Agent-Based Modelling and the Environmental Complexity Thesis

Anil K Seth
The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA
seth@nsi.edu

Abstract

A variation of Godfrey-Smith's 'environmental complexity thesis' is described which draws together two broad themes; the relation of functional properties of behaviour to environmental structure, and the distinction between behavioural and mechanistic levels of description. The specific idea defended here is that *behavioural and/or mechanistic complexity can be understood in terms of mediating well-adapted responses to environmental variability.* Particular attention is paid to the value of agent-based modelling within this framework.

1. Introduction

"The function of cognition is to enable the agent to deal with environmental complexity". So states, with admirable brevity, the 'environmental complexity thesis' (ECT) of Peter Godfrey-Smith, a philosophical project which first appeared in his 1996 book, *Complexity and the Function of Mind in Nature,* and which has subsequently been the focus of considerable and largely positive attention (for example Hardcastle, 1999; McShea, 1996; Oyama, 1996; Bedau, 1996; Belew, 1996).

My aim in this paper is to use a discussion of the ECT as a way of drawing together two broad themes: the relation of functional properties of behaviour to environmental structure (central to the ECT as described by Godfrey-Smith), and the distinction between behavioural and mechanistic levels of description (*not* so central). The motivation for this is to describe a framework within which the relations between behaviour, mechanism, and environment, as they appear in agent-environment systems in general, can be usefully articulated. I shall also give examples throughout of how agent-based modelling techniques complement this framework in advancing our understanding of adaptive behaviour.

The upshot of all this will be a new interpretation of the ECT, which is that *behavioural and/or mechanistic complexity can be understood in terms of mediating well-adapted responses to environmental variability.* Undoubtedly this is similar to Godfrey-Smith's own formulation, but there are differences, the significance of which will be illustrated by critiquing the original ECT from three directions: (i) the use, and meaning, of the term 'cognition', (ii) the relation of the ECT to W. Ross Ashby's 'law of requisite variety', and (iii) the role played by agents in the 'construction' of their environments. Whilst distinct in substance, these three elements share a common interest, unlike Godfrey-Smith's original thesis, in distinguishing between behavioural and mechanistic levels of description. It is nevertheless important to emphasise that the purpose of this paper is not to try to undermine the ECT *per se,* but rather to exploit some of its many riches.

2. Setting the stage

We begin with some definitions, which may be regarded as assumptions of this paper, and with a discussion of the distinction between behaviour and mechanism.

Behaviour is defined in this paper as *observed ongoing agent-environment interactivity,* and *mechanism* as the *agent-side structure subserving this interactivity.* Defined this way, all behaviours (eating, swimming, fleeing, building-a-house) depend on continuous patterns of interaction between agent and environment; there can be no eating without food, no building-a-house without bricks, no swimming without water. In addition, it is ultimately up to the external observer to decide which segments of agent-environment interactivity warrant which behavioural labels. In principle, different observers may both privilege different junctures in observed activity and label the same segments differently, either way causing problems for any explanation of mechanism framed in the language of behaviour. Behaviour is a product of the joint activity of agent, environment, and observer, therefore - and this is the crucial point - the (agent-side) mechanisms underlying the generation of any behaviour should not be assumed to be identical to the behaviour itself. This is what is meant by distinguishing between behavioural and mechanistic levels of description.

An instructive example of the importance of this distinction comes from Lorenz's (1937) observations of the 'parenting behaviour' of ducks (see also Hendriks-

Jansen, 1996). Mother ducks engage in a number of different patterns of interaction with their offspring, and Lorenz subsumed all such observed patterns under the label of 'parenting behaviour'. This is clearly valuable for descriptive classification, but it should *not* be taken as reason to believe in the existence of any 'behavioural icon' for parenting behaviour, internal to the mother duck, as a mechanistic explanatory locus. And indeed, as Lorenz subsequently discovered, from the perspective of the mother duck the various interaction patterns are all triggered by quite different stimuli. The only point at which they 'intersect' in any sense is on the duckling itself, as an object in the eyes of the external observer.

Moreover, if a behaviour appears *complex* (to an external observer), this does *not* imply that the underlying internal mechanisms are also complex. The classic illustration of this is Herbert Simon's description of an ant on a beach (Simon, 1988). The internal mechanism of this (hypothetical) ant consists of a simple obstacle-avoidance rule - if there is a clump of sand to the left, go right, and vice-versa. Thus the ant responds to every tiny clump of sand, veering first left then right as it negotiates its terrain. Simon's point is that from the perspective of the external observer - who cannot perceive the small-scale heterogeneity of the beach surface, and who is not aware of the simplicity of the ant's 'algorithm' - the trajectory traced by the ant is strikingly and perhaps irretrievably complex.

So how can one begin to trace the relationship between behaviour and mechanism? A first step is to be clear about the meaning of environment, which we may think of in two ways: the environment as it appears to us, as external observers, and the environment as it appears to the agent that we (as external observers) are observing. The former (the 'external' environment, following Brandon, 1990, or 'distal' environment, following Brunswik, 1952, and Nolfi, 1998) is that which features in the definition of behaviour as agent-environment interactivity, and is perhaps the most intuitive way to understand the term. The latter is perhaps best labelled by the term *Umwelt* ('proximal' will also do), coined by Jakob von Uexküll (1934) to refer to the space of sensorimotor cues relevant to an organism, containing those features which constitute stimuli for the organism, to which the organism can potentially muster a response.[1] Consider, then, that behaviours, being elements in the vocabulary of an observer, are located in the external environment. In the case of Lorenz's duck, the descriptions of the various interaction patterns that constitute parenting behaviour are framed in terms of the external environment, and intersect only on the duckling *as an entity in the external environment.* The relation of these behavioural descriptions to mechanism may then require that the relations between the external environment and the *Umwelt* be traced - Lorenz had to identify what constituted stimuli for the mother-duck in order to understand the nature of its parenting behaviour.[2]

[1]See Ziemke and Sharkey (2001) for an interesting commentary on von Uexküll and recent developments in adaptive robotics.

3. The environmental complexity thesis

We may now move on to the second of our primary themes, the relation between behaviour and environmental structure. This is the province of the ECT.

