Simulating Autonomous Anticipation:
The Importance of Dubois’ Conjecture

John Collier
University of KwaZulu-Natal
School of Philosophy and Ethics
Durban 4041, South Africa
collierj@ukzn.ac.za

Abstract
Anticipation allows a system to adapt to conditions that have not yet come to be, either externally to the system or internally. Autonomous systems actively control their own conditions so as to increase their functionality (they self-regulate). Living systems self-regulate in order to increase their own viability. These increasingly stronger conditions, anticipation, autonomy and viability, can give an insight into progressively stronger classes of models of autonomy. I will argue that stronger forms are the relevant ones for Artificial Life. This has consequences for the design of and accurate simulation of living systems.

Keywords: autonomy, modelling, function, simulation, anticipation, emergence

DUBOIS’ CONJECTURE

Autonomy basically means self-regulation. Self-regulation implies internal control of system states to achieve greater functionality, either internally or interactively with the environment. Functionality, as a teleological notion, implies that the system must be directed by likely future states; that is, it must anticipate and (possibly) adapt to likely future states. Functionality does not require autonomy (assuming functional goals are externally set), but merely that current states of the system can select suitable future states on the basis of suitable input. In this
restricted case of functionality, selection of states is under the control of external inputs, and not under internal control. Specifically, the system in the minimal case does not add information to the input-output relation. If the inputs are strings of symbols, for example, and the outputs are strings of actions, then for the minimal case there is a many-one mapping of input strings to output strings (a function from input string to output) that completely describes the non-stochastic aspects of the system. This is equivalent to a behaviourist system on the well-known approach of B.F. Skinner. The simplest form of self-regulation must permit the selection of states over and above input-output relations: for example, the same input, depending on selection of internal states, may produce differing outputs, or the same output may be selected by differing inputs, or both. That is, the rules governing input-output relations vary, and must themselves be selected internally. Alternatively, self-regulation might be limited to the selection of internal states.\(^1\) More complex forms of self-regulation would involve higher order control of this internal selection of rules. The requirements of anticipation, functionality, and internal selection of rules will turn out to put some non-obvious constraints on what can be properly considered to be autonomous. In order to see this, the relevant concepts and their relations need further analysis.

Daniel Dubois (2003) makes a very useful distinction between external and internal control of future states which he calls, respectively, weak and strong anticipation. Dubois tentatively defines anticipation: “an anticipatory system is a system for which the present behaviour is based on past and/or present events but also on future events built from these past, present and future events”. Weak anticipation relies on externally produced data to internally model future states of

\(^1\) Information theoretic criteria for these cases are given by Bertschinger, Olbrich, Ay and Jost (this volume). They assume that “[a]utonomy might mean the freedom of a system to set its own goals, to construct its own rules of operation, or to select the methods for achieving its aims according to some internal procedure or set of rules that is shielded from the control of the environment the system happens to be situated in.” I will argue that their approach gives a measure of independence (which is correlated with autonomy, and might be considered to be a weak form of autonomy), but not directly of the sort autonomy found in living systems.
the environment, while strong anticipation uses internally produced data to model future internal states. In general, anticipation requires that some future state determines or guides present dynamics. Since the future state is not directly accessible, it must be modelled some place within the system and its environment. Weak anticipation relies on externally produced data (and environmental regularities) to internally model the environment and project its future states, while strong anticipation uses internally produced data to model internal states and project their future states. The notion of modelling here requires Rosen’s (1991) idea (well established and used in science) that goes back at least to Hertz, according to which logical (inferential) relations are used to represent the supposed causal structure of another system. The idea is that one system (the object of the modelling) has a certain causal structure, and another system (it could be the same one if the modelling is internal, i.e., in strong anticipation) has logical relations that mirror the causal relations of the object, so the model is a set of relations mirroring the set of relations constituting the structure of its object. Anticipation further requires that the logical structure of the model can be projected to possible future states, thus allowing the current modelling of future states. Functionality implies that these models are used to aim at some goal.

