Genic Representation: Reconciling Content and Causal Complexity

Michael Wheeler and Andy Clark

Abstract

Some recent cognitive-scientific research suggests that a considerable amount of intelligent action is generated not by the systematic activity of internal representations, but by complex interactions involving neural, bodily, and environmental factors. Following an analysis of this threat to representational explanation, we pursue an analogy between the role of genes in the production of biological form and the role of neural states in the production of behaviour, in order to develop a notion of *genic representation*. In both cases an appeal to *normal ecological context* is used to balance multi-factoral, interactive causal determination against the intuition that certain aspects of the causal nexus play a special role in promoting adaptive success. Certain worries about this vision help us to get a better grip on the concept of genic representation itself. We end with a puzzle concerning the relation between cognition and representation.

1   Introduction: Hard Times For Inner Vehicles

These are hard times for the notion of internal representation. Increasingly, theorists are questioning the explanatory value of the appeal to internal representation in the search for a scientific understanding of the mind and of intelligent action.[1] What is in dispute is not, of course, the status of certain intelligent agents as *representers* of their worlds. It is surely right and proper to say, for example, that a certain person represents as sad the death of Mother Theresa. What is in dispute is not our status as representers, but the presence within us of identifiable and scientifically well-individuated *vehicles* of representational content. Recent work in neuroscience, robotics, philosophy, and developmental psychology suggests, by way of contrast, that a great deal of (what we intuitively regard as) intelligent action may be grounded not in the regimented activity of inner content-bearing vehicles, but in complex interactions involving neural, bodily, and environmental factors and forces.[2]

This is an intriguing claim, and one that has occupied both of the present authors elsewhere.[3] One quick response (which strikes us as correct, as far as it goes) is to draw attention to the special class of cases where such ongoing agent-environment interaction is necessarily absent:

---

[1] See, e.g., Thelen and Smith (1993), van Gelder (1995), Webb (1994), Cliff and Noble (1997).

[2] E.g., in neuroscience see Globus (1992). For related claims in robotics, philosophy and developmental psychology, see note 1 above.

[3] E.g., Clark (1997a, 1997b, 1998), Clark and Toribio (1994), Wheeler (1994, forthcoming).

cases such as counterfactual rehearsal and so on.[4] Such cases are important and worthy of special study, but they do not constitute the bulk of daily, *on-line* intelligent behaviour. The primary expression of natural, biological intelligence, it seems fair to say, consists in the production of an ongoing sequence of actions that flexibly respond to a constant stream of incoming environmental stimuli — think of driving a car, preparing a meal, or engaging in a lively conversation. In cases such as these, the allegedly problematic profile of complex, multi-factor interactivity clearly obtains. In this paper we investigate the claim that complex causal interactions cause trouble for the notion of inner representational vehicles. We review some of the cases supposed to put pressure on a representational-vehicle based understanding and conclude that the threat, even in these ongoing, interactive cases, is more apparent than real. The main contribution of the present paper, however, is to go beyond this negative thesis[5] and to begin to develop a somewhat different vision of the nature of internal representation itself. For reflection on the more densely interactive and distributed (across brain, body and world) class of cases suggests, we believe, a more sensitive and broadly applicable notion of the way inner states contribute to intelligent behaviour, and hence of the notion of an inner representational vehicle itself. A crucial part of our strategy will be to pursue an extended analogy between the role of genes in the causal determination of biological form, and the role of neural states in the causal determination of intelligent action.[6]

2   Spreading the Cause

What does it mean to say that intelligent behaviour comes about through complex interactions between neural, bodily, and environmental factors? Here is a compelling example, from research taking place at the interface between biology and robotics. To attract mates, male crickets produce an auditory advertisement, the frequency and temporal pattern of which are species-specific. Conspecific females track this signal. So what sort of mechanisms in the female might underlie her tracking capacity? By combining existing neuroethological studies with robotics-inspired ideas, and by implementing a candidate mechanism in an actual robot, Webb (1994, 1996) has developed and defended an elegant story that begins with the anatomy of the female cricket's peripheral auditory mechanism.

