COMPUTATION AND THE EXPLANATION OF INTELLIGENT BEHAVIOURS: ETHOLOGICALLY MOTIVATED RESTART

GUGLIELMO TAMBURRINI AND EDOARDO DATTERI

Abstract. Computational theorizing is fruitfully pursued in the investigation of sensorimotor coordination mechanisms of simple biological systems, such as unicellular organism chemotaxis. These investigations undermine the sweeping "dynamicist" claim according to which intelligent and adaptive behaviours in biological systems are to be accounted for in terms of continuous dynamical systems. Moreover, these investigations suggest the opportunity of developing a more fine-grained framework for analyzing the hierarchical interplay between computational, dynamical, and hybrid models of adaptive behaviours in both biological systems and machines. Key epistemological issues arising in this context of inquiry are clearly identified in Turing's and von Neumann's early reflections on the computational modelling of intelligent behaviours and brain functions.

§1. Introduction. A variety of sensorimotor coordination mechanisms are being successfully modelled on the basis of continuous dynamical system approaches (Beer [1997]; Steinhage and Bergener [2000]; Turvey and Carello [1995]). This work is invoked as empirical support for a sweeping "dynamicist" thesis: intelligent and adaptive biological behaviours are to be ultimately accounted for in terms of continuous dynamical systems; properly computational investigations make approximate simulation tools available for dynamical theories but play no essential theoretical role (Port and Gelder [1995]). Similar claims about mathematical theorizing in cognitive ethology and biology at large can be found in (Beer [1995], Steels [1995]) and (Longo [2003]), respectively. These claims are critically examined here, in the light of theoretical models of simple sensorimotor adaptive behaviours. These case studies are particularly suited to our purposes, for continuous dynamical approaches are supposedly at their best in the modelling of sensorimotor coordination mechanisms. Computational approaches, we submit, are being fruitfully pursued there too.

This critical examination suggests the opportunity of developing a more detailed analysis of interactions between (and respective roles of) computational

We gratefully thank an anonymous referee, Ernesto Burattini, and Franco Giorgi for valuable criticisms and suggestions.

Logic Colloquium '04
Edited by A. Andretta, K. Kearnes, and D. Zambella
Lecture Notes in Logic, 29
© 2006, Association for Symbolic Logic

and ~~dynamical approaches to the investigation~~ of intelligent and adaptive behaviours. A key epistemological problem arising in this context of inquiry was already identified by Alan Turing, in his early reflections on computational intelligence. In particular, Turing distinguishes between "continuous" and "discrete" machinery:

> We may call a machine 'discrete' when it is natural to describe its possible states as a discrete set, the motion of the machine occurring by jumping from one state to another. The states of 'continuous' machinery on the other hand form a continuous manifold and the behaviour of the machine is described by a curve on this manifold.[1]

Strictly speaking, Turing claims, there is no genuinely discrete machine: everything moves continuously, and thus every machine should be ultimately regarded as a "continuous" system. He is careful to add, however, that "there are many kinds of machine which can profitably be *thought of* as being discrete state machines". John von Neumann's open-minded speculations on computational modelling of brain functions reflect a strikingly similar viewpoint. Although analog processing is a pervasive feature of brain processing, for some functional modelling purposes this feature may turn out to be an *inessential complexity* (Von Neumann [1958]), p. 58.

Turing's and von Neumann's remarks point to the right kind of questions about the interplay between ~~computational and dynamical approaches~~. When is the discrete-state machine viewpoint profitably adopted to theorize about mechanisms underlying intelligent and adaptive behaviours? When is analog processing an inessential complexity of such mechanisms, which may be disregarded without loss of empirical and explanatory adequacy?

Only a few years before Turing and von Neumann presented these reflections, pioneers of Cybernetics Arturo Rosenblueth and Norbert Wiener developed a sensible epistemological framework for an examination of the above questions (Rosenblueth and Wiener [1945]). One may attempt to locate within this broad framework a proper role for "discrete-state machine" theorizing, for in this framework relatively few requirements are set on kinds of machines to be used in the investigation of adaptive behaviours (apart from the familiar cybernetic emphasis on feedback). ~~Accordingly,~~ Moreover, the pioneers of Cybernetics emphasized the hierarchical organization and the explanatory role of "box-and-arrows" theoretical models of adaptive biological

[1] See (Turing [1992]). See also (Turing [1950]) p. 439: "The digital computers considered in the last section may be classified amongst the 'discrete state machines'. These are the machines which move by sudden jumps or clicks from one quite definite state to another. These states are sufficiently different for the possibility of confusion between them to be ignored. Strictly speaking there are no such machines. Everything really moves continuously. But there are many kinds of machine which can profitably be *thought of* as being discrete state machines." For discussion of this distinction in connection with the closely related digital-analog distinction, see (Trautteur [2005]).

behaviours. Finally, significant experimental uses of machines are suggested for the purpose of testing and revising these theoretical models.[2] Let's see.

Please check the short running title.

§2. The ~~c~~Cybernetic legacy: mechanistic models and machine experiments. Significant machine-based modelling efforts are being carried out in cognitive ethology. A graphic example is the investigation, carried out by means of robotic models and experiments (Lambrinos *et al.* [2000]), of the homing behaviour of desert ant *Cataglyphis*, which is capable of reaching its nest following a straight path after wandering for dozens of meters in search of food.

Cybernetics addressed the problem of explaining this sort of adaptive system capacities (why is desert ant *Cataglyphis* capable of straight-line homing, even when the nest is not in the range of its sensory apparatus?) and behavioural events (why did ant #3 follow a certain trajectory on trial #6?) by means of special kinds of interconnected regularities, that we call here *mechanism schemata* ~~*or mechanism descriptions*~~[3]. Cybernetic mechanism schemata are theoretical models grounded into functional analyses of behaviour, to be pursued from an initial "closed box" view of the system S under investigation (Rosenblueth and Wiener [1945]). A closed box view of S is adopted when the relationships between a fixed finite number of its input and output variables are investigated irrespective of what processes S carries out in order to bring about these relationships (whence the closed box metaphor). Rosenblueth and Wiener are careful to point out that a closed-box approach presupposes that there is only "loose coupling" between the investigated variables on the one hand, and the other system variables on the other hand. Loose coupling ensures that the interactions of the selected variables with other system variables are negligible for approximate modelling purposes[4].