"The function of cognition is to enable the agent to deal with environmental complexity". A strong claim, and one which Godfrey-Smith is keen to place in the context of what he labels 'externalism', the explanation of internal organism properties in terms of their relations to the external. In fact Godfrey-Smith describes the ECT as expressing 'c-externalism', which he defines as the attempt to understand internal *complexity* in terms of external *complexity.* The intuition is that 'cognition' is more complex than 'no cognition', thus 'cognition' is argued to be a complex response to a complex environment. Importantly the ECT is also a claim about necessity, not sufficiency. The idea is that environmental complexity is necessary for there to be cognition, but *not* that environmental complexity will always result in there being cognition.

A wider aim that Godfrey-Smith holds for his book is the exploration of externalist explanation in general, and it will repay us to understand something of this aim. In this broad context the ECT stands along with adaptationist explanations in biology, empiricism in philosophy, and associationism and behaviourist learning theory in psychology. The obvious (and equally influential) complement of externalism is internalism, the attempt to explain internal organism properties in terms of (their relation to) other internal properties. Biological internalism is manifest in the structuralist accounts of Goodwin (1994), and, earlier, d'Arcy Thompson (1917). Psychological internalism is strident in Chomsky, and in philosophy the rationalist tradition is internalist insofar as it argues for the necessity of mental 'pre-structure' in the formation of beliefs and judgements.

Of course any distinction between internalism and externalism naturally presupposes a distinction between the external and the internal, yet as many have argued this can be difficult to justify, and theories which take as their explanandum the *existence* (whether apparent or 'real') of a distinction between external and internal naturally resist classification as either externalist or internalist (see, for example, Oyama, 1996, 1985). In view of this it can be argued that the ECT ought to ex-

[2]To take another example, the catching behaviour of cricketers has been greatly elucidated by appreciating as a stimulus the acceleration of the tangent of elevation of gaze from player to ball; if this acceleration is kept at zero, the cricketer *will* meet the ball before the ball meets the ground (McLeod & Dienes, 1996).

press a 'pragmatic externalism', retaining the emphasis of accounting for internal properties in terms of external properties, but at the same time displaying consistency with arbitrary boundaries between external and internal. Importantly, not all versions of the ECT are pragmatically externalist in this sense.

Getting back to the ECT itself, Godfrey-Smith is quick to point out that its plausibility rests entirely on the meanings ascribed to the terms 'function', 'cognition', and 'environmental complexity'. The vision of *environmental complexity* held by Godfrey-Smith is one of *heterogeneity*:

> Complexity is changeability, variability; having a lot of different states or modes, or doing a lot of different things. Something is simple when it is all the same. (1996a., p.24)

This definition of complexity is not without its difficulties, and in particular bears a controversial relationship with unpredictability (McShea, 1996), but for present purposes we may let it stand. 'Function' is interpreted 'teleonomically'; the teleonomic function of something is the effect it has which explains why it is there, usually in view of some selective process, so from this vantage cognition is understood to be an *adaptation* to the problem of environmental complexity, the ECT an adaptationist thesis. Finally, what Godfrey-Smith means by 'cognition' is, unfortunately, less clear, despite being that which the ECT is ostensibly directed towards accounting for. We will have more to say about this in section 4.1.

Godfrey-Smith also identifies an informative historical context for the ECT which takes the form of a contrast between the positions of the philosophers Herbert Spencer and John Dewey. Spencer believed in a general 'law of evolution' which prescribed a universal dynamic from states of "indefinite, incoherent homogeneity" to states of "definite, coherent heterogeneity". Increases in environmental heterogeneity were supposed to lead to corresponding increases in internal heterogeneity by "the continuous adjustment of internal relations to external relations". Godfrey-Smith interprets this as suggesting that "organic properties are not seen so much as solutions to environmental problems but rather as bearing the imprint of the environment's pattern" (ibid., p.89). There is therefore no sense, for Godfrey-Smith, in which a Spencerian ECT implies that cognition 'deals with' environmental complexity by way of judiciously deployed responses, rather, organism 'accommodates' environment in the sense of the external somehow being captured 'inside' the internal. Interpreted this way, the Spencerian ECT is endlessly problematic. Most seriously, the idea of organic properties 'accommodating' environmental properties demands a reification of the distinction between external and internal, one cannot bring the external within the confines of the internal without asserting a pre-existing distinction between the two. Because of this the Spencerian ECT is inconsistent with pragmatic externalism.

Dewey receives greater favour from Godfrey-Smith, and it is in his later work - most particularly in *Experience and Nature* (1929) - that a version of the ECT is identified. For example:

> The world must actually be such as to generate ignorance and inquiry; doubt and hypothesis, trial and temporal conclusions; [...] The ultimate evidence of genuine hazard, contingency, irregularity and indeterminedness in nature is thus found in the occurrence of thinking. (1929, p.69, quoted in Godfrey-Smith, 1996a, p.100)

It is Dewey's pragmatism that makes him more acceptable. As Godfrey-Smith explains, "Dewey was opposed to the idea that in solving problems organic systems merely accommodate environmental demands. Rather, they intervene in the environmental processes which generated the problem, and alter the environment's intrinsic course" (ibid., p.139).

Godfrey-Smith's reading of this version of the ECT improves on Spencer's version in at least two ways. First, consistency with pragmatic externalism is ensured because the boundary between external and internal can be retained as arbitrary (to the extent that the response is *not* framed in terms of the environment itself, otherwise one is once again forced to 'internalise' the environment). Second, Dewey helps sharpen our ideas about the kinds of environment that might provide adaptive advantage for cognition, focusing on a mixture of predictability and unpredictability:

> The incomplete and the uncertain give point and application to the ascertainment of regular relations and orders. (1929, p.160, quoted in Godfrey-Smith, 1996a, p.130)

Godfrey-Smith restates this idea in the form of two conditions to be satisfied if a role for cognition is to be mandated. First, that there exists *variability* "with respect to distal conditions that make a difference to the organisms's well-being". Second, that there be *stability* "with respect to the relations between these distal conditions and proximal and observable conditions" (ibid., p.118). We shall return to these conditions in section 4.2.