In order to connect the model (logical structure) to the world, the model itself must have a dynamical embodiment (a dynamical structure, defined in terms of forces and flows – the mathematical definition of dynamical system is insufficient here since it is a logical definition, and as such would get us no closer to the world). Elements of the model serve as signs (vicariants) for elements in the world, with their dynamical embodiment causally mirroring the logic, so that the dynamical processes embodying the model correspond to the logic of the model. To be effective in control, the model must also be connected to the system under control, though this may be very indirect. In weak anticipation inputs and outputs to the system are
transmitted to a model of the external world, computations (digital or analogue) are done, and an output is generated. The output is a function of the input and the logic of the model, up to any stochastic variation. Weak anticipation, then, takes the form of a behaviourist system, and control is determined by the input, or rather the sequence of inputs. In strong anticipation, there is an additional model of the system itself, which has inputs from the internal state of the system, and outputs modifying that state, allowing the system to control itself, certainly a necessary condition for autonomy understood as self-regulation.

Homeostasis is perhaps the simplest version of self-regulation. In temperature control in mammals, for example, there are signals from internal states to the hypothalamus, which compares those signals to internal conditions and computes a difference, which is then output to the system, either causing the internal state of the system to alter to bring the temperature in line with a set point (difference vanishes), or altering the behaviour of the system (e.g., seeking shade) to the same effect, or perhaps both. The alert reader will perhaps have noted that a thermostat performs the same function, but we do not consider a thermostat, or a thermostatically controlled system, to be autonomous. Why should we have different standards for living systems, or their simulations? The straight-forward answer is that we should not. An internal model controlling internal states is not sufficient for autonomy. At best it ensures independence of operation, where independence involves shielding of internal processes from external inputs.

Despite this problem, Dubois (2003) conjectured that an autonomous system is one that shows strong anticipation. The rationale behind it is that self-regulation is impossible without acting on the self and that this requires an (at least partial) model of the self to achieve in any consistent way (except for the degenerate case of a system that is its own model – Collier 2004a). If we consider systems that only model their environment, the logical rules of the model permit...
only environmental inputs and outputs, since that is all that the model represents. Some independence from the environment may result, for example if the environmental cues result in outputs that move the system to a safer location, or to a source of energy required for continuing operation. However, the behaviour of such a system is determined by its environment, so it is a behaviourist system as described above. If there is a self-model, however, mappings of internal states onto internal states are possible, which allows the system to control its internal condition, and possibly also to control the rules that govern its responses to environmental signals. It appears then that an internal model is sufficient for self regulation. If it is also necessary, then Dubois’ conjecture is true. Most of the rest of this paper is directed at establishing the necessity condition and its consequences. First, though, there is an issue that needs to be cleared up about the nature of Dubois’ conjecture. I think that a modification, or rather clarification, of Dubois’ conjecture is in order.

Noting that a thermostat is a behaviourist system as defined above; perhaps homeostasis is as well. As discussed so far, it might be. The role of the internal model can be eliminated in favour of a more complex external model involving a more complex computation. I suggest that strong anticipation be understood in terms of the need for an ineliminable internal model. A model is ineliminable if and only if the set of relations it corresponds to cannot be found among the relations defined on all possible pairs of strings of inputs and outputs.\(^2\) This ensures that the model will contain information not in relations between its inputs and outputs, and the system will not be a behaviourist system. As Rosen (1991) noticed, such a model must be in the cross (Cartesian) product of the set of pairs of strings of inputs and outputs (or more generally, of ordered n-tuplets of observables). If we call an internal model that can be eliminated in favour of

\(^2\) The appropriate reference set is the set of all possible input-output relations, not some finite past set as in (Bertschinger et al). The latter can give only independence of the internal processes from past conditions.
a more complex external model to which it is equivalent a trivial internal model, and require of strong anticipation that the internal model be nontrivial, then we can keep Dubois’ conjecture.\(^3\) I believe this was the intended form (see Collier 2004b, 2006).

Consider homeostasis again. What might have gone wrong in treating it as a paradigm of autonomy? There are several possibilities. It could be that homeostasis is not always autonomous in living creatures. The supposed model would imply this, if the model is correct homeostasis is not really autonomous, despite appearances. Or it could be that the model is not accurate. Given that the functionality of the temperature regulation system in mammals is quite a bit more subtle than that of a thermostatically controlled room, this is not unlikely. Or it could be that despite our intuitions a thermostatically controlled room really is autonomous. At the very least there is something like autonomy here. In the example above of environmental modelling to obtain safety or food I called this independence. A thermostatically controlled room does act independently. Is independence enough for autonomy? It is not enough if the internal model must be ineliminable.