The female cricket has two ears (one on each of her two front legs). A tracheal tube of fixed length connects these ears to each other, and to other openings (called 'spiracles') on her body. So sound arrives at the ear-drums both externally (directly from the sound source) and internally (via the trachea). According to Webb, what happens when the male's auditory signal is picked-up is as follows: On the side nearer the sound source, the externally arriving sound travels less distance than the internally arriving sound, whereas, on the side further away from the sound source, the two sounds travel the same distance. This means that the external and internal arrivals of sound are out of phase at the ear-drum on the side nearer to the sound source, but in phase at the ear-drum on the side further away from the sound source. The practical consequence of this difference is that the amplitude of ear-drum vibration is higher on the side nearer the sound source. In short, when the male's signal is picked up, there is a direction-dependent intensity difference at the female's ears, with the side closer to the sound source having the stronger response.

---

[4]For a full treatment of this class of cases, see Clark and Grush (forthcoming).

[5]The negative point is also pursued by Clark (1998).

[6]A similar analogy is put to work by Keijzer (1997). However, as we shall see, Keijzer's inter-related accounts of how genes contribute to the generation of phenotypic form, and of how neural states contribute to the generation of behaviour, are substantively different from ours. Indeed, they embody precisely that scepticism about representation which we aim to resist.

So far so good. All the female has to do, in order to find the male, is to keep moving in the direction indicated by the ear-drum with the higher amplitude. This is achieved via the activity of two dedicated inter-neurons in the female's nervous system, one of which is connected to the female's right ear, the other of which is connected to her left. Each of these neurons fires when its activation level reaches a certain threshold, but because the neuron which is connected to the ear-drum with the higher vibration amplitude will be receiving the stronger input, that neuron will reach threshold more quickly, and hence fire first. So the idea is that, at each burst of sound, the female moves towards the side of the neuron that fires first. In effect, therefore, her tracking mechanism responds only to the *beginning* of a sound burst. It is here that the specific *temporal pattern* of the male's signal becomes important. On the one hand, the gaps between the syllables of the male's song must not be too long, otherwise the information about his location will arrive too infrequently for the female's mechanisms to track him reliably. On the other hand, the gaps between the syllables of the male's song must not be too short. Each of the two dedicated interneurons in the female's nervous system has a decay time during which, after firing, it gradually returns to its non-activated rest-state. During this recovery period, a neuron is nearer to its firing threshold than if it were at rest. In consequence, if a neuron receives input during the decay time, it will fire again more quickly than if it receives that input while at rest. So if the gaps between the syllables of the male's song were shorter than the total decay time of the neurons, or if the song were continuous, then it would often be unclear which of the two neurons had fired first. Thus the temporal pattern of the male's song is constrained by the activation profile of the dedicated interneurons in the female.

The humble cricket — or its robotic relation — serves as a first demonstration of the way in which cognitive science encounters an effect that we shall call *causal spread*. Causal spread obtains when some phenomenon of interest turns out to depend, in unexpected ways, upon causal factors external to the system previously/intuitively thought responsible.[7] We confront the effect here because a standard cognitive-scientific explanation would surely have understood the tracking behaviour as depending (at least in any interesting sense of 'depend') on the systematic activity of neurally-located representational and computational processes (e.g., an inner map of the environment which might be used both to encode the estimated position of the target, and to calculate a satisfactory route).[8] In Webb's explanation, things are different. Adaptive success is secured by a system of organized interactions involving significant causal contributions not only from the female cricket's nervous system (the dedicated interneurons), but also from her body (the fixed-length trachea), and from her environment (the specific structure of the male's signal). Thus the outcome is generated by a network of co-contributing causal factors that extends across the systemic boundaries between brain, body, and environment. This is the distinctive signature of causal spread in the mechanisms underlying intelligent action.