A closed box can be "opened" by iteration of closed box analysis. This iteration may enable one to identify a mechanism schema formed by functionally characterized and interconnected parts: one divides the original closed box into "smaller" compartments, thereby fixing their contribution to overall system behaviour by means of their mutual I/O interconnections. Deciding how far to pursue the "opening" of some closed box depends on various factors, including current theoretical interests and system variables that happen to be measurable by available techniques.

The explanatory adequacy of functional mechanism schemata depends on the following features. First, mechanism schemata enable one to provide a *causal* account for how some behavioural events are produced or why the

[2] See (Tamburrini and Datteri [2005]) for more extensive discussion of Cybernetic methodology and its import on contemporary biorobotics.

[3] This terminology dates back at least to (Nagel [1961]) Chapter 8.

[4] The notion of nearly decomposable system, in the sense of (Simon [1996], p. 197 ff.), provides a somewhat more precise understanding of Rosenblueth and Wiener's idea that "a number of variables are only loosely coupled with the rest of those belonging to the system".

system is endowed with some behavioural capacities. Second, overall system behaviour is accounted in terms of the behaviour of *simpler* functional parts. Third, *unified* understanding is achieved, insofar as the same causal factors account for behaviour observed in structurally different systems. Indeed, the characteristic unification promise of Cybernetics is based on the identification of mechanism schemata harnessing the behaviour of both living organisms and machines, that is, the behaviour of systems differing from each other in their "structural" properties (Rosenblueth *et al.* [1943]). In passing, let us notice that abstraction from material support of behaviour is also invoked by Turing in order to go beyond shallow analogies between computers and nervous systems[5].

Cybernetic mechanism schemata are naturally accommodated by the mechanistic model of explanation (Machamer *et al.* [2000]; Glennan [2002]), but are significantly related the covering-law model (Hempel and Oppenheim [1948]) of scientific explanation too, insofar as cybernetic schemata are theoretical descriptions taking the form of coordinated *regularities*. Ernest Nagel, a major advocate of the covering law model, went as far as to assert that Cybernetic accounts of homeostatic system behaviours are special cases of covering-law explanations: "all such systems can in principle be analyzed into parts which are causally relevant to the maintenance of some feature in those systems, and which stand to each other and to environmental factors in determinate relations capable of being formulated as general laws. The discovery and analysis of such detailed mechanisms is the task of specialized scientific inquiry" (Nagel [1961], p. 418).

The uniform functional analysis of both machines and living organisms sought for by Cybernetics paves the way to a principled use of machines in experiments about mechanism schemata. A material model is the representation of a complex system by another system "which is assumed simpler and which is also assumed to have some properties similar to those selected for study in the original complex system". (Rosenblueth and Wiener [1945], p. 317). The crucial relationship between the original complex system and its material model is one of similarity. In this context, distinct material systems are similar to the extent their functional organization is harnessed by the same abstract theoretical model. Accordingly, Rosenblueth and Wiener are careful to point out that the basis for passing similarity judgments between the original system and its material model is provided by some theoretical model, e.g., a mechanism schema, of the original system. Similarity is a theoretical

[5] "Importance is often attached to the fact that modern digital computers are electrical, and that the nervous system also is electrical. . . . In the nervous system chemical phenomena are at least as important as electrical. In certain computers the storage system is mainly acoustic. The feature of using electricity is thus seen to be only a very superficial similarity. If we wish to find such similarities we should look rather for mathematical analogies of function." (Turing [1950], p. 439).

notion, and similarity judgments are always relative to some sets of properties: the theoretical model is the yardstick in terms of which similarity judgments are issued.

A material model singled out on the basis of some mechanism schema can be used to test the empirical adequacy of the mechanism schema, insofar as it may enable one to carry out "experiments under more favorable conditions than would be available in the original system". The similarity between material model and original system is a crucial *empirical assumption* underlying these experimental enterprises. In addition to corroboration and falsification, material models may enable one to obtain results which "could not have been easily anticipated on the basis of the formal model alone" (Rosenblueth and Wiener [1945], p. 317).

In our view, this broad cybernetic picture of the relationship between theoretical modelling, explanation, and machine experiments provides a sensible methodological framework for much work in contemporary biorobotics, which is characteristically concerned with the problem of conducting experiments on adaptive biological behaviours by means of robotic material models (Tamburrini and Datteri [2005], Datteri and Tamburrini [2005]). Let us now turn to compare, within this overall methodological framework, ~~continuous dynamical system approaches with the "discrete state machine" viewpoint~~ in the theoretical modelling of simple sensorimotor coordination mechanisms.

§3. Production-rule mechanisms and simple organism chemotaxis.

3.1. Bacterial chemotaxis. Bacterial chemotaxis has been extensively studied since the end of the 19th century. It can be observed by pushing a capillary tube containing a solution of glucose into a suspension of *E. coli* on a slide; bacteria will massively go inside the tube (Adler [1969]). By considering only bacteria positions and chemical concentrations, the behaviour of E. coli is readily modelled as a gradient ascent: bacteria go towards the region of maximum chemical concentration (Staddon [2003], Chapter 2), where the quantities involved (positions and chemical gradients) are naturally represented as continuous state variables.