**DEFINING AUTONOMY**

A common definition of an autonomous system in Artificial Intelligence and related research such as robotics is that the system can use an internal state to adjust some internal state of own. One example given in a standard introductory AI text is a robot that adjusts its timekeeper to local time when it crosses time zones, rather than relying on external setting by a programmed input or by data from a local clock. Obviously this example has problems, and workers in the field are well aware of this, but the idea gives a rough and ready operational distinction between external and internal control. The source of the information is external in the nonautonomous

---

\(^3\) I use ‘trivial’ here specifically to invoke von Foerster’s (1960) distinction between trivial and nontrivial machines. It is obvious (in the mathematical sense) that behaviourist systems are trivial machines, and that nontrivial models
case, and internal in the autonomous case. It should be noted that this sort of autonomy, which I will call autonomy$_1$, requires that the system have an internal model of itself (or rather part of itself) in order to control itself. Unfortunately, autonomy$_1$ doesn’t require that the internal model be nontrivial, so implementations of autonomy$_1$ are at best poor simulations of autonomy. It is too weak to require the sort of organic autonomy that Kant conjectured in his *Critique of Judgement* (which requires self-determination of goals as well as means). We need a stronger, self-determinedly teleological autonomy. We could define autonomy in terms of Dubois’ conjecture, but this would make it a definition, not a conjecture, which would teach us little if anything.

To satisfy the stronger notion of autonomy, I will later introduce autonomy$_2$ and then autonomy$_3$ which satisfy stronger conditions of internal closure. I then argue that any naturally evolved system that satisfies Dubois’ condition for strong anticipation will almost certainly also satisfy the conditions for autonomy$_2$ or autonomy$_3$. I next argue that in the context of Artificial Life, the stronger notions of autonomy are the relevant ones. I then discuss some of the consequences of these forms of autonomy, and some rather strong constraints on how they can be realized artificially. But first we need to approach autonomy from another direction, or the question of the truth of Dubois’ conjecture will be begged.

**THE LOGICAL STRUCTURE OF AUTONOMY**

Dubois’ conjecture implies that autonomy adds information to the information processed by an autonomous system. Is there anything like this, and can we get to it analytically without presupposing it? I start by constructing an idea of autonomy by starting with well-known philosophical notions, and then restricting these by adding further conditions until we arrive at a
suitable definition. The desiderata of a suitable definition are a) it accounts for the identity and individuation of the autonomous entity, b) it accounts for the functionality of the autonomous entity, on pain of triviality, c) it has a suitable account of self-governance, and d) it is interpretable in dynamical terms (so we can detect and possibly build such an entity by intervening in the world). I start with identity (Collier 2002, 2004a, 2004b, 2006), since it is the broadest of the four desiderata (function must be grounded in some entity which the function is for, self-governance implies a self to govern. The reflective character of autonomy requires that it contains a notion of identity. This may seem trivial, but it is amazing how often this triviality is ignored. Start with logical identity (dynamical identity is a special case of identity).

1. Identity, A = B

   a. Is a logical condition, same for all things
   b. Identity is an equivalence relation: symmetric, transitive, closed
   c. A = B implies that B has every property that A has, and vice versa

This tells us virtually nothing, since it is a purely logical relation, but it does put these logical constraints on any concept of autonomy. An autonomous system must map states of itself onto itself, requiring preservation of the above conditions. The next move is to look at what makes parts of something parts of that thing. This is provided by the unity relation.

2. Unity, U(A)

   Unity is the relation among the parts of a thing A such that:
   a. If a and b are parts of A, then aUb, and bUa (symmetric)
   b. If a and b are parts of A, then aUb and bUc implies aUb (transitive)
   c. By a. and b., U is an equivalence relation
   d. U(A) is the closure of U, given any initial part.
By a. to d., \( U(A) \) contains all and only the parts of \( A \).

It is empirical question what satisfies \( U(A) \) for a given \( A \). Typically the type of unity relation will depend on the sort of thing \( A \) is. Autonomous entities, as noted above, must be dynamical entities, so the sort of unity relation we need a dynamical form of unity. In previous writing, I have called this cohesion (Collier 1986, 1988, 2003):

3. Cohesion, \( C(A) \), dynamical unity

Cohesion is the unity relation for dynamical objects, such that:

a. All parts \( a \subset \subset b \) are dynamical

b. \( C \) is dynamical

Simple examples of cohesion:

A quartz crystal -

the closure of intermolecular interactions gives the boundary of the crystal, external interactions being much weaker than internal interactions

A gas in a box -

the cohesion of the box that defines the boundaries of the gas.