In summary, we may say that Godfrey-Smith favours an adaptationist, teleonomic version of the ECT in which complexity is interpreted as heterogeneity, in which the environment is seen to comprise of a mix of stability and variability, and in which cognition is interpreted as response rather than accommodation. As such, the ECT provides a clear way to think about how the functional properties of behaviour relate to environmental structure: they cope with environmental complexity.

4. Critiquing the ECT

Whilst both the ECT and the behaviour/mechanism distinction make good sense when considered individually, the picture they paint when considered together is hardly an integrated whole. Our efforts are now in this direction in the form of a critique of Godfrey-Smith's version of the ECT. The first element of this critique concerns 'cognition'.

4.1 Cognition

Without a clear understanding of what Godfrey-Smith means by cognition it is not possible to understand what the ECT is trying to explain. Is it just 'response' and nothing more? A good place to begin is the Precis of his book:

> A core set of cognitive capacities, including the capacities for perception, internal representation of the world, memory, and decision making, have the function of making possible complex patterns of behavior that enable organisms to deal effectively with complex patterns and conditions in their environments. (1996b, p. 453)

Elsewhere throughout the book itself, however, ambiguity returns, with cognition discussed at a variety of levels of abstraction and in several different senses (a worry partly shared by Belew, 1996). In one place it is interpreted "as a means to the production of behavioural complexity" (1996a, p.26), in another, the possibility of cognition being an advanced kind of homeostatic device is advanced (ibid., pp.76-79), and there is also reference to a "basic mental tool-kit" (ibid., p.127). This variety makes it clear that although the ECT may provide a detailed discussion of how the functional properties of behaviour may relate to environmental structure, there will be less to say about how such behaviour is to be related to underlying mechanisms. For Godfrey-Smith, both levels of description are subsumed under the term 'cognition', a term which he loads with many of its customary connotations such as internal representation and decision making (cf. Neisser, 1967).

This dalliance with what one may call 'classical' cognitive terminology leads his version of the ECT into some unfortunate territory. Shapiro (1999), for example, gets things very much the wrong way around, suggesting that the ECT may be used to answer the question of whether the observation of a particular behaviour warrants the ascription of *psychological* mechanism, rather than just *physiological* mechanism. If there is sufficient environmental complexity, his argument goes, then psychological (read 'cognitive') mechanism must be involved, the implication being that physiological (read 'non-cognitive') mechanism just could not cope in such circumstances). Unfortunately, in proposing this argument, Shapiro succeeds only in reviving a prejudice effectively dispelled by Pavlov in the 1920s; that any process analysable in physiological terms *ipso facto* cannot be a psychological process (see Lorenz, 1948, p.204).[3]

This, then, is the first element of my critique of Godfrey-Smith's formulation of the ECT. Slippery usage of the term 'cognition' enables him to gloss over the distinction between behaviour and mechanism. The functional properties of behaviour are the proper targets of the adaptationist claims of the ECT, mechanisms are certainly required to subserve behaviour, and so mechanisms will also be indirectly subject to the same forces of selection that operate directly on the functional entities of behaviour, but there is no necessary, direct link between complex behaviour and complex mechanism, cognitive or otherwise (recall Simon's ant). Therefore, there is no way the ECT should be used to justify claims for the existence of 'psychological mechanism' over 'physiological mechanism'. In other words, part of the appeal of the ECT comes from the apparent link between environmental structure and internal mechanism, yet this misses out the essential link provided by behaviour. Part of the difficulty with Godfrey-Smith's formulation is that the term 'cognition' obscures any clear distinction between behaviour and mechanism, rendering any direct implications of the ECT for internal mechanism unclear.

There are some straightforward implications for agent based modelling here, quite simply that such models allow distinctions between behaviour and mechanism to be fully realised and empirically interrogated, obviating the need for problematic catch-all terms such as cognition. In any but the most trivial concrete model there will be both observable behaviour and analysable behaviour-generating mechanisms. Causal networks linking behaviour, mechanism, and environment can in principle be fully unravelled, and the explanatory leverage of the ECT in accounting for *mechanistic* complexity empirically assessed.

To take a concrete example, consider Beer's work on the evolution of 'minimally cognitive' behaviour (Beer, 1996, 2000). Beer analysed an agent equipped with seven sensory 'rays', set at various angles relative to the midline, reporting the distance to the edges of a falling object, either a circle or a diamond. The agent could move along a horizontal line and was faced with the task of avoiding diamonds and centering circles. Beer used a genetic algorithm to evolve successful agents, con-

[3] Dennett (1987) makes a similar point, arguing that behaviour of a sufficient apparent complexity drives observers to take the 'intentional stance' with respect to the behaving agent, such that in order to understand and/or predict the behaviour of the agent it is necessary to (or at least it *helps* to) ascribe 'classical' cognitive states to the agent. However, as Dennett is careful to emphasise, the act of taking the intentional stance does *not* imply that the actual agent-side behaviour-generating mechanisms are cognitive in this sense; the intentional stance is an *observational* strategy.

trolled by recurrent neural networks, and then asked how they worked. Without going into details, he observed that "subtle interplay between sensory input and internal state is crucial to accurate discrimination" (Beer, 2000, p.94), thus accounting for an arguably 'cognitive' phenomenon in terms of dynamic interrelations between behaviour, mechanism, and environment. A second example which more directly probes the relationship between environmental structure and mechanism complexity is provided by Biró and Ziemke (1998), who analysed simple evolved neural networks controlling agents able to locate objects within a walled arena. They noticed that the dynamics of successful networks could often be analysed in terms of clusters of activity that correlated well with distinct behavioural segments (search, tracking, avoidance), in this case arguing for a simple relationship between behaviour and mechanism (see also Ziemke, 2000). One can easily imagine a productive extension of this work, challenging this simple relationship by varying both the structure of the environment and the sensory mapping from the environment to the controlling network.