A second taste of causal spread comes from the developmental literature on infant walking. Hold a newborn human infant upright and she will produce coordinated stepping movements. However, at about 2 months of age, these movements seemingly disappear from the child's repertoire of actions, only to reappear again at about 8 to 10 months, when she acquires the capacity to support her own weight on her feet. At about 12 months of age, the infant takes her first independent steps. How might this reliable sequence of transitions be explained? A certain tradition in developmental theorizing would understand it as, broadly speaking, the relentless playing out of a set of developmental instructions encoded in the infant's genes. On this view, behavioural change follows from the gradual maturing and coming 'on-line' of a

---

[7]For a related use of the term, see Hurley (forthcoming). Many thanks to Susan Hurley, Christopher Hookway, and an anonymous referee for specific comments which have helped us to improve our characterization of causal spread. For an earlier characterization, see Clark and Wheeler (forthcoming).

[8]What we mean by the qualification 'interesting sense of depend' should become clear in the next section.

genetically-specified central pattern generator which is housed in the central nervous system, and which determines patterns of walking by issuing directions to the appropriate muscles. This 'instructionist' style of explanation is roundly rejected by Thelen and Smith (1993), who suggest that locomotion and locomotive development are examples not of central control (neural or genetic), but of *self-organization* in an extended brain-body-environment system, that is, in a system exhibiting large-scale causal spread.

The concept of self-organization will play an important role in what follows, so we need to cash it out. Self-organization occurs in energetically open complex systems where large numbers of components interact with each other in nonlinear ways to produce the autonomous emergence and maintenance of structured order.[9] Popular illustrative examples include the Belousov-Zhabotinski reaction, lasers and slime-molds (see, e.g., Goodwin, 1994; Kelso, 1995; Thelen & Smith, 1993). To appreciate just one way in which, on Thelen and Smith's account, self-organization makes a crucial contribution to an understanding of infant walking, recall the developmental stage in which infants who are being held upright *do not* produce coordinated stepping movements. Thelen and Smith found that, during this stage, 7 month old infants *will* immediately produce coordinated stepping movements, *if* they are held upright *on a slow-moving motorized treadmill*. Moreover, the infants are able to compensate for increases in the overall treadmill speed; and they are able to make asymmetrical leg adjustments to preserve coordinated stepping, if the two belts of the treadmill are set to move at different speeds. Thelen and Smith's hypothesis (op cit. pp. 96-7) is that the mechanical action of the treadmill brings about a shift between the non-stepping and stepping modes of behaviour. The treadmill does this simply by replacing certain leg-dynamics that occur naturally in mature locomotion. Just after the infant's legs 'touch down' onto the moving treadmill, her centre of gravity is shifted over the stance leg and the trailing leg is stretched backward. In effect, the trailing leg behaves like a spring, the stretch imparting energy into the leg which then swings forward with little or no extra muscle activity. The initiation of the swing appears to be triggered by the proprioceptive, biomechanical information available to the central nervous system at the point of maximum stretch, while the achievement of maximum stretch is itself jointly determined by the action of the treadmill and the infant's capacity to make flat-foot contact with the belt, this capacity depending in turn upon the degree of flex in the infant's leg muscles. So, given the context provided by the treadmill, the basic patterns of mature walking dynamically self-organize in real-time, during a phase of development in which such patterns were thought to be lost to the infant.

From this sort of example, Thelen and Smith draw the following conclusions.