These examples are at the two extremes of cohesion, with the crystal having strong internal forces that keep it relatively rigid, whereas the gas has little internal stability, and is held in place by its container. Most dynamical systems sit someplace in between. There are several things to note about cohesion:

- In each case cohesion is not absolute; it is a matter of degree. Some things are more cohesive, and others less. It is the balance of inward and outward tendencies that matters.
- We should expect difficult intermediate cases in which the balance is close to equal.
- Cohesion can differ in strength in different dimensions (factors)
We really need a multidimensional and statistical cohesion profile to individuate an object.

Cohesion both unifies a dynamical object, and distinguishes it from other dynamical objects.

Thus, it is effective as a criterion of individuation.

Its real strength, however, is in the way it forces us to look for dynamical closure whenever we want to claim that something is individuated.

This is especially significant in the case of autonomy.

Autonomy is a special type of cohesion that is maintained actively (through work – or in a simulation, computation) through the contributions of component processes to the continued existence of the system, either directly, or through intermediate processes. The requirements of autonomy place certain restrictions on what sort of organized system might be autonomous. It should be obvious that neither a rock nor a gas in a box is autonomous, since they don’t act on themselves in any sense. In particular, neither a rock nor a gas in a box is autonomous because they cannot alter their own state, especially not to respond to processes that go across their boundaries. Thus they are unable to adapt to their own conditions or conditions around them, and certainly not to anticipate them.

In order to have this sort of self control, a system must be internally differentiated (this condition is relatively trivial, but its justification caused problems for one of the referees: self control requires action on the self; homogeneity permits no locus of action, and the system can merely act, which is not action on itself). Furthermore, it cannot be in a uniform steady state, but must have a number of internal states that are dynamically accessible. This requires a certain flexibility that systems whose cohesion is based in high energy differentials cannot maintain.
Being active (doing work: self action requires doing work on oneself, or there is no effect – again I mention this only because one of the referees of this paper was puzzled) requires the existence of non-equilibrium conditions. This means that there must be available resources to make use of.

I cannot stress too much that autonomy is impossible unless there are sufficient resources available for use. Thus we can expect it to be characteristic of autonomous systems that energy is not their primary concern, but rather organization of their processes so as to divert matter and energy as suitable for their survival. It would be proper, then, to describe autonomous systems, and the degree of autonomy itself, in terms of relative organization rather than in terms of relative energies of interactions. This is coherent with the intuition behind autopoiesis (Varela 1979, Maturana and Varela 1980) that organization is of central significance, and also with Rosen’s (1991) idea of reflexive networks of efficient causation.

Furthermore, since processes contributing to autonomous cohesion must be coordinated so as to achieve viability, we should expect autonomous systems to show holistic organization of a hierarchical sort in which open aspects of lower level processes are closed at higher levels. However, unlike in autopoietic systems, this closure need never be complete. While internal process closure (sometimes called operational closure) is to some degree essential, there will also be interactive closure among processes, both within the internal infrastructure and with features of the external environment. This means that autonomy must be open to the outside. It requires only that the internal organizational closure is greater than the interactive closure. This is in line with the general idea that the individuation of dynamical entities is a matter of degree. It is a mistake to think that operational closure must be complete. Dynamical systems in general are open in all respects, or else they cannot evolve or develop in those respects that they are closed.
In practice, application of this autonomy concept presents some difficulties. Comparing degrees of organization is non-trivial. In algorithmic complexity theory, logical depth, or the number of steps required to produce a surface structure from a deep structure, is often taken as a measure of organization. This value, though, is difficult to compute under the best of circumstances, and often impossible. Henri Atlan (Koppel and Atlan 1991) uses a somewhat different notion that he calls sophistication (which bears some similarities to logical depth) to argue that some living biomolecular systems have infinite sophistication. Such systems are not reducible, and must emerge through spontaneous “self-organization” (Maturana agrees on the spontaneity but not the ‘self’ part, see Collier 2004a, footnote 1). Comparing organization of infinitely deep structures can only be done in terms of overall complexity of the organization, which itself is hard to determine.