4.2 The law of requisite variety

The second element of the critique analyses the ECT in terms of Ashby's (1956) 'law of requisite variety' (henceforth LRV). A product of the mid-twentieth century cybernetic school, the LRV may be considered intuitively as asserting that only agent-side variation can cope with - or 'force down' - environmental variation (as opposed to, for example, agent-side *stasis*) . This is clearly similar to the basic premise of the ECT, however although the LRV is mentioned by Godfrey-Smith it is never for him a focus of attention. This I argue is a missed opportunity for the reason that the LRV can help the ECT bring more clearly into focus the important concepts of variability and homeostasis.

Let's start with a more formal account of the LRV. Consider that for an agent to maintain relative stability in certain (internal) essential variables (for example heart rate, body core temperature), it must *prevent* the transmission of environmental variability through to these essential variables. In the same way that a good thermostat prevents the transmission of environmental variations in temperature through to a particular object (for example, the interior of a refrigerator should remain at a constant cool temperature despite the fluctuating temperature of a kitchen on a midsummer day), a well adapted agent prevents the transmission of certain environmental variables (for example the prevalence or scarcity of food, the proximity, or otherwise, of predators) through to such essential internal variables as blood sugar level or heart rate. With this in mind, the LRV can be easily formulated: Consider a set of possible environmental disturbances D, a set of possible responses on the part of the agent, R, and a set of possible outcomes, O. Consider also that for each D_i, there is distinct outcome O_i, and a particular response R_i. Stability in the essential variables requires minimising the variation in O, and this then requires that the variety in D is matched by the variety in R. Ashby himself provides a more concise summary:

> If R's move is unvarying, then the variety in outcomes will be as large as the variety in D's moves; only variety in R's moves can force down variety in the outcomes.[4] (1956, p.206)

So how can the LRV express a version of this? Let's rehearse once again: "the function of cognition is to enable the agent to deal with environmental complexity".

Most obviously, environmental complexity may be interpreted as the set of environmental disturbances D. 'Function' can be interpreted, teleonomically, as the maintenance of stability in essential variables (the minimisation of variation in O). This leaves 'cognition', which instead of being associated with the variety of interpretations favoured by Godfrey-Smith, is here related directly to the regulation effected by the various responses (the set R) deployed by the agent in 'forcing down the variety' in O.

In this way the LRV expresses a quite specific version of the ECT, one with several interesting properties. First, there is a clear emphasis on environmental (distal) variability, which for Ashby is significant in two distinct manifestations:

> There is that which threatens the survival of the gene-pattern. This part must be blocked at all costs. And there is that which, while it may threaten the gene-pattern, can be transformed (or re-coded) through the regulator R and used to block the effect of the remainder (1956, p.212).

This may be usefully compared with Godfrey-Smith's dual condition for the adaptive significance of cognition. For Godfrey-Smith, without distal variability, cognition is not necessary, and without stability in distal-proximal relations, cognition is not possible. Ashby offers a third

[4]A small qualification is worthwhile. It is not usually necessary for essential variables to be maintained at a precise value; variation within a certain range is usually permissible and often inevitable. In mammals, for example, changes in heart rate and blood sugar concentration are in fact essential contributions to adaptive behaviour, for example in ensuring physiological preparedness for 'fight' or 'flight' responses. McFarland and Sibly (1975) speak of a physiological state-space, with essential variables defining axes in this space such that its dimensionality is determined by the number of essential variables. It would be more accurate to interpret the LRV in terms of maintaining the state of the organism *not* at a single point in physiological state-space, but rather in a region of this space circumscribed by a set of 'lethal boundaries'. However, for the present purposes this distinction may be overlooked; I shall be concerned with a broadly construed stability in essential variables only.

alternative, identifying the importance of 'potentially beneficial' variability, the suggestion being that without this kind of variability, the agent would not be able to act - or respond - at all. One example of this kind of variability would be a change in odour intensity that correlates with proximity to a food source, another would be some kind of environmental stochasticity that enables protean behaviour in a predator-prey situation. The perspectives of Ashby and Godfrey-Smith are therefore complementary, and an interpretation of the ECT emerges in which all three characteristics combine: stability in distal-proximal relations, and distal variability that both threatens the agent and facilitates its activity.

Secondly, by framing the components of the ECT in terms of potentially *measurable* variety - in environment, response, and outcome, - the LRV emphasises the value of concrete agent-based modelling. Fletcher, Zwick, and Bedau (1998, 1996), for example, investigate how the manipulation of environmental texture in a toroidal grid-world environment (interpolating between flat, random, and sinusoidal resource distributions) relates to the variety of responses deployed by well-adapted agents. They employ Shannon entropy to track three distinct measures of variety: the information content of the environment (from the perspective of the agent), the information content of look-up tables of sensorimotor rules that constitute agent mechanism (which represents the variety of response), and the *between-agent* variety in look-up table structure. Although they do not mention the ECT, they take two results to exemplify the LRV. First, that response variety approaches environment information content; thus variety in R is matching variety in D. Second, that between-agent variety falls to zero, indicating that the variety in R is not random, such that for every D_i there is indeed a particular response R_i.

In a second example, Seth (in press) explores an evolutionary iterated prisoner's dilemma (IPD) model, focusing on the relationship between the introduction of noise at various loci in the model and the evolution of strategy memory (which can informally be thought of as strategy complexity). Analysis in terms of the LRV predicts which loci promote the evolution of memory when noise is applied, and which do not, enabling a distinction to be drawn between *adaptive* and *non-adaptive* evolution of memory, without reference to fitness statistics.

It is worth considering why the LRV was given such short shrift by Godfrey-Smith; one possible reason is the suspicion that the LRV implies that cognition be understood as *homeostasis,* and this is an implication that Godfrey-Smith tries hard to avoid (pp.76-79). Ever since Cannon (1932), the term 'homeostasis' has been used to describe mechanisms in which stasis in certain properties of a mechanism is maintained by variation in others. Cannon originally discussed homeostasis in a physiological context, but it is clear that the notion generalises, and the LRV certainly has a homeostatic interpretation in terms of the maintenance of stability in essential variables (in whatever context) through the deployment of appropriate responses. So let us briefly consider the question: can 'cognition' be construed as homeostasis?