Further observation of *E. coli* behaviour reveals additional features of their chemotaxis capacities (Bourret and Stock [2002]; Sourjik [2004]; Webre *et al.* [2004]). *E. coli* exhibit *only two* actions: they swim in a straight line or tumble down. The relative frequencies of swimming and tumbling are the key to understanding how *E. coli* finds a chemical source. In solutions characterized by uniform, non-changing attractant concentrations, bacterial motility exhibits typical patterns: bacteria swim in fairly straight lines; occasionally they tumble and start swimming in completely new directions. When an attractant is inserted in the solution, bacterial motility appears much smoother and better co-ordinated than normal: the frequency of straight-line swimming increases with respect to tumbling and randomly changing direction.

This produces an overall motion upward along the concentration gradient. A few minutes after stopping the flow of attractant, the frequency of tumbling is restored to its steady-state value and bacterial motility shows the typical patterns described above. Negative attractant gradients produce opposite behaviours: tumbling increases over straight-line swimming, and bacterial motion appears erratic and non-coordinated (Macnab and Koshland [1972]). Chemotaxis is thus achieved by switching between two distinct actions, tumbling and swimming.

Let us consider current explanations of how these two actions are produced. Each *E. coli* is endowed with some flagella (five to ten per cell). Flagella can rotate clockwise or counterclockwise, producing tumbling and smooth swimming respectively. Transmembrane receptors bind to attractant molecules producing a chain of protein phosphorylation processes, that ultimately leads to variation in the concentration of phosporylated protein CheY (CheY-P). The latter binds to protein FliM, which is thought to be a "switch" protein, responsible for changes of flagellar motion direction (Bren and Eisenbach [2001]). The resulting behaviour appears to be governed by simple rules:

> The question of how the chemotaxis system is regulated [...] reduces to that of how the direction of flagellar rotation is controlled. The default direction of rotation of the flagellar motor is counterclockwise. A switch to rotation in the clockwise direction occurs in response to a sensory signal, transduced by the chemotaxis protein CheY (Eisenbach and Caplan [1998]).

Two rules are included for two distinct and mutually exclusive behaviours. This hypothesis is also at the core of a computer simulation of bacterial chemotaxis, which takes into account a wide variety of processes that are known to occur within the cell, from membrane receptors to motor activation via phosphorylation signalling pathways (Bray *et al.* [1993]). In their model, the state assumed by the flagellar motor complex M (run, pause, and tumble) depends on different binding possibilities between M and phosphorylated CheY: M can bind up to four different phosphorylated CheY molecules, thus resulting in five distinct molecules. Counterclockwise motion is elicited by two of these molecules (M alone, and M bound with just one CheY-P molecule); motion pause is elicited by the binding of M with two CheY-P molecule; the other two combinations elicit clockwise behaviour. Predictions obtained with this computer simulation match with observations in many significant cases[6].

In the simple and well-studied case of bacterial chemotaxis, the "same" system (*E. coli* exemplars) can be modelled either as a discrete or else as

[6]The two-actions behaviour of bacterial chemotaxis has served as an inspirational source for a recent robotic system, capable of locating gradient sources, and applied to the field of environmental monitoring (Dhariwal *et al.* [2004]).

a continuous-state mechanism. Which is chosen as the more suitable explanatory source depends on the character of the *explanandum*. If one wants to understand the relation holding between chemical gradients and *E. coli* positions, a ~~continuous-state~~ approach may be preferred over a discrete one. Accordingly, quantization of gradients and positions in space — introduced, say, for practical reasons concerning the handling of simpler mathematical or simulation tools — is regarded only as an *approximation* of quantities that, in the theoretical account of the system, take values in the real field. Discrete-state theoretical models provide empirically supported explanations of the relations holding between the reception of chemical molecules by *E. coli* transmembrane proteins, and the direction of flagellar motion. In the model reviewed above, the flagellar behavioural repertoire of *E. coli* is represented as a discrete set of mutually exclusive actions; quantization is empirically (rather than practically) justified, insofar as the discrete theoretical model is supported by empirical testing. Further opening of the boxes into which this discrete model functionally decomposes (for example, by looking at the chemical dynamics involved in binding and reaction processes occurring within the cell), may induce one to switch to a continuous-state approach again. In specific empirical investigations, discrete state machine descriptions and continuous dynamic system descriptions may happen to be interleaved in hierarchical, multilevel mechanism schemata. Which level of the hierarchy is chosen as an appropriate source of explanations depends on the nature of explananda and pragmatic considerations about explanation needs.

Similarly, production-rule mechanisms operating on quantized variables may have to be combined with other mechanisms in a horizontal (i.e. non-hierarchical) fashion, in order to capture some biological action selection mechanism.[7] The opportunity of introducing such hybrid mechanisms may emerge even under the assumption - embedded into models of *E. coli* chemotaxis - that actions are mutually incompatible. For example, McFarland's hybrid models of animal behaviour (McFarland and Sibly [1975]) rest on a similar strong assumption:

> it is always possible to classify the behavioural repertoire of a species in such a way that the classes are mutually exclusive in the sense that members of different classes cannot occur simultaneously

Animals do *only* action x at time t, where x can be classified as an *unit of*

[7] "Imagine a zebra in the African savannah. At each moment in time this zebra has to weigh up alternative courses of action before deciding which will be most beneficial to it. For instance, it may want to graze because it is short of food, or it may want to head towards a water hole because it is short of water, or it may want to remain motionless in order to avoid detection by the predator it can see lurking nearby. This is an example of the problem of *action selection*: how to choose, at each moment in time, the most appropriate out of a repertoire of possible actions" (Tyrrell [1993]).

behaviour, in McFarland's words. Units of behaviours are identified on the basis of a number of their distinctive features. McFarland is not alone in this suggestion. Tyrrell, Hull, Tinbergen, Lorenz and Baerends propose models of action selection conforming to this oft-maintained assumption, insofar as the behavioural repertoire is given by a finite set of mutually incompatible actions. In these theoretical models, however, variables are included which range over some real interval and participate in various ways to the selection of a winning action. McFarland's causal factors are a significant case in point, as well as Hull's *drives*, construed as internal motivation variables associated to behaviours (Hull [1943]). Each behaviour has a drive, whose magnitude is determined by the strength of internal and external stimuli and by other factors that change with learning. One action is chosen from a set of competing actions on the basis of its drive strength. Baerends is more explicit in postulating a hierarchical organization of behaviours in his study of digger wasp and herring gull: the animal's 'higher-level instincts' activate some 'lower-level instinct', which in turn activates some 'fixed action pattern' (Baerends [1976]).