Fortunately, differences in organization are often large, and are correspondingly easy to recognize. Thus we can pick out many autonomous systems, and even compare their degree of autonomy, and, further, to compare autonomy in various respects. Because cohesion, and thus autonomy, comes in degrees (and degrees in various dimensions, at that), there will be cases that are difficult to classify, for example, slime mould. But this is as it should be, since the world does really contain such intermediate cases. This is quite different from autopoiesis, which is an all or nothing, quite indiscriminate condition (Varela 1979).

In summary, autonomy requires non-equilibrium conditions, internal dynamical differentiation, hierarchical and interactive process organization, incomplete closure, openness to the world, and openness to infrastructural inputs from the material basis of the organization. The existence of autonomy, like any cohesion, is identical to the corresponding process closure, and is not something complementary to, or over and above, this closure. When we want to look for
autonomy, we should look for the appropriate types of process closure. Since one of the central elements of the notion of unity, and thus of cohesion and autonomy, is an empirical component, we cannot decide a priori what things or systems are autonomous. This is at least in part a scientific question that requires observation and experiment.

From this account of autonomy, we can determine some likely properties of autonomous systems. First, they will tend to exhibit levels of both organization and modularity. This is because self-interaction is essential, which is most efficiently done with a certain degree of modularity. Secondly, these modules must be largely closed (cohesive), but preferably should also be organizationally open in order to play a part in larger organization. Thirdly, the system will show levels of organization (and perhaps structure) because the modules must be integrated so that the organization can be closed. Just as with the modules, the levels must be cohesive to maintain their identity, but they must also be open to lower level and higher level influences in order to perform their integrative tasks.

**FUNCTIONALITY RECONSIDERED**

Functionality can be defined externally (as allowed in the discussion at the beginning of this article), but natural functionality must defined internally. Natural functionality derives from autonomy through its contribution to self-preservation, or viability of an autonomous entity. This dependence can be either direct or indirect; lower level functions can serve higher level functions, and so on, but ultimately the functions have to be for the sake of an autonomous entity. Because autonomous entities are usually, if not always, emergent (not reducible), functionality cannot typically be reduced to the purely mechanical processes of the substrate of the organization. Furthermore, autonomous entities can emerge from other autonomous entities,
either from individuals or interacting collections. These higher level autonomous entities can have functions that conflict with lower level functions.

The requirement that functionality must be grounded in autonomy (and is, furthermore, partly constitutive of autonomy) would be satisfied it involves Dubois’ (2003) strong anticipation, as noted above. Weak anticipation, on the other hand, is directed purely at external systems. The etiological concept of functionality (Wright 1973, Millikan 1989, Neander 1991) requires no more than this, since functionality is defined in terms of what something does – its input-output relations. The organizational account, however, requires internally directed action. Does it require that the internal direction is ineliminable? If so, then Dubois’ conjecture is shown to be necessary.

On the present account so far, even though functionality best fits the idea of strong anticipation, there need be no model of the system of itself, except in a very abstract sense. In effect, the autonomy of the system could be its own model; it is genuinely self-directing. On Aristotelean accounts of action, nothing can be self-directing (Juarrero 1999). The Aristotelean view requires a logical separation of agent and subject. Dubois’ self modelling account of anticipation appears to require this separation. Juarrero shows, though, that the dynamical approach obviates the need for this separation: there is no need for separation of system and model in order to embody self-direction, if the dynamics of the system are emergent. The autonomy account of functionality, by dynamically embodying functionality in the very processes that make up autonomy, unifies both agent and action inextricably. This is possible only because the non-mechanical nature of emergent autonomy allows us to escape the causal model of bumps and nudges that must be traced back to some further source. Genuine self-direction requires dynamical emergence of autonomy. This requirement ensures that that the
internal model (formal structure corresponding to the dynamical self) must be ineliminable, showing that Dubois’ condition is not just sufficient, but is also necessary.

It is common to talk of the function of some component or process, but the autonomy account of functionality allows that functionality can be diffuse, and not easily localized, so talk of specific function is not always warranted even though there may be functionality. Proper functions certainly exist, but they are possible only when the functional subsystem has become stabilized and modularized, for example, as the heart has become a separate and more or less modular organ through evolution. Many aspects of function, however, are not so highly modularized, and all we can talk about is functionality, rather than the proper function of some part or aspect of the organism.