For Godfrey-Smith, the answer is 'sometimes'. In some cases, cognition will lead to actions that are genuinely homeostatic, for example in the intelligent use of fire to maintain bodily warmth. 'Genuine' homeostasis, for Godfrey-Smith, obtains when "there is some intermediate organic property [body temperature] such that complex activity contributes to the maintenance of stasis in this intermediate property, where this property makes a real contribution to survival" (1996a, p.79). However, in other cases there will be no non-trivial homeostatic interpretation of the action of cognition, for example in the adept use of perception and coordination to evade a sudden rock slide. In these cases, cognition "is like hibernation - it is adaptive, but the explanation for why it is adaptive goes *directly* from organic variation to survival" (ibid., p.79).

Godfrey-Smith is therefore right to be cautious of *equating* cognition with *non-trivial* homeostasis, even if some instances of cognition - or indeed many - can be understood in this way. However, reading the LRV as a version of the ECT does *not* necessitate this commitment. Admittedly, there is a temptation to equate Ashby's essential variables with Godfrey-Smith's 'intermediate organic properties' - body temperature, for instance, has been used in examples of both - and such an equivalence would indeed encourage an interpretation of the LRV as non-trivially homeostatic. But at the abstract level of the LRV, it does not appear possible to uphold any distinction between an 'intermediate organic property', and a property that is *constitutive* of survival. After all, for Ashby, the essential variables are *defined* as precisely those variables for which their stability is a condition of survival. So to the extent that essential variables are constitutive of survival, rather than mediatory of it, the LRV asserts only a 'trivial' homeostasis: as Godfrey-Smith says, the explanation for why a response is adaptive can go *directly* from variation to survival (survival being, by definition, stability in essential variables). The mechanisms underlying the generation of behaviour may be genuinely homeostatic in some cases, not so in others. An interpretation of the ECT drawn from the LRV is consistent with this position, and provides a useful conceptual tool for exploring those situations in which genuine homeostasis *does* apply. (For a recent agent-based exploration of non-trivial homeostasis see Di Paolo, 2000, in which neuronal homeostasis underlies adaptation to visual field inversion in a model of phototaxis.)

It may be suggested, in summary, that the ECT fails to avail itself of the extended understanding offered by the LRV in several ways: (1) a particular emphasis on distal variability, together with the distinction between that which threatens the organism and that which facilitates its activity, (2) a potential for quantitative expression in terms of measurements of variety, and for empirical exploration in terms of instantiation of disturbances, responses, and essential variables, and lastly (3) an engagement with the idea of cognition as homeostasis, but *not* a blanket commitment to this proposition.

4.3 Construction

Externalism and internalism together define a third position, which we have not so far discussed, that of *construction;* the explanation of properties of the external in terms of properties of the internal. The third and final element of our critique concerns the role of construction in the ECT.

The most obvious interpretation of construction is when the actions of an agent alter structures of the external environment, for example when a beaver builds a dam. Something about the external environment has changed, and, because this change also figures in the *Umwelt* of the beaver, it can elicit subsequent responses from the animal. For Godfrey-Smith, the key feature of this 'narrow', or 'causal' sense of construction is that "some change is made to an intrinsic property of something external to the organic system" (ibid., p.146). For reasons explored below, this is the only sense that Godfrey-Smith entertains in his formulation of the ECT.

However, construction can also be interpreted in a 'constitutive' or 'ontological' sense. For Godfrey-Smith this is the sense in which "[f]eatures of the environment which were not physically put there by the organism are nonetheless dependent on the organisms's faculties for their existence, individual identity or structure" (ibid., p.145). The key features of this sense of construction are that properties of an organism entail the existence of features in the *Umwelt* of that organism, that the relations between these features and features of the external environment need not be straightforward, and that no change to intrinsic properties of things external to the agent need be involved.

Constitutive construction can be manifest in several ways. Simply by moving around, an agent can influence what features of the external environment can influence its activity, without necessarily altering these features as they appear in the external environment (although relational properties of the external environment may have changed, intrinsic properties of things external to the agent will have remained the same). Even the fact that an agent is a particular *size* can influence the statistical structure of its *Umwelt*, for example by determining whether or not a field is homogeneous with respect to temperature, light intensity *etc.* A third manifestation of constitutive construction attaches to the way in which organisms "transduce the physical signals that reach them from the outside world" (Lewontin, 1983, p.100). For example, an (external) environmental change - an approaching rattlesnake - that may entail a change in the rate of vibration of air molecules, which is then transduced by the organism into some feature of the agent's *Umwelt*, perhaps associated with - or identifiable with - changes in the concentration of particular chemicals, which are themselves "transformed by the neuro-secretory system into the chemical signals of fear" (ibid., p.100).

Constitutive construction therefore describes the process by which *Umwelt* is generated from the external environment, and as we have argued previously, an understanding of this process may facilitate - or perhaps be essential - in effectively relating behavioural and mechanistic levels of description. Any interpretation of the ECT that respects the distinction between behaviour and mechanism must therefore be considered weak unless it entertains a role for constitutive construction (similar concerns are entertained in Oyama, 1996; Bedau, 1996).

Why, then, does Godfrey-Smith explicitly limit his formulation of the ECT to entertain only narrow, causal construction? One reason seems to be that he fears an admission of constitutive construction would necessitate an interpretation of the mechanisms underlying behaviour as *accommodatory.* This is evident from his choice of an example of constitutive construction:

> Suppose an organism develops a way to detoxify some chemical in its environment which was formerly highly poisonous to it. This organism has made an internal change to its chemistry, and it has also made a change to the relational properties of the external chemical. There is now one less thing the chemical can poison. But if the organism has not, in doing this, made any change to an intrinsic property of any external feature then this is a paradigm case of an internal accommodation of the environment. It is the type of thing to be *contrasted* with [narrow] constructive actions such as physically removing the chemical from the environment or spraying something on it to change its intrinsic nature. (1996a, p.147)

Two points arise here. First, this example is *not* a paradigmatic example of internal accommodation in the sense in which accommodation is understood in this paper (the external somehow becoming 'internalised', a transgression of pragmatic externalism). The 'internal change to its chemistry' could perhaps be interpreted as a response, and if so, then the kind of constitutive construction alluded to in this example would *not* necessitate accommodation, would be consistent with pragmatic externalism, and as such should not pose a problem for Godfrey-Smith.