> Leg coordination patterns are entirely situation-dependent ... There is ... *no essence* of locomotion, either in the motor cortex or in the spinal cord. (pp.16-17, original emphasis)

> There is no logical or empirical way to assign priority in the assembly of ... [kicking] movements to *either* the organic components that "belong" to the infant *or* to the environment in which they are performed. Both determine the collective features of the movement which we call a kick. (p.83, original emphasis)

> [Locomotor] development can only be understood by recognizing the multidimensional nature of this behavior ... the organic components and the context are equally causal and privileged. (p.17)

---

[9]The term 'autonomous' is commonly used in connection with self-organization, to capture the fact that the order observed in a self-organizing system is a product of the intrinsic nature of that system, rather than of the activity of some external organizing agent.

These conclusions make fully explicit the claim that the causal contributions made to on-line adaptive success by environmental and (non-neural) bodily happenings are on a par with those made by neural happenings. But there is also a more radical claim appearing here (one which, in truth, was bubbling just under the surface of the cricket example), namely that, in cases of on-line intelligence, no aspect of the extended brain-body-environment causal system should be *explanatorily* privileged in the cognitive-scientific understanding of the observed behaviour. This sharing-out of the explanatory weight — call it *explanatory spread* — is manifestly at odds with the cognitive-scientific tradition, which is to seek explanations of intelligent behaviour that appeal fundamentally, and usually exclusively, to strictly agent-internal neural phenomena.

Once the distinction between causal spread and explanatory spread has been made, a question arises about the relationship between the two. Reading between the lines, it appears that the message from Thelen and Smith is that causal spread and explanatory spread come as a package: thus we learn that "the organic components and the context are equally causal and privileged". If this co-occurrence is supposed to indicate a necessary link between the two forms of spread, then perhaps the most obvious and plausible way in which that link might be made intelligible is if explanatory spread were *entailed by* causal spread. Under such circumstances, any non-neural features of the agent's body (e.g., tracheal tubes, muscle dynamics), or any features of the agent's environment (e.g., distinctively structured auditory signals, treadmills), that figured as full-blooded factors in the underlying causal story, would thereby assume an explanatory status equal to that of any neural features that so figured. Whether this is the right way to think about things is an issue to which we shall return.

3   Explanation Without Representation?

In addition to laying stress on the causal and explanatory contributions made by the agent's non-neural body and her environment, theorists who embrace causal spread often display a marked hostility to the representation-centred explanatory vocabulary that is at the heart of most cognitive science. For example, in discussing her robot implementation of the theory of cricket phonotaxis described earlier, Webb declares her mistrust of representational interpretations.

> The robot operates without any attempt to build an internal model of its environment: there is no centralised representation of the sensory situation, not even in a distributed sense ... It could be argued that the robot does contain 'representations', in the sense of variables that correspond to the strength of sensory inputs or signals for motor outputs. But does conceiving of these variables as 'representations of the external world' and thus the mechanism as 'manipulation of symbols' actually provide an explanatory function? It is not necessary to use this symbolic interpretation to explain how the system functions: the variables serve a mechanical function in connecting sensors to motors, a role epistemologically comparable to the function of the gears connecting the motors to the wheels. (Webb, 1994, p.53)

Thelen and Smith are even more emphatic:

> How do minds change? Where does new knowledge, new understanding, new behavior come from? How does the organism continually adapt and create new solutions to new problems? The answer we present ... makes no use of representation or representation-like processes. (Thelen & Smith, 1993, p.42)

For sure, these theorists (and others who pursue a similar line) inherit some of their scepticism about representational styles of explanation from aspects of their overall theoretical outlook other

than their embrace of causal spread (aspects that can sometimes be traced to the influence of Continental phenomenology or Gibsonian psychology). But there is little doubt that causal spread is a key factor, and that it is often taken to be implied by, or supportive of, those other aspects (see Wheeler, 1996). So it is certainly appropriate to ask to what extent a distrust of representationalism might reasonably be fuelled directly by a recognition of causal spread. In other words, does the mere presence of causal spread provide sufficient grounds for rejecting the claim that contributing neural states should be viewed as inner representations?

