussion by arguing that the fact that there are very marginal cases of autonomy is not a problem for the approach, in fact it is an advantage. At the lower bounds of autonomy the system will be 'just barely' autonomous. Such marginal cases will be difficult to diagnose in practice, and pose strong tests for the interpretation of the principles of diagnosis. For instance, although ecosystems fairly clearly have the properties of collective benefit and process interdependency, assessing the extent to which they impose collective constraints on membership is more difficult, both from a theoretical and practical perspective. Determining what counts as 'collective constraint on membership' for an ecosystem is not a simple matter: is outright exclusion of some species required, or only some modification of species to enhance ecosystem integrity? If the latter, how much modification is required to call the ecosystem autonomous? We don't propose to offer determinate a priori answer to these kinds of questions, instead, what we hope to do is raise the issues and outline ways of meaningfully developing answers.

In very marginal cases of autonomy the system's identity will be so tenuous that speaking of benefit to the system, let alone 'serving a function' for the system, may appear overly grandiose, and jar on intuitions built up around central examples. However recognizing marginal cases is not problematic so long as the dimensions that determine increasing centrality can be reasonably clearly specified. Moreover, recognizing marginal cases has strong theoretical advantages, not least because it allows us to understand how transitions occur.

Conclusion

We have indicated some of the factors involved in autonomous systems analysis, hopefully in enough detail to give a sense of how it can be applied, how the autonomy of a given system can be evaluated, and the adaptive issues it helps to illuminate. The concept of autonomy focuses on the process characteristics of functional integration and system integrity. This provides us with a basic normative functional concept of serving a function for an autonomous system, and it also allows us to understand important adaptive issues such as resiliency and propagating dysfunction.

The openness of our account may be worrisome for some people, in the sense that it ascribes autonomy and functionality in cases that are counterintuitive. However the account does make principled distinctions, including with respect to determining system boundaries, and in this respect it is entirely coherent. Moreover there are major advantages to adopting an open dynamical process approach to characterizing biological systems. One of the problems with etiological and statistical attempts to define systems, such as in terms of clusters of etiological functions or 'species-typical' traits, is that they are static definitions that apply, at best, for relatively short periods of time, until change

occurs and there is a mysterious quantum transition in the nature of the system. Static definitions of species, for instance, are fine so long as phylogenetic trajectories are straight non-branching lines. In the terminology of Lewontin (1969), they presuppose an equilibrium conception of biology. However, transformational approaches to biology focus on understanding change: for example, a transformational approach might examine questions such as whether a segmented body plan imposes constraints on subsequent evolution, or, in the case of Raff (1996), what the effects of developmental constraints are on evolutionary modification of early, middle and late development.¹³ Consequently, there is a need in this form of analysis to track systems through transformations. For transformational biology, therefore, developing dynamical ways of characterizing systems is imperative.

Thus, whilst our account is in its most basic form a-historical, it actually allows for more thoroughly historical forms of analysis than does the etiological approach. It models biological systems directly as integrated systems in a way that is at least partly independent of the mechanisms responsible for producing the organization. As a result, it then becomes possible to examine the relative effects that various factors can have to induce changes to the system, thus making possible a more sophisticated account of constraints and evolutionary mechanisms.¹⁴ With respect to these kinds of issues, the problem with the etiological account is that it is not historical enough.

¹³ For a philosophical discussion of some of the key issues associated with transformational approaches in biology see Griffiths 1996.

¹⁴ Cf. Amundson and Lauder (1998), who also argue for the utility of 'causal-role' functional analysis for transformational

Bibliography

Amundson, R. and Lauder, G.V. (1998). Function Without Purpose: The Uses of Causal Role Function in Evolutionary Biology. In Hull, D. and Ruse, M. (eds.) *The Philosophy of Biology*. Oxford: Oxford University Press.

Bickhard, M.H. (1993). Representational Content in Humans and Machines. *Journal of Experimental and Theoretical Artificial Intelligence* **5**: 285-333.

Bickhard, M. H. (2000). Autonomy, Function, and Representation. *Communication and Cognition — Artificial Intelligence*, 17(3-4), 111-131.

Bigelow, J. and Pargetter, R. (1987). Functions. *The Journal of Philosophy* **84**: 181-196.

Christensen, W., Collier, J.D., and Hooker, C.A., (2000). Autonomy. In Christensen, *An Interactivist-Constructivist Approach to Adaptive Intelligence and Agency*. PhD thesis. Available at:
<http://www.newcastle.edu.au/departement/pl/staff/WayneChristensen/dissertation.htm>

Collier, J.D. (1988). Supervenience and reduction in biological hierarchies. *Canadian Journal of Philosophy*, supplementary volume: Biology and Philosophy, Calgary: University of Calgary Press.

Cummins, R. (1984). Functional Analysis. In E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge: MIT Press.

Depew, D. and Weber, B. (1995). *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection*. Cambridge MA: MIT Press.

Griesemer, J. (forthcoming). The Informational Gene and the Substantial Body: On the Generalization of Evolutionary Theory by Abstraction.

Griffiths, P. (1996). The historical turn in the study of adaptation. *British Journal for the Philosophy of Science* **47**: 511-32.

Griffiths, P. and R. Gray (1994). Developmental Systems and Evolutionary Explanation. *Journal of Philosophy* **91**: 277-304.

Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposium on Quantitative Biology* **22**: 415-427.

Lewontin, R.C. (1969). The Bases of Conflict in Biological Explanation. *Journal of the History of Biology* **2**: 35-45.

- Millikan, R.G. (1989). In Defense of Proper Functions. *Philosophy of Science* **56**: 288-302.
- Millikan, R.G. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA.: MIT Press.
- Millikan, R.G. (1999). Wings, Spoons, Pills and Quills. *Journal of Philosophy* **96**(4): 191-218.
- Raff, R. (1996). *The Shape of Life: Genes, Development, and the Evolution of Animal Form*, Chicago: University of Chicago Press.
- Rosen, R. (1985). Organisms as Causal Systems Which Are Not Mechanisms: An Essay into the Nature of Complexity. In R. Rosen (ed.), *Theoretical Biology and Complexity*, London: Academic Press, Inc.
- Sober, E. (1993). *Philosophy of Biology*. Boulder: Westview Press.
- Sterelny, K. (2001). Niche Construction, Developmental Systems and the Extended Replicator. To appear in: Russell Gray, Paul Griffiths and Susan Oyama (eds) *Cycles of Contingency*, MIT Press, 2001.
- Ulanowicz, R. (1986). *Growth and development: ecosystems phenomenology*. New York: Springer.
- Wimsatt, W.C. (1972). Teleology and the Logical Structure of Function Statements. *Studies in the History and Philosophy of Science* **3**: 1-80.
- Wimsatt, W.C. (1997). Functional Organization, Functional Analogy, and Functional Inference. *Evolution and Cognition* **3**(2): 102-132.
- Wimsatt, W. C. and J. C. Schank (1987). Generative entrenchment and evolution. In Fine, A. and Machamer, P. (eds), *PSA 1986*. East Lansing, Mich.: Philosophy of Science Association.
- Wright, L. (1973). Functions. *Philosophical Review* **82**: 139-68.