Secondly, and more importantly, it is certainly not the case that this example is representative of constitutive construction in general. Consider another hypothetical example. Suppose an organism has a property such that some types of fluctuation in light intensity constitute relevant stimuli for it, whereas others do not. The *Umwelt* of the agent thereby contains these relevant stimuli (or transformations of them). The organism responds to these stimuli (or their transformations) in order to maintain some internal variable within a certain range (perhaps by moving away from certain kinds of bright lights in order to maintain body temperature). In this example, as in the examples described in the previous section, it is clear that constitutive construction plays a role - the generation of *Umwelt* - in determining how the organism should respond to (not accommodate) the environment. Godfrey-Smith is, however, reluctant to discuss this potentially useful interpretation of constitutive construction:

> What do we say about the role which organisms play in determining which properties of the environment are relevant to them? [...] We should say that and nothing more: *relevance* is a good concept to capture these phenomena [...] The organism plays a role in making it the case that its environment contains *relevant complexity* or not. (ibid., pp.148-154)

"But what then is left of the ECT?" Godfrey-Smith asks this very question (ibid., pp.154), and he answers with a 'concession' and a 'bet'. The concession - which he describes as a concession to internalism - is that "the organic system in question does play a role in determining whether or not a given environmental pattern is relevant to it" (ibid, p.155). The bet is that once this role has been played, "there will be other organic properties that can be explained in terms of this environmental pattern" (ibid., p.155). This 'externalist bet' is to be contrasted with its internalist counterpart, that once the first role has been played (of generating *Umwelt*), there would be little or nothing left to explain about the organic system. This way of putting things seems to suggest that constitutive construction be understood as a *precondition* of application of the ECT, the generation of *Umwelt*, not as something to be explained *by* it. But this, as we have seen, is a step Godfrey-Smith appears unwilling to take, for reasons that seem to stem from an attachment to the idea that construction must feature, if at all, as a consequence of the ECT, and that as such, must involve accommodation.

But that's not all. It may also be possible to interpret construction as a consequence of the ECT, *without* this entailing an identification of construction with accommodation, by considering the generation of an appropriate *Umwelt,* in some cases, as constituting a *response* to external environmental variability. Before placing much faith in this idea, however, we must face some difficulties to do with mechanism. As remarked above, to the extent that the generation of *Umwelt* is understood as a precondition of application of the ECT, those aspects of mechanism involved in this generation can no longer be explained by the ECT; only those aspects of mechanism mediating responses, given the composition of the *Umwelt,* would fall within its explanatory domain. However, if the generation of *Umwelt* is construed as response, then those mechanism structures mediating the generation of *Umwelt do* fall within the explanatory domain of the ECT after all. The problem is that there may be no *a priori* way of disambiguating these two interpretations for any given agent-environment system. Furthermore, there may be no way of unambiguously identifying those aspects of mechanism involved in the generation of *Umwelt* and those involved in responding to stimuli in the *Umwelt* (indeed to expect this to be possible at all would be to place questionable faith being able to meaningfully distinguish between perception and action).

Here the value of concrete agent-based models is perhaps clearest of all. If there is no escaping the need for pre-existing internal structure in subserving the translation from external environment to *Umwelt* - and there isn't - then this must simply be accepted, and as we have said before (section 4.1), any concrete model of agents and environments in continuous interaction and mutual specification must *necessarily* endow agents with some initial structure, if the model is not to be infinitely trivial. In this way, once one starts building concrete models of this sort, issues of constitutive construction attach to operational details of the model and cease to be philosophical obstacles.

To give some examples, models that endow agents with stable sensors (but other potentially modifiable aspects of mechanism) can treat the constitutive construction mediated by these sensors as a precondition of application of the ECT. This then clears the way for understanding - in terms of the ECT - the behaviour patterns that respond to features of the external environment, and for understanding - also in terms of the ECT - the mechanistic structures (other than the sensors) that respond to features of the *Umwelt.* The work of Beer (1996, 2000) again comes to mind (see section 4.1), as does Seth (1998) in which reliable and noisy sensors were compared in an evolutionary robotics task involving homing navigation; it was found that evolved mechanisms integrated multi-modal sensory data in the latter case, but utilised only a single modality in the former, and the qualitative complexity of behaviour also differed across the two conditions.

By extension, models that allow the sensors themselves to adapt can admit the constitutive construction associated with this adaptation into the explanatory domain of the ECT. Harvey, Husbands, and Cliff (1994), for example, explore the artificial evolution of sensor morphologies for robots faced with the task of discriminating between triangles and squares, as they appear in the external environment. This, on the face of it a difficult problem, becomes trivial for the robot once it has adapted its sensors to generate a very simple *Umwelt*. Going even further, models that allow the entire agent morphology to adapt can be interpreted in the same way (Pfeifer, 2000; Pfeifer & Scheier, 1999).

In either case, constitutive construction associated with agent *movement* can fall within the explanatory domain of the ECT to the extent that this movement can itself be interpreted a response to the external environment. Todd and Yanco (1996), for example, explore several ways in which the adaptive significance of (externally apparent) resource 'clumps' depends upon the movement of simulated agents in a simple concrete model, and Nolfi and Parisi (1993) assess the performance of systems which have the ability to expose themselves only to sub-classes of stimuli to which they can effectively respond.

A last comment on this issue is that I am nowhere denying that 'narrow' or 'causal' construction is also important (and the distinction between causal and constitutive varieties is certainly worth keeping). By intervening in formerly autonomous external environmental processes, an agent may well bring about modifications in that environment - which may entail changes in the *Umwelt* - which can then elicit distinctive responses. The point is simple that causal construction on its own *is not always enough.*

5. Agent-based modelling

We have discussed the utility of agent-based models at several stages in this paper. To summarise: they explicitly realise the distinction between behaviour and mechanism, getting around the slipperiness of 'cognition'. They allow the quantitative exploration of the ECT from the perspective of Ashby's LRV, and, by opening *Umwelt* generation to empirical analysis, they find a place for constitutive construction in the ECT, as both precondition *and* explanandum.