A raw affirmative cannot be the right answer here. To see why, consider a simple case.[10] A standard LISP program for sorting a set of references into alpha order might comprise, let's say, 30 lines of code. These lines of code will include items such as (cons d (abc)) where CONS names a function that takes an item (in this case 'd') and appends it to a list (in this case 'abc'). We think it is fair to assume that the 30 lines of code, in standard usage, is thought to constitute a representational specification of how to proceed so as to solve the problem. In fact, the 30 lines surely constitute a bona fide *algorithm* for the task. *But the algorithm or specification is effective only in the context of a machine set-up to compile LISP.* The operation of the function CONS is not defined as part of what is intuitively the program. It is part of the backdrop against which the program is expected to do its stuff.[11] It seems, then, that even the most routine and paradigmatic examples of programs, recipes and algorithms already display a good amount of causal spread, a fact which indicates that causal spread cannot directly undermine the claim that certain factors constitute a kind of coded set of instructions (i.e., a body of problem-solving representations) for solving a problem.

For the anti-representational case to get off the ground, then, it is not enough merely to demonstrate that extra-neural factors and forces are necessary for certain forms of behavioural success. Indeed, such an observation verges on the trivial, and any dependence of behaviour on those extra-neural factors would be of an essentially uninteresting kind. The fact is, however, that there are different types of causal spread, and not all of them are as mundane and as obviously harmless to standard modes of representational explanation as the variety highlighted by the list-sorting example. What the best of the recent empirical work (e.g., Webb's work on cricket phonotaxis) clearly demonstrates is that extra-neural factors and forces can be responsible for aspects of behaviour that might otherwise be expected to be under the direct control of some more heavily representational style of problem solution. One such non-trivial form of the phenomenon occurs when the additional factors and forces reveal themselves as the unexpected root of a good deal of the *kind* of flexibility and subtlety normally associated with representation-based control. Thus consider the idea — due to Ballard, Hayhoe, Pook, and Rao (1997) — that biological organisms use bodily orientations (such as saccadic eye motions) to directly yield the kinds of binding (of object features to spatial location) classically achieved by the use of computational pointers (data structures that link sites in memory). As the authors comment:

> [The] external world is [here] analogous to computer memory. When fixating a location, the neurons that are linked to the fovea refer to information computed from that location. Changing gaze is analogous to changing the memory reference in a silicon computer. (Ballard et al., 1997)

In this case our conditions are surely met. Aspects of body and world are here doing the kind of work normally associated with a purely inner computational resource. And the work done is responsible, in part, for the fluency and flexibility of the observed behaviors of the system.[12]

---

[10]This example appears, in a different form, in Clark (1998).

[11]This idea of a 'backdrop' will be important later.

[12]In citing the binding example at this point in the text, we do not mean to suggest that Ballard et al. are themselves hostile to all forms of representational explanation. Rather, their model provides an elegant example

Given that it is clearly possible for body and world to steal the adaptive thunder from inner representations in this way, the following argument has some prima facie force: in instances where the various additional causal contributions made by non-neural bodily and/or environmental factors account for much of the richness and the adaptive flexibility of the observed behaviour, the contributing neural states or processes cannot count as fully specifying that behaviour (i.e., as encoding the instructions for what to do). An especially potent application of this style of reasoning becomes possible where the form of the behaviour stems, to some large extent, not from instruction-sets issued by a centrally located controller, but from processes of generic self-organization that extend across brain, body, and environment (as in Thelen and Smith's account of infant walking). The general position is nicely stated by Keijzer:

> The neural system ... [does not] ... incorporate a complete set of behavioral instructions ... Instead, the neural system uses the order which is already present in the musculo-skeletal system and the environment. Behavior is the result of the interactions between the pre-existing order in these systems, and consists of the mutual modulation of neural, bodily and environmental dynamics. (Keijzer, 1997, p.204)

What all this shows, we think, is that there is some plausibility to the claim that certain forms of large-scale causal spread threaten to undermine representational forms of explanation in cognitive science.[13]

In what remains of this paper, we hope to stem the tide of anti-representationalism, without losing sight of the important lessons that have emerged from the foregoing discussion. Such a project recommends itself, in part because, despite the foregoing identification of a prima facie threat to representational explanation, there is evidence that representation-based modes of understanding remain useful tools, even where we confront ongoing, interactive behaviours involving large degrees of non-trivial causal spread. Here we shall consider the details of just one illustrative example (for others, see, e.g., Agre, 1988; Ballard et al., 1997; Mataric, 1991).