These various modelling enterprises suggest that an adequate epistemological analysis of the interplay between ~~computational and dynamical~~ approaches requires a detailed study of the hierarchical interleaving of continuous and discrete mechanism descriptions as well as their "horizontal" combination into hybrid mechanism descriptions. This detailed study goes clearly beyond the limits of this paper, but we can here strengthen this motivation by reference to a historically prominent case study.

3.2. Stentor. In *Behaviour of the low organisms*, H. S. Jennings proposes a mechanism schema to explain the behaviour of Stentor, a unicellular Protozoan organism. Stentor resembles a "trumpet" with cilia surrounding its larger opening (Jennings [1906]). By moving the cilia, Stentor gives rise to a water current that carries food into its body. Jennings accounts for the behaviour of Stentor when a noxious chemical tray reaches the opening in terms of a simple adaptive mechanism. This mechanism is adequately described by means of a few production rules[8]. Each production rule encodes, in its action part, a different reaction to the noxious stimulus. And which action is chosen (which rule wins) depends on the organism's past history.

Let us consider Jennings' model in more detail. After sensing the noxious tray,

> [t]he animal bends to one side ... It thus as a rule avoids the cloud of particles, unless the latter is very large. [...] If the first reaction is not successful, it is usually repeated one or more times. If the

[8] Jenning's discrete-state model of Stentor has been discussed in (Cordeschi [2002]), in the context of machine-based investigations on adaptive biological behaviours.

> repeated turning to one side does not relieve the animal, so that the particles of carmine continue to come in a dense cloud, another reaction is tried. The ciliary movement is suddenly reversed in direction, so that the particles against the disk and in the pouch are thrown off. [...] If this fails to relieve the organism, the next reaction — contraction — usually supervenes.

Finally, if the stream reaches the cilia again, Stentor tries another very expensive strategy:

> it ceases to extend, but contracts repeatedly and violently while still enclosed in its tube. In this way the attachment of its foot to the object on which it is situated is broken, and the animal is free. Now it leaves its tube and swims away.

Jennings isolates a number of stereotyped actions (turning away; ciliary reversal; contract into tube; break and go away), all of them being possible responses to the same stimulus. The actions are tried sequentially (see also Staddon [2003], Chapter 3).

How does Stentor chooses the action to perform from its fixed repertoire $a_1, \ldots, a_n$ of actions when the noxious stimulus s is present? Jennings postulates the existence of some set of distinct physiological states, or "varying internal conditions of the organism". This is the reason why Stentor reacts to the same stimulus in different ways, depending on its current physiological status:

> After this [*the chemical stimulation*] has taken place once or twice we find that the Stentor no longer reacts as before. All the external conditions remain the same, the stimulus applied is the same. Nevertheless, the Stentor does not react. Therefore, we must conclude that the Stentor itself has changed. Its physiological condition is now different from what it was original.

A convenient production-rules representation of Jenning's account takes the form:

$$\mathrm{s} \wedge \mathrm{p_1} \rightarrow \mathrm{a_1}$$
$$\ldots$$
$$\mathrm{s} \wedge \mathrm{p_n} \rightarrow \mathrm{a_n}$$

where each p_i, $1 \leq i \leq n$, is a physiological state taking into account Stentor's recording of recent interactions with the environment.

By means of this mechanism, Jennings can answer explanation requests about Stentor's behaviour, concerning, e.g., the observation that Stentor cannot perform more than one action at a time, or the causal factors determining its action selection. Whether this mechanism is a source of contextually appropriate explanations depends on the character of explanation requests. If one

wants to understand, down to biochemical details, how sensory stimuli are processed by Stentor, and eventually affect its motor system, the functional accounts one may provide on the basis of this production-rule mechanism turn out to be unsatisfactory. A more satisfactory explanation will most likely involve a hierarchy of mechanisms, with some lower-level mechanism specified in terms of biochemical entities and processes which give rise to the behaviour functionally characterized by the higher-level production-rule mechanism. Once again, ~~discrete state machine descriptions and continuous dynamic system descriptions~~ may turn out to be interleaved in a hierarchy of mechanism schemata.

The opportunity of developing a fine-grained framework for analyzing hierarchical and horizontal combinations of ~~computational and dynamic approaches to the investigation~~ of adaptive behaviours emerges also from an examination of biologically inspired robotics. Sweeping dynamicist theses were advanced in this area of investigation too, especially building on the development of behaviour-based systems, which are both inspired to and provide heuristic insight for functional models of adaptive biological behaviours. Dynamicist theses stand in need of serious qualifications and amendments in this case too. Let's see.

§4. Biologically inspired robotics: ~~computational and dynamical~~ models.

4.1. ~~**A dynamicist thesis about behaviour-based systems.**~~ The broad heuristic idea that many adaptive behaviours are adequately modelled and explained in terms of continuous ~~dynamical systems~~ is a source of inspiration for much recent work in bio-robotics (Beer [1995], [1997]) and biologically inspired robotics (Steels [1995]). To exemplify, let us consider ~~continuous dynamical system~~ approaches in the framework of behaviour-based architectures. These are also used for modelling adaptive capabilities of many biological systems (from insect navigation to human limb control).