Closure places severe restrictions on what can qualify as autonomous. This has consequences for determinations of functionality and partship, which in turn has implications for satisfactory explanations. First, explanations of parts must be open-ended in order to allow for the open character of their interactions with other parts. However, complete explanations must not leave functional processes open, but should at least define the classes of appropriate inputs and outputs in a way that is compatible with their grounding in dynamical organization. Explanations should be constructed or rendered so that modular and nesting connections can be made, and explanations in terms of organization or energy and matter flows alone are incomplete; the two must correspond in every detail at the appropriate level. As with explanations, these are also desiderata for simulations.

THREE DEGREES OF AUTONOMY
In weak anticipation future states of the environment are projected using a model of the environment and environmental data alone. In strong anticipation, however, the system requires a model of itself (or part of itself) in order to project its future states from its current internal state. The idea behind Dubois’ conjecture is that self-control requires the use of an internal model of the self. This can take on three related forms:

**Autonomy\(_1\):** The internal model computes internal states, thereby determining the future states. This is not the autonomy described above that supports functionality, since it doesn’t require organization or viability.

**Autonomy\(_2\):** Maturana/Varela, or Rosen style organization, which is closed and does not come in degrees; it is either there or it is not (though complexity might give some relative values of internal organization). The closure *is* the autonomy.

**Autonomy\(_3\):** The dynamical interactive version given above, in which the degree of autonomy is measured by relative values of internal to external organization involved in interactions. Again, closure is the autonomy, but it holds only in degrees.

The problem with autonomy\(_1\) is that unless the internal model is irreducible at least in the sense of autonomy\(_2\), all causes can be traced back to external events, so weak anticipation is sufficient for explanation of the activity of the system. Both Rosen (1991) and Maturana and Varela were well aware of the deficiency of autonomy\(_1\). Whereas the latter gave general arguments about the need for self-production to be separated from underlying mechanisms, Rosen showed that some models of some systems cannot be reduced to linear combinations of
their components or to input-output relations. Such systems, he said, typify living systems. Whether he was right that only living systems are like this is dubious, but he is likely correct in thinking that living systems must be closed to efficient causation. So autonomy\textsubscript{1}, though apparently fitting Dubois’ conjecture, is really a case of weak anticipation, unless it is also one of the stronger cases. There are still some gaps in the arguments, most evidence from work that considers the issue of irreducibility points to living systems not being merely autonomous\textsubscript{1}.

Autonomy\textsubscript{3} differs from autonomy\textsubscript{2} in that it requires that the closure is not complete, but just dominant. The essential openness of autonomy\textsubscript{3} allows that functionality can depend on future states involving constructed closures, despite the fact that autonomy is itself an invariant. This is especially relevant in dealing with unexpected circumstances (Collier 2001) in which the autonomy of a system is reshaped by impinging influences so that similar influences can be accommodated in the future. The maintenance of autonomy under such conditions requires the construction of a state that maintains autonomy (and thus functionality) into the future. Autonomy\textsubscript{2} is too closed; it does not allow for organized and organizing interactions between organisms and their environments. Operational closure and closure to efficient causation, if complete, make living systems mechanistic in their own operation, though this mechanism is not fully commensurable with the processes that underlie them and/or impinge upon them. Autonomy\textsubscript{3} allows for the autonomous system to be open in its own logic, as well as being open to energy and matter flows. Recent developments in cognitive science pointing towards distributed cognition and its advantages suggest that living also involves distributed organization, since it would have the same advantages.

**SIMULATION**
What does this mean for simulation? First, behavioural simulation cannot be sufficient, since any finite set of behaviour of a strongly anticipative device can be duplicated with a weakly anticipative device. Nonetheless, just as behaviourism has been very productive as a technology in psychology, behavioural simulation is no mean feat, and helps us to understand the problems involved in creating artificial life. Second, autonomy is too weak for Alife cases, since it can be reduced to cases of weak anticipation. Inasmuch as it differs from the stronger versions of autonomy, ultimately it is equivalent to behavioural simulation. Nonetheless, the idea of an internal model used for predicting internal states is a useful one because such models can approximate more realistic cases of strong anticipation. For example, a robot that has a model of the physics of its own sensors and articulators and will be able to stream data from input to output much more efficiently than one that has to move, then correct, and so on (see, for example Bongard et al 2006). It is possible to get similar smooth behaviour by using mappings from inputs to outputs in the manner of Braitenberg’s (1984) and Brooks’ robots, but these devices are known to be fairly inflexible and difficult to scale, and results are disappointing, even for those working in the field (Brooks 2001). Internal models of the physics of the robot would enable greater flexibility, and are should be pursued. Living systems have a very integrated “body-sense” that is worth imitating. However, if Rosen., Maturana and Varela are right about closure and nonreducibility, then this approach will eventually come up against limitations.