In work by Franceschini, Pichon, and Blanes (1992), detailed neurobiological knowledge of the visual system of the housefly was used to guide the design of a mobile robot that is capable

---

of the sort of adaptive solution that might, in certain hands, be used to construct an anti-representational case.

[13]Having arrived at this claim concerning causal spread, we should pause to note the existence of a different but related phenomenon, namely *continuous reciprocal causation*. As defined by Clark (1997a, 1998), this is causation that involves multiple simultaneous interactions and complex dynamic feedback loops, such that (a) the causal contribution of each component in the system of interest is determined by, and helps to determine, the causal contributions of large numbers of other components, and, moreover, (b) those contributions may change radically over time. In certain extreme cases of continuous reciprocal causation, the sheer number and complexity of the causal interactions occurring within the agent's nervous system and/or between the agent and its environment can, it seems, make it difficult to apply standard representational explanatory strategies. (For further discussion of this point, see the aforementioned papers by Clark, plus the treatment by Wheeler (forthcoming).) So although causal spread and continuous reciprocal causation are clearly separable phenomena — as indicated by the fact that one could have a system in which causal spread is rife, but in which the multiple causal factors concerned play distinct roles and combine linearly to yield action— they can apparently have similar implications. Indeed, both can be used independently to provide an argument for the adoption of a thoroughgoing *dynamical systems approach to cognitive science*. Dynamical systems theory is, in many ways, a natural mathematical language in which to formulate explanations of (i) causal systems that range across brain, body, and environment (see, e.g., Beer, 1995; Thelen & Smith, 1993; van Gelder & Port, 1995), and (ii) the subtle interactions that are characteristic of continuous reciprocal causation (see, e.g., the discussion of the Watt Governor by van Gelder, 1995). Although, in this paper, we do not explicitly pursue issues surrounding the dynamical systems approach, it will not escape notice that we make use of certain dynamical systems concepts in developing our position. (The canonical introduction to the dynamical systems approach is the collection edited by Port and van Gelder (1995). The view has been developed and defended by, among others, Beer (1995), Husbands, Harvey, and Cliff (1995), Kelso (1995), Smithers (1995), Thelen and Smith (1993), van Gelder (1995, forthcoming), and Wheeler (1994). For critical discussion of the field, see Clark (1997a, 1997b) and Clark and Toribio (1994).)

of navigating its way to a goal (a light source), in real time, whilst avoiding obstacles. This feat is achieved via a series of simple movements, each made according to a specific direction-heading generated in the following way. The robot is endowed with a compound eye featuring a layer of photosensors connected to a parallel array of analogue elementary motion detectors (EMDs). This EMD-array drives an obstacle avoidance layer (a parallel-processing, analogue network) that integrates the motion signals from the EMDs to help control the steering of the robot. Since the only information available to the obstacle avoidance layer concerns motion, as registered by the EMD-array, in an environment of stationary obstacles, the robot is, for the purposes of obstacle avoidance, blind at rest. However, the EMD-array is able to use relative motion information, generated by the robot's own bodily movements during the preceding step, to build a temporary 'snap map' of detected obstacles. This map, which is made available to the obstacle-avoidance layer, is constructed in a polar coordinate system defined in relation to the compound eye. (Since the eye is mounted in a fixed position, it rotates with the wheels, so a frame of reference centred on the eye is also one centred on the direction of movement.) Then, in a local and short-lived motor-map, information concerning the angular bearings of the obstacles (as supplied by the snap map) is fused with information concerning the angular bearing of the goal (as supplied by a supplementary visual system), and the direction-heading of the next movement is generated. This heading is as close as possible to the one that would take the robot straight towards the goal, adjusted so that the robot avoids all detected obstacles.