The functional structure of behaviour-based systems (Arkin [1998], Brooks [1986]) is usually described in terms of a set of m interconnected layers, called *behaviours*, each layer being associated to a function from sensor data to motor actions. ~~T~~The motor response r_i of behaviour b_i ~~at time t~~ is

$$r_i = \gamma_i \times b_i(s_t) ,$$

We added the highlighted comma. Please check.

where s_t is the state perceived by the sensor associated with behaviour b_i at time t, γ is a real-valued gain, and $0 \leq i \leq m$. The system is parallel and asynchronous. If n behaviours control the same actuator, a coordination function C, taking as input the outcomes of those n behaviours, outputs a command for the actuator (see Figure 1). A typical coordination function (termed cooperative) blends together (typically sums) multiple responses, previously weighted by means of some set of n real-valued gains. Function C, together with the

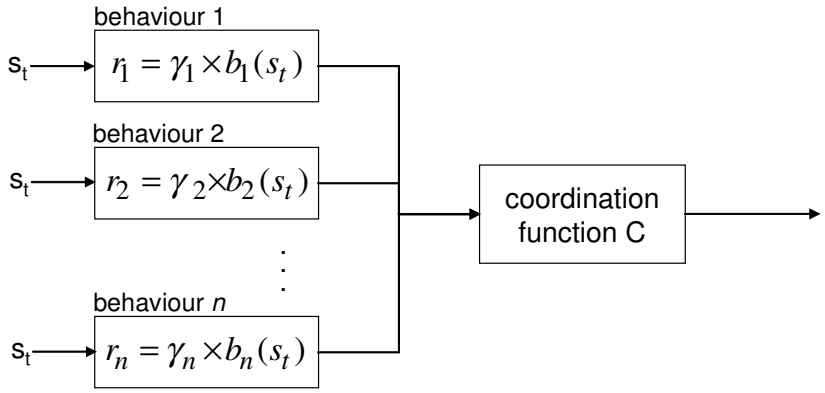


FIGURE 1. Functional structure of behaviour-based architectures.

various gains and mappings characterizing each behaviour, determines how the system acts in response to perceived environmental conditions. Let f be the overall agent control function. Distinct sets of gains determine distinct reaction policies to environmental stimuli, i.e. distinct control functions.

A continuous-time, dynamical-system approach has been proposed as unifying treatment for behaviour-based systems. Notably, Steels (Steels [1995]) focuses on control programs in behaviour-based systems, and claims that these are best viewed as "dynamical systems which establish a continuous relation between the time varying data coming out of a sensor and a stream of values going to the actuators". Furthermore, he proposes to apply equilibrium analysis to the study of behaviour-based systems: a behaviour "reaches an equilibrium state if it exerts no influence on the dynamics" of the whole dynamical system, i.e. of the whole assemblage of behaviours. By means of equilibrium analyses one may identify, for example, whether a behaviour will reach equilibrium or the conditions under which a behavioural assemblage will reach some configurations in the state space.

Let us proceed to a critical examination of this approach to behaviour-based architectures in the light of adaptive robot navigation problems.

4.2. Symbol processing and potential field navigation. Adaptive navigation requires the capacity of orienting oneself to goals (whose position is directly perceived or known), while avoiding static or dynamic obstacles. Basic manifestations of these capacities, and their combination, are readily modelled in the framework of some behaviour-based system A, which is provided with the following behaviours:

- One behaviour, called *attractive*, is mathematically represented as a vector, whose direction points towards the goal, and whose magnitude is directly proportional to the distance between current point and goal or some sensory cue. For example, the attractive behaviour of a phototropic organism might be represented by a mapping from the perceived orientation of a light source to the attractive vector orientation, and by a mapping from light intensity to the attractive vector magnitude;
- another behaviour, called *repulsive*, is represented as a vector whose direction points away from perceived obstacles, and whose magnitude is inversely proportional to the distance between robot and obstacles. For example, if some distance sensors are available, their orientation and distance reading might be mapped to the two components of the repulsive vector.

A cooperative function sums the two vectors; at each given time, A's orientation is the orientation of the resulting vector, and its velocity is the vector magnitude. The contribution of the *repulsive* behaviour prevents A from colliding with obstacles, while the *attractive* behaviour directs it to the goal.

This modelling strategy has been more comprehensively formalized as the *potential field* method of navigation, according to which the system is represented as a point in space, acting under the influence of an *artificial potential field* U (Latombe [1991]). This potential field is typically defined as the vector sum of an attractive and a repulsive potential. More precisely, given an agent A (a single rigid object) and an environment E (a Euclidean space represented as R^2 or R^3), at each point q in space $U(q) = U_{att}(q) + U_{rep}(q)$. Typical equations for calculating the magnitude of attractive and repulsive potentials are

$$U_{att}(q) = \frac{1}{2} k \rho^2 \, goal$$

$$U_{rep}(q) = \begin{cases} \frac{1}{2} j \left(\frac{1}{\rho}(q) - \frac{1}{\rho_0} \right) & \text{if } \rho(q) < \rho_0 \\ 0 & \text{if } \rho(q) \geq \rho_0 \end{cases}$$

where ρ is the Euclidean distance between q and the goal, k and j are positive scaling factors, and ρ_0 is a positive constant called the *distance of influence* of the obstacles. The potential field approach is conceptually akin to Hull's goal-gradient hypothesis (Hull [1932]). Several computer simulation or robotic experiments have shown its significance in the modelling of adaptive navigation and adaptive interaction with the environment, over and above the robotic navigation domain (Khatib [1986]).

This navigation method shows that some navigation behaviours are trajectories predictable on the basis of some simple ~~dynamical~~ law; and in turn, such

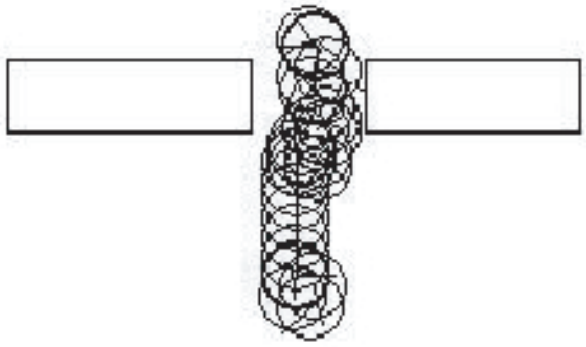


FIGURE 2. The robot gets stuck within the door aperture.