We have not said much about the particular kinds of agent-based models best suited for these purposes beyond the general requirements that they separate intuitively into agent and environment components and display activity at both behavioural and mechanistic levels of description. Whilst it is beyond the scope of this paper to comprehensively cover all relevant paradigms (for example the dynamical systems approach well articulated in Beer 1995, or the various agent-based methodologies described in Pfeifer & Scheier, 1999), it is worth mentioning one broad class of model that may be particularly useful, indeed many of the examples previously discussed have been of this kind. Artificial evolution models not only satisfy the above general requirements, but also they are *not* required to prefigure the relations between behaviour and mechanism (see Nolfi, 1998). The designer can specify a (teleo-functional) fitness function, and leave the evolutionary process to work out the details of the underlying mechanism (it remains incumbent on the designer, of course, to specify in advance *some* aspects of mechanism structure). They are also (pragmatically) externalist in character, at least superficially, in that properties of the internal are moulded, by selection, to engage with properties of the external. Indeed a useful case can be made for understanding such models as extensions of biological 'optimal foraging' models in biology, which operate on exactly this premise but which are overly constrained with regard to the relations between behaviour and mechanism (Seth, 2000b).

Agent-based modelling (in general) also allows the issue of pragmatic externalism to be tackled head on. Whilst it has been stated as a requirement that such models separate intuitively into agent-environment components, and should initially be analysed as such, further analysis could examine how the explanatory story of a model would differ if *different* boundaries were demarcated between internal and external, an enterprise which could shed light on the nature of pragmatically externalist explanation, and possibly also on our intuitions about how agents and environments should normally be distinguished.

Lastly, it would be very remiss not to note that Godfrey-Smith does what few philosophers dare (as Belew, 1996, remarks) in presenting some concrete models of his own in the second part of his book. There are certainly points of contact between these models and those advocated here, one examines the costs and benefits of phenotypic plasticity, for example, and another asks with signal detection theory how much sensory acuity an agent should 'invest in' in different environments, but in general the focus is on mathematically quantifying agent and environmental variability, and less on agent-based elucidations of the relations between behaviour and mechanism. Nevertheless, they are to be highly recommended, not only for their specific insights, but also for their alternative perspectives on modelling the ECT, and not least for reassurance that effective modelling of a philosophical project is indeed possible.

6. Summary

The organising feature of this paper has been a courtship between two broad themes: distinguishing behaviour from mechanism, and relating behavioural function to environmental structure. In its simplest form, the resulting message is this: the ECT articulates the pragmatically externalist hypothesis that environmental complexity can incur behavioural complexity, but to understand how such behavioural complexity relates to underlying mechanism it is of enormous importance to understand how the agent perceives its environment, how the external environment is translated into *Umwelt.* This kind of understanding can figure both as a precondition, and as a consequence of application of the ECT.

More specifically, we have considered three related critiques of Godfrey-Smith's ECT which, in combination, lead us to a new interpretation of his idea which is that *behavioural and/or mechanistic complexity can be understood in terms of mediating well-adapted responses to environmental variability.* This interpretation differs from Godfrey-Smith's in three important ways, corresponding to the three elements of the critique offered above. The first is that there is no commitment to an interpretation of 'cognition' which obscures the essential distinction between behaviour and mechanism. This is important because the adaptationist claims of the ECT attach to behaviours, yet any behaviour can be subserved by a variety of mechanisms. This is why there an obvious ambiguity in my interpretation of the ECT; the relations between behaviour and mechanism cannot be prespecified in advance of consideration of any particular agent-environment system.

The second difference is that the interpretation of the ECT favoured here accepts - and exploits - the parallel with Ashby's LRV. This extends the understanding of the role of homeostasis, lights the way to quantitative modelling, and highlights the distinction between environmental variability threatening to the agent and environmental variability which facilitates its activity.

The third difference is the attempt to explicitly incorporate constitutive construction, on the one hand as a precondition of application of the ECT (fixed sensor structure, for example), and on the other as a response explicable by the ECT (adaptive sensors, for example, or *Umwelt* generation through movement). This third difference is intimately related to the first insofar as understanding the relations between behaviour and mechanism can be facilitated - or may even require - tracing of the relations between the external environment and the *Umwelt.*

Finally, two other distinguishing features are apparent from the wording. The term 'variability' is preferred over (environmental) complexity, largely in view of the specific ideas about variability in the perspectives of Dewey and Ashby, and the greater handle afforded for empirical analysis. Also, I speak of behavioural/mechanistic *complexity;* this is perhaps the most easily falsifiable element of the present formulation (recall, one final time, Simon's ant), but this only goes to emphasise that the ECT is, above all, a hypothesis, hypotheses are after all to be tested, and a wonderful thing about the ECT is that falsifying evidence is likely to be just as interesting as evidence adduced in its support.

Acknowledgements

I am grateful to the CCNR, the Neurosciences Research Foundation, my anonymous reviewers, and also to Peter Todd and Andy Clark for a combination of financial support and constructive discussion. The material in this paper is drawn in part from my D.Phil. thesis (Seth, 2000a).

References

Ashby, W. (1956). *An introduction to cybernetics.* Chapman Hall, London.

Bedau, M. (1996). The extent to which organisms construct their environments. *Adaptive Behaviour*, *4*(3/4), 476–483.

Beer, R. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, *72*, 173–215.

Beer, R. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behaviour. In Maes, P., Mataric, M., Meyer, J., Pollack, J., & Wilson, S. (Eds.), *From animals to animats 4: Proceedings of the Fourth International Conference on the Simulation of Adaptive Behavior*, pp. 421–429 Cambridge, MA. MIT Press.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*, 91–99.

Belew, R. (1996). Developments across the internalist/externalist dichotomy. *Adaptive Behaviour*, *4*(3/4), 483–486.

Biró, Z., & Ziemke, T. (1998). Evolution of visually-guided approach behaviour in recurrent artificial neural network robot controllers. In Pfeifer, R., Blumberg, B., Meyer, J., & Wilson, S. (Eds.), *From animals to animats 5: Proceedings of the Fifth International Conference on the Simulation of Adaptive Behavior*, pp. 73–77 Cambridge, MA. MIT Press.