It seems that the adaptive success of Franceschini et al.'s robot is a paradigmatic case of on-line intelligent behaviour being generated by a distributed causal process featuring non-trivial causal spread. The routes that the robot traverses are never fully specified or pre-determined at any global level by centrally located 'neural' processes. Rather, they emerge from ongoing dynamic feedback between brain, body and environment. So factors located in each of the relevant causal domains — brain, body, and environment — contribute to the distinctive form and adaptive flexibility of the goal-seeking/obstacle-avoidance behaviour, and, moreover, make their contributions during the real-time execution of that behaviour.[14]

We now face a certain tension: on the strength of the case presented earlier, one might expect that any attempt to make an explanatory appeal to internal representations in the case of the Franceschini et al. robot will court problems; yet the explanation of the agent's behaviour as governed by a series of *maps* seems to speak *in favour* of a recognizably representational understanding of that behaviour. At this point care is needed, not least because the internal structures in question are clearly not maps in a certain familiar sense commonly appealed to within artificial intelligence, in that they are not symbolic representations of the objective environment which are stored and recalled for the off-line advance planning of a route. Instead, they are short-lived, task-specific, egocentric (i.e., robot-centred) structures, which are built on-line, during each round of sensing and action. That said, the fact remains that nothing about this particular profile of features seems to undermine, in any obvious way, their status *as maps* —

---

[14]Imagine, by way of maximal contrast, a robot which uses sensory inputs to build an internal model of the environment, and which then uses that model to plan an appropriate path to the target, encoded as a set of movement-instructions. Assuming everything goes 'to plan' (the model is accurate, the environment doesn't change too much etc.), the motors then follow the pre-specified movement-instructions, without any additional or ongoing sensory feedback from the environment. (This is a rather extreme example of a robot that works according to the principles of what Brooks (1991) has dubbed the 'sense-model-plan-act' framework. Brooks takes this framework to be characteristic of the approach to robotics adopted in classical artificial intelligence.) This hypothetical robot does not rely on regular sensory feedback to complete its task, but rather on the information extracted from the initial sensory input. Moreover, although additional factors are clearly essential, in some sense, for the behaviour to occur as it does (e.g., the robot needs motors and wheels in order to move), the source of any adaptive flexibility or context-appropriateness in the behaviour is here rooted firmly in the inner route-planning mechanism.

that is, as some sort of representation[15] — which does not, of course, rule out the possibility that some other feature or features of the scenario might threaten to have such an effect.

In this tension lies the hope of a rapprochement between causal spread and representational explanation. Time, then, to reconsider our options, and to try to determine the true implications of causal spread. One way to do this, we think, is to turn to a somewhat different arena — one in which we see essentially the same debate, but in which the conceptual and empirical terrain has been better explored.

4 Genes, Codes and Programs

The received orthodoxy in contemporary biology is that genes, or complexes of genes, should be understood as *encoding for* phenotypic traits (such as bodily forms and neural architectures). Thus an organism's genotype is standardly characterized as a *set of instructions for*, or as *a program for*, or as *storing the information for*, the building of that organism's phenotype. In short, the standard view in biology is that *genes represent phenotypic traits*, and that *genotypes represent phenotypes.* We have seen that representational explanatory strategies in cognitive science might be put under pressure by the presence of causal spread in the systems that underlie on-line intelligent behaviour. One might wonder, therefore, if there is any way in which causal spread might threaten the representational understanding of genes? In order to see how this might be a real issue, we need to get clear about how, in the context of this discussion, to map the phenomena studied by cognitive scientists onto the phenomena studied by developmental biologists. In relation to genes and their phenotypic effects, it seems that the correlates of neural states are genes and gene-complexes, and that the correlates of behavioural outcomes are phenotypic traits. Therefore the system in which the presence of causal spread might threaten the idea that genes encode for traits would be the system that underlies the ontogenetic development of phenotype from genotype. First, then, we need to identify causal spread in the developmental system.