~~dynamical~~ laws can be involved in the specification of a black-box mechanism schema, insofar as these laws identify the I/O behaviour of the attractive and repulsive modules of the potential-field system. Let us now explore limitations of potential-field navigation methods and the opportunity of combining this method with symbol processing mechanisms, by reference to experiments involving robotic navigation behaviours.

Let us consider a potential-field based system, provided with some kind of distance sensor (like sonars or infrared). Koren and Borenstein (Koren and Borenstein [1991]) identify some of the main drawbacks of potential-fields models in cyclic behaviour due to local minima, no passage between closely spaced obstacles, oscillations in the presence of obstacles, and oscillations in narrow passages. These behaviours are claimed to be independent of implementations. And Latombe (Latombe [1991]) claims that

> [...] local minima remain an important cause of inefficiency for potential field methods. Hence, dealing with local minima is *the* major issue that one has to face in designing a planner based on this approach.

Even doors constitute a difficulty for behaviour-based agents based on potential fields. As Koren and Borenstein explain, depending on the size of the aperture (assuming it is larger than the robot body), on the depth of the walls, and on the position of distance sensors on the system, doors are perceived as walls, and the robot cannot go through them. Figure 2 shows superimposed snapshots from a simulation of a sonar-based robot which is commonly used

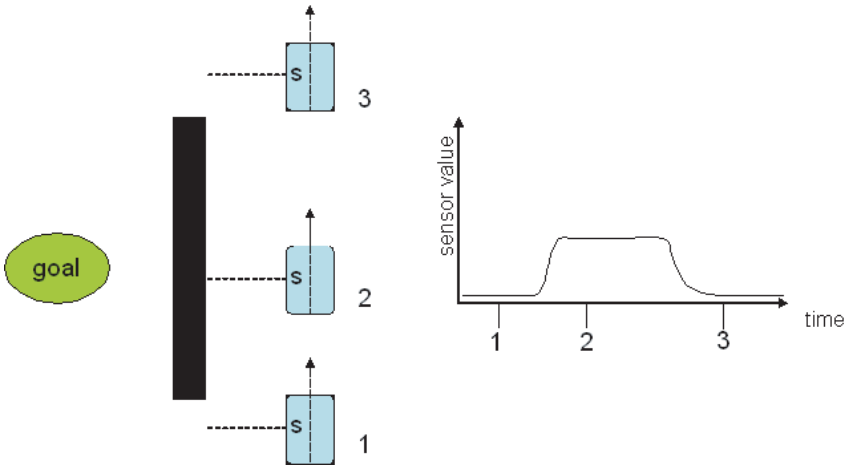


FIGURE 3. Values of sonar readings when the robot walks close to a wall.

in robotic investigations. This robot, represented by a circle, gets stuck within a door aperture.

To overcome problems of this sort, one may try and determine action on the basis of data that go beyond instantaneous sensory readings. A promising approach requires fitting the robotic system with an event detection mechanism based on sensor histories. Such event-detection mechanisms can give rise to genuine symbol-processing and more sophisticated planning behaviour. Let's see.

The graph on the right-hand side of Figure 3 shows the values signalled over time by a sonar of a behaviour-based robot. By considering sensor history, the presence of a wall can be hypothesized and distinguished from the presence of some, say, round obstacle: in the case of walls, the sensor curve typically shows three distinct periods, with a time lag of "high" value interleaved between two periods of "low" value.

This observation is the basis for devising a history-based mechanism, allowing the robot to perceive, in addition to local (current) potential fields, the presence of particular obstacles configurations, like walls or doors. Behaviour-based systems provided with this mechanism can modify their behavioural arrangement by adjusting their behavioural gains or even by completely suppressing behaviours[9] in order to cope with doors or walls. For example, as

[9]Inhibition of behaviours is characteristic of a particular coordination function (see Section 4.1), termed *competitive* control function.

soon as the perceptual apparatus detects the presence of a wall on the left, by detecting the characteristic sensor curve shown in the figure above, the system may adopt the following strategy in order to go around the wall: suppress for a while the attractive behaviour (by setting to 0 the corresponding gain), and activate at the same time a *wander* behaviour, directing the system in random directions; reactivate, after some time lag, the attractive behaviour. The *wander* behaviour is a source of noise, allowing the robot to escape from the local minimum induced by the wall-target configuration.

One can encode the first segment of the new navigation strategy as a production rule:

We added the highlighted period. Please check.

$$\text{wall_left} \rightarrow (\text{on}(\text{wander}) \wedge \text{off}(\text{attractive}))$$

Similarly, if the perceptual system has the capacity of detecting doors, it can decrease the strength of the repulsive behaviour and increase the strength of the attractive behaviour, thus allowing the robot to escape from stalemate.

This strategy has been expanded and implemented in (Schönherr *et al.* [2001]), where histories of behavioural activation values are tracked and grouped in stereotyped patterns, corresponding to particular environmental situations. Behaviours are defined as in Section 4.1, and activation values correspond to the γs in the definition: behaviour activation values indicate the strength of their motor actions in relation to the magnitude of sensory inputs, and can be changed over time by other system modules (or behaviours), in order to inhibit, increase or completely suppress activity. The approach of (Schönherr *et al.* [2001]) is only apparently different from tracking *sensor values*, insofar as behaviour activation values strictly depend on sensor values, and are changed according to the following differential equation[10]:

We added the highlighted comma. Please check.

$$\dot{\gamma}_b = g_b(\gamma_b, On_b, Off_b)$$

where γ_b is the activation value for behaviour b, g_b is a nonlinear function, On_b and Off_b represent the contribution of conditions positively and negatively related with the activation of b. The latter two parameters depend on the behaviour sensor values.