The distinction between autonomous anticipation and derivative or weak anticipation might seem to be of only theoretical interest, given our present design abilities and the capacities of current computers. The distinction should be kept in mind, however, or else the capacities of systems with the only derivative anticipation are likely to be overestimated, and those of systems
with the strong anticipation (our goal for modelling) to be underestimated. This can lead to overly optimistic claims for derivative systems, and disappointing results that might harm the whole project of developing autonomous anticipative systems, derivative or not. In the longer run, experiments with self-maintaining robots and the development of autonomous machines may lead to trainable robots.

Although autonomy cannot in general be reduced, the fundamental differentiation of internal and external processes has no common dynamical model. One could set up an artificial distinction in a dynamical model, but that is ad hoc, rather than a property of the model itself. This may be sufficient for simulation. Since the complete closure implies a mechanism underlying autonomy, it may well be possible to imitate living organization and superimpose it on dynamically based robots of the Braitenberg and Brooks sort. This would be a development of the combined models described above, but the problem of internal modelling is not to make more directly effective connections between sensors and actuators and actual behaviour, but to get this as a side effect of modelling living organization. To take a well-worn example, the flow of sound in speech is not divided into simple patterns consistent with syntax. In order to recognize language in the clamour of the world, Chomsky has proposed that there is a Universal Grammar (UG) that guides our learning of language. UG is, he says, a finite state machine, logically the most simple of machines, but powerful enough to impose an order on aural and visual stimulation so that language filters through. Another example is David Marr’s work on vision. Perhaps a similar approach could be applied as an overlay to dynamically grounded robots to make them both more flexible and efficient. One problem with Marr’s model is that it is based on workings of the optic cortex, and ignores the role of the cerebellum in the fine control of sensors and actuators in mammals. But one can add modules. Some version of computational models
overlaying dynamical systems, perhaps as multiple interacting modules comprising numerous layers of control, is a promising route towards simulating autonomy, and may be able to overcome the limitations of autonomy. Autonomy in a simulation would be real autonomy. The simulation of emergent properties typically requires a similar identity: simulated emergence is possible on contemporary computers, but it breaks down just at the critical points where emergence is possible. The breakdown is indicative of possible mathematical irreducibility at the limit, but this can be misleading. Mathematician Jonathan Smith (1999) has shown that the period doubling “route to chaos” via the Feigenbaum map can be represented by a convergent mathematical structure, so appearances can be deceiving. We do not presently know how to make accurate models of emergent systems in any general way. That doesn’t mean, though, that we can’t come up with some interesting specific cases. The best way to produce an autonomous device is to let it grow under the right conditions. Thus the problem of devising a self-modifying anticipatory device that can develop modifications even to what it can recognise and control is more analogous in some respects to horticulture than to mechanical manufacture.

One striking example of how we might model emergent systems is given by Peter Cariani (1993). Gordon Pask, in the 1950s, invented a device that could modify its electrical and electrochemical ferrous sulphate substrate to distinguish tones and magnetic fields. These capacities formed spontaneously through changes in the malleable ferrous sulphate substrate under exposure to appropriate stimuli, rather than being designed in from the beginning. Unfortunately, this work was not pursued further, perhaps due to the dominant computational model of mind. It does suggest, though, that one way to model emergent systems is to start with materials that are known to be able to exhibit emergence, and constrain them in ways that are
both biologically realistic and functional. It should be possible to simulate such processes on high capacity modern computers, but one should keep in mind the caveat that digital modelling can break down just where things start to get interesting. One promising approach along the above lines is due to Ikegami and Suzuki (this volume). Their second model uses chaotic population dynamics to induce a self-moving state, in which sensors and recursive coupling occur. A possible modification in line with Dubois’ conjecture would be to find a way to induce self-sensing and self-modelling of movement and perhaps location. Chaotic dynamics is the right sort of thing to permit self-organization and emergence, but it isn’t sufficient itself for the stronger forms of autonomy. Some version of modulating control at a higher level that is independent of specific movements, but interacts with sensors of movement and position of the system could enhance the autonomy and possibly permit more directed motion.

REFERENCES


Full article in *Biosystems* Vol 91/2 pp 346-354, reprints available from the author