Brandon, R. (1990). *Adaptation and environment.* Princeton University Press, Princeton.

Brunswik, E. (1952). *The conceptual framework of psychology.* University of Chicago Press, Chicago.

Cannon, W. (1932). *The wisdom of the body.* Norton, New York.

Dennett, D. (1987). *The intentional stance.* MIT Press, Cambridge, MA.

Dewey, J. (1929). *Experience and nature.* Dover, New York. revised edition.

Di Paolo, E. (2000). Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In Meyer, J., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. (Eds.), *From animals to animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*, pp. 440–449 Cambridge, MA. MIT Press.

Fletcher, J., Zwick, M., & Bedau, M. (1996). Dependence of adaptability on environmental structure. *Adaptive Behaviour*, *4*(3-4), 275–307.

Fletcher, J., Zwick, M., & Bedau, M. (1998). Effect of environmental texture on evolutionary adaptation. In Adami, C., Belew, R., Kitano, H., & Taylor, C. (Eds.), *Proceedings of the Sixth International Conference on Artificial Life*, pp. 189–199 Cambridge. MA. MIT Press.

Godfrey-Smith, P. (1996a). *Complexity and the function of mind in nature.* Cambridge University Press, Cambridge.

Godfrey-Smith, P. (1996b). Precis of *Complexity and the function of mind in nature*. *Adaptive Behaviour*, *4*(3/4), 453–466.

Goodwin, B. (1994). *How the leopard changed its spots: The evolution of complexity.* Phoenix, London.

Hardcastle, V. (Ed.). (1999). *Where biology meets psychology: Philosophical essays.* MIT Press, Cambridge, MA.

Harvey, I., Husbands, P., & Cliff, D. (1994). Seeing the light: Artificial evolution, real vision. In Cliff, D., Husbands, P., Meyer, J., & Wilson, S. (Eds.), *From animals to animats 3: Proceedings of the Third International Conference on the Simulation of Adaptive Behavior*, pp. 392–401 Cambridge, MA. MIT Press.

Hendriks-Jansen, H. (1996). *Catching ourselves in the act: Situated activity, interactive emergence, and human thought.* MIT Press, Cambridge, MA.

Lewontin, R. (1983). The organism as subject and object of evolution. In Levins, R., & Lewontin, R. (Eds.), *The Dialectical Biologist*, pp. 85–106. Harvard University Press, Cambridge, MA.

Lorenz, K. (1937). The nature of instinct: The conception of instinctive behaviour. In Schiller and Lashley (1957), pp. 129-175.

Lorenz, K. (1948). *The natural science of the human species.* MIT Press, Cambridge, MA. edited from author's posthumous works (1944-48) by A. von Cranach, trans. R.D. Martin, 1996.

McFarland, D., & Sibly, R. (1975). The behavioural final common path. *Philosophical Transactions of the Royal Society of London: Series B*, *270*(907), 265–293.

McLeod, P., & Dienes, Z. (1996). Do fielders know where to go to catch the ball, or only how to get there?. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 531–543.

McShea, D. (1996). Unpredictability! and the function of mind in nature. *Adaptive Behaviour*, *4*(3/4), 466–471.

Neisser, U. (1967). *Cognitive psychology.* Appleton, New York.

Nolfi, S. (1998). Evolutionary robotics: Exploiting the full power of self-organisation. *Connection Science*, *10*(3/4), 167–185.

Nolfi, S., & Parisi, D. (1993). Self-selection of input stimuli for improving performance. In Bekey, G. (Ed.), *Neural networks and robotics*, pp. 403–418. Kluwer Academic, Boston, MA.

Oyama, S. (1985). *The ontogeny of information.* Cambridge University Press, Cambridge.

Oyama, S. (1996). The ins and outs of nature and mind. *Adaptive Behaviour*, *4*(3/4), 471–476.

Pfeifer, R. (2000). On the role of morphology and materials in adaptive behavior. In Meyer, J., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. (Eds.), *From animals to animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*, pp. 23–32 Cambridge, MA. MIT Press.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence.* MIT Press, Cambridge, MA.

Seth, A. (1998). The evolution of complexity and the value of variability. In Adami, C., Belew, R., Kitano, H., & Taylor, C. (Eds.), *Artificial Life VI: Proceedings of the Sixth International Conference on the Simulation and Synthesis of Living Systems*, pp. 209–221 Cambridge, MA. MIT Press.

Seth, A. (2000a). *On the relations between behaviour, mechanism, and environment: Explorations in artificial evolution.* Ph.D. thesis, University of Sussex.

Seth, A. (2000b). Unorthodox optimal foraging theory. In Meyer, J., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. (Eds.), *From animals to animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*, pp. 471–481 Cambridge, MA. MIT Press.

Seth, A. (in press). Distinguishing adaptive from non-adaptive evolution using Ashby's law of requisite variety. *To appear in Proc. 2002 IEEE Congress on Evolutionary Computation (CEC2002).*

Shapiro, L. (1999). Presence of mind. In Hardcastle, V. (Ed.), *Where biology meets psychology: Philosophical essays*, pp. 83–99. MIT Press, Cambridge, MA.

Simon, H. (1988). *The sciences of the artificial.* MIT Press, Cambridge, MA. 3rd edition.

Thompson, W. (1917). *On growth and form.* Cambridge University Press, Cambridge.

Todd, P., & Yanco, H. (1996). Environmental effects on minimal behaviours in the minimat world. *Adaptive Behavior*, *4*(3/4), 365–413.

von Uexküll, J. (1934). A stroll through the worlds of animals and men. In Lashley, K. (Ed.), *Instinctive behavior.* International Universities Press, New York.

Ziemke, T. (2000). On the parts and wholes of adaptive behavior: Functional modularity and diachronic structure in recurrent neural network robot controllers. In Meyer, J., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. (Eds.), *From animals to animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*, pp. 115–124 Cambridge, MA. MIT Press.

Ziemke, T., & Sharkey, N. (2001). A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robotics and artificial life. *Semiotica*, *134*, 701–746.