The usefulness of monitoring activation values for achieving adaptive behaviour is illustrated by an example, involving a behaviour-based system. Six behaviours are used (*turn-to-door*, *go-through-door*, *follow-wall*, *avoid-collision*, *wander*); three additional, "higher-level" behaviours are taken into account, whose outputs do not affect motor commands, but rather change the value of some other global variables. The experimental task is to pass through a door. The vector of the six activation values shows typical patterns in time, in correspondence with specific robot/environment configurations.

[10]This is a slightly simplified version, omitting variables connected to higher-level modules of the system.

For example, the sequence of behaviours activation values when the robot is in the vicinity of the door, goes through it and then exits, is characterized by

- a initial increase of one of the *follow-wall* behaviour;
- a subsequent increase of the *close-to-door* behaviour;
- during high activation of the *close-to-door* behaviour, a brief activation of *turn-to-door* followed by the simultaneous activation of *go-through-door* and *avoid-collision*.

Behavioural sequences of this kind are adequately represented by means of a suitable first-order formalism, and even simpler production rules such as

$$(\mathrm{HIGH}(\text{wall-left}) \text{ or } \mathrm{HIGH}(\text{wall-right})) \rightarrow (\mathrm{ON}(\text{wander}) \wedge \mathrm{OFF}(\text{attractive}))$$

suffice to overcome local minima difficulties arising with doors and walls. During robot motion, behavioural activity is logged continuously on in order to detect the occurrence of certain patterns. This pattern matching mechanism allows the system to extract "facts" from activation values, and to select actions on this basis. Note, however, that more than mere pattern recognition is required to qualify a perceived chronicle as a fact:

> The fact extraction technique does not presume or guarantee anything about the consistency of the facts that get derived over time. Achieving and maintaining consistency, and determining the ramifications of newly emerged facts remain issues that go beyond fact extraction. (Schönherr *et al.* [2001]).

Perception in these behaviour-based systems naturally takes the form of a hypothesize-and-test process: incoming chronicles are tested against the database of accepted facts and, possibly, against other facts that are expected in virtue of accepted facts. As long as temporal constraints characterizing adaptive action in the world can be properly met by the required reasoning systems, one has here a clear indication of the opportunity of combining symbol processing with continuous ~~dynamical system modelling in~~ behaviour-based architectures.

Various modes of combination are suggested by a closer examination of the examples provided so far. The event-detection module of the behaviour-based system described in this section is appended, in a cascade, to sensory modules whose state variables are assumed to range over some real interval. The cases of *E. coli* and *Stentor* suggest that continuous and discrete state mechanisms can be used as theoretical models of the same system, at different levels of the "box-and-arrows" modelling hierarchy[11]. In general, biologically inspired robotics provides additional motives to relinquish sweeping dynamicist theses, and compelling reasons for developing a more fine-grained analysis

[11] A neuro-symbolic language for combining continuous-state neural networks with production rules in a hierarchical fashion is described in (Burattini *et al.* [2005]).

of both hierarchical and horizontal combinations of ~~computational and dynamic~~ approaches in the modelling of adaptive and intelligent behaviours. The selection of an appropriate explanatory source from these hierarchies of models (differing from each other in terms of functional primitives and organizations) involves pragmatic decisions that are motivated by particular explanation needs driving scientific inquiry.

REFERENCES

J. Adler [1969], *Chemoreceptors in bacteria*, ***Science***, vol. 166, no. 3913, pp. 1588–1597.

R. C. Arkin [1998], ***Behavior-Based Robotics***, The MIT Press.

G. Baerends [1976], *The functional organisation of behaviour*, ***Animal Behaviour***, vol. 24, pp. 726–735.

R. D. Beer [1995], *A dynamical systems perspective on agent-environment interaction*, ***Artificial Intelligence***, vol. 72, pp. 173–215.

R. D. Beer [1997], *The dynamics of adaptive behavior: A research program*, ***Robotics and Autonomous Systems***, vol. 20, pp. 257–289.

R. B. Bourret and A. M. Stock [2002], *Molecular information processing: Lessons from bacterial chemotaxis*, ***The Journal of Biological Chemistry***, vol. 277, no. 12, pp. 9625–9628.

D. Bray, R. B. Bourret, and M. I. Simon [1993], *Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis*, ***Molecular Biology of the Cell***, vol. 4, pp. 469–482.

A. Bren and M. Eisenbach [2001], *Changing the direction of flagellar rotation in bacteria by modulating the ratio between the rotational states of the switch protein flim*, ***Journal of Molecular Biology***, vol. 312, pp. 699–709.

R. A. Brooks [1986], *A robust layered control system for a mobile robot*, ***IEEE Journal of Robotics and Automation***, pp. 14–23.

E. Burattini, E. Datteri, and G. Tamburrini [2005], *Neuro-symbolic programs for robots*, forthcoming in the *Proceedings of NeSy'05 Neural-Symbolic Learning and Reasoning*, Workshop at *IJCAI-05*, Edinburgh, Scotland, August 1st, 2005.

R. Cordeschi [2002], ***The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics***, Kluwer Academic Publishers, Dordrecht, The Netherlands.

~~E. Datteri and G. Tamburrini [2005], *Bio-robotic experiments and scientific method*, forthcoming in *Proceedings of European Computing and Philosophy Conference (ECAP) 2004*, University of Pavia.~~

A. Dhariwal, G. S. Sukhatme, and A. A. G. Requicha [2004], *Bacterium-inspired robots for environmental monitoring*, ***Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2004)***, pp. 1436–1443.

M. Eisenbach and S. R. Caplan [1998], *Bacterial chemotaxis: Unsolved mystery of the flagellar switch*, ***Current Biology***, vol. 8, pp. R444–R446.

S. Glennan [2002], *Rethinking mechanistic explanation*, ***Philosophy of Science***, no. 69, pp. S342–S353.

C. G. Hempel and P. Oppenheim [1948], *Studies in the logic of explanation*, ***Philosophy of Science***, vol. 15, no. 2, pp. 135–175.

C. L. Hull [1932], *The goal-gradient hypothesis and maze learning*, ***Psychological Review***, vol. 39, pp. 25–43.

C. L. Hull [1943], ***Principles of Behaviour: an Introduction to Behaviour Theory***, D. Appleton-Century Company.

H. S. Jennings [1906], ***Behavior of the Lower Organisms***, Columbia University Press, New York.

O. KHATIB [1986], *Real-time obstacle avoidance for manipulators and mobile robots*, ***The International Journal of Robotics Research***, vol. 5, no. 1, pp. 90–98.

Y. KOREN AND J. BORENSTEIN [1991], *Potential field methods and their inherent limitations for mobile robot navigation*, ***Proceedings of the IEEE Conference on Robotics and Automation***, Sacramento, California, pp. 1398–1404.

D. LAMBRINOS, R. MÖLLER, T. LABHART, R. PFEIFER, AND R. WEHNER [2000], *A mobile robot employing insect strategies for navigation*, ***Robotics and Autonomous Systems***, vol. 30, pp. 39–64.

J. C. LATOMBE [1991], ***Robot Motion Planning***, Kluwer Academic Publishers, Norwell, MA.

G. LONGO [2003], *Computer modelling and natural phenomena*, ***Proceedings of the European Software Engineering Conference and ACM SIGSOFT Symposium***, Helsinki.

P. MACHAMER, L. DARDEN, AND C. F. CRAVER [2000], *Thinking about mechanisms*, ***Philosophy of Science***, vol. 67, pp. 1–25.

R. M. MACNAB AND D. E. KOSHLAND [1972], *The gradient-sensing mechanism in bacterial chemotaxis*, ***Proceedings of the National Academy of Science USA***, vol. 69, no. 9, pp. 2509–2512.

D. J. MCFARLAND AND R. M. SIBLY [1975], *The behavioural final common path*, ***Philosophical Transactions of the Royal Society of London Series B, Biological Sciences***, vol. 270, no. 907, pp. 265–293.

E. NAGEL [1961], ***The Structure of Science: Problems in the Logic of Scientific Explanation***, Routledge and Kegan Paul, London.

R. F. Port and T. van Gelder (editors) [1995], ***Mind as Motion: Explorations in the Dynamics of Cognition***, Bradford Books, The MIT Press.

A. ROSENBLUETH AND N. WIENER [1945], *The role of models in science*, ***Philosophy of Science***, vol. 12, pp. 316–321.

A. ROSENBLUETH, N. WIENER, AND J. BIGELOW [1943], *Behavior, purpose and teleology*, ***Philosophy of Science***, vol. 10, pp. 18–24.

F. SCHÖNHERR, M. CISTELECAN, J. HERTZBERG, AND T. CHRISTALLER [2001], *Extracting situation facts from activation value histories in behavior-based robots*, ***KI-2001: Advances in Artificial Intelligence, LNAI 2174***, pp. 305–319.

H. SIMON [1996], ***The Sciences of the Artificial***, The MIT Press, Cambridge, MA.

V. SOURJIK [2004], *Receptor clustering and signal processing in E. coli chemotaxis*, ***Trends in Microbiology***, vol. 12, no. 12, pp. 569–576.

J. E. R. STADDON [2003], *Adaptive behavior and learning*, Internet Edition (`http://psychweb.psych.duke.edu/department/jers/abl/TableC.htm`).

L. STEELS [1995], *Mathematical analysis of behavior systems*, ***Proceedings of the Conference "From Perception to Action" (Perarc)***, pp. 88–95.

A. STEINHAGE AND T. BERGENER [2000], *Learning by doing: A dynamic architecture for generating adaptive behavioral sequences*, ***Proceedings of the Second ICSC Symposium on Neural Computation NC2000, Berlin***, pp. 813–820.

G. TAMBURRINI AND E. DATTERI [2005], *Machine experiments and theoretical modelling: from cybernetic methodology to neuro-robotics*, forthcoming in ***Minds and Machines***.

G. TRAUTTEUR [2005], *Beyond the super-turing snare: analog computation and digital virtuality*, ***New Computational Paradigms*** (S. B. Cooper, B. Löwe, and L. Torenvliet, editors), LNCS, vol. 3526, Springer, Berlin, pp. 507–514.

A. M. TURING [1950], *Computing machinery and intelligence*, ***Mind***, vol. 59, pp. 433–460.

A. M. TURING [1992], *Intelligent machinery*, ***Collected Works of A. M. Turing: Mechanical Intelligence*** (D. C. Ince, editor), Elsevier Science Publishers.

M. T. TURVEY AND C. CARELLO [1995], *Some dynamical themes in perception and action*, (R. F. Port and T. van Gelder, editors), pp. 374–401.

T. TYRRELL [1993], ***Computational Mechanisms for Action Selection***, Ph.D. thesis, University of Edinburgh, Centre for Cognitive Science.

J. VON NEUMANN [1958], ***The Computer and the Brain***, Yale University Press, New Haven, CT.

D. J. Webre, P. M. Wolanin, and J. B. Stock [2004], *Modulated receptor interactions in bacterial transmembrane signalling*, ***Trends in Cell Biology***, vol. 14, no. 9, pp. 478–482.

Please provide the complete addresses of the authors.

DIPARTIMENTO DI SCIENZE FISICHE
UNIVERSITÀ DI NAPOLI FEDERICO II

DIPARTIMENTO DI FILOSOFIA
UNIVERSITÀ DI PISA