Our first stop is a compelling explanation of biological development due to Brian Goodwin (1994), in which phenotypic forms are conceptualized as the characteristic and robust modes of order that emerge through self-organization in (what Goodwin calls) morphogenetic fields. Morphogentic fields (like the physicist's fields in general) are dynamical systems extended in space. But they are fields which involve combinations of physical and chemical properties in ways which are distinctive of living systems. Thus, in his account of the developmental system that generates the phenotypic form of *Acetabularia* (a type of algae), Goodwin gives key roles to the ways in which cells control the concentration of calcium in the cytoplasm (the plasmic content of the cell other than the nucleus), and the ways in which calcium affects the mechanical properties of the cytoplasm.

Goodwin drives home his message via discussions of many phenotypic traits, including tetrapod limbs and the vertebrate eye. Here we shall concentrate on his account of phyllotaxis (leaf arrangements) in higher plants (op cit. pp.105-119). Despite the enormous variety of phyllotactic arrangements that we see in nature, there are really only three generic forms. The most common of these is spiral phyllotaxis, in which successive leaves on the stem appear at a fixed angle of rotation relative to each other. Amazingly, the angles of rotation found in natural instances of spiral phyllotaxis tend to take one of only a few values, of which the most common is 137.5 degrees. How might one explain these facts? Goodwin offers the following account: As leaf-tissue grows, it places pressure on an elastic surface layer of epidermal cells. This pressure causes the epidermal cells to synthesize cellulose microfibrils to resist the force. Where the next

---

[15]Cf. the discussion of representation, egocentricity and task-specificity by Clark (1997b).

leaf will grow is determined by the fact that, as a result of exactly where the stress has been placed, and exactly how the cellulose defences are laid down, the resistance to growth will be stronger in some areas of the epidermal layer than in others. The idea, then, is that the pattern of leaves in the phenotype results from a sequence of mechanical interactions between (i) the growing leaves under the epidermal surface, and (ii) the barricades of defending cellulose microfibrils. To provide support for such a view, Goodwin cites modelling studies (due to Green) which show that the phyllotactic arrangements observed in nature are stable patterns produced by such a system. This done, the challenge is to explain why these arrangements are the *only* stable arrangements generated by the proposed system. Here Goodwin appeals to a second model (due to Douady and Couder) which demonstrates that if (a) the rate of leaf formation is above a critical value, and (b) the system starts with the most-commonly-found initial pattern of leaf primordia in the growing tip, then the developing plant will tend overwhelmingly to settle on spiral phyllotaxis with an angle of rotation of 137.5 degrees! In other words, given certain parameter-values and initial conditions, the most common arrangement found in nature is the dominant generic form produced by the self-organizing developmental dynamics in the model. Moreover, with different values for certain key parameters in the model (e.g., the growth rates and the number of leaves generated at any one time), the other phyllotactic arrangements observed in nature reveal themselves as less-likely generic forms of the system. The evidence is indeed persuasive: it seems that phyllotactic development is plausibly understood as a process of self-organization in a morphogenetic field in the growing tip of a plant.

In the present context, the most important point to take away from all this is that the sorts of chemical and mechanical processes singled out by Goodwin, as essential causal factors in development, provide further evidence of causal spread in the developmental system. The effects of calcium on the mechanical properties of the cytoplasm, mechanical stresses in elastic sheets of cells, spatial displacements of growing forms in the face of physical resistance: all of these processes are extra-genetic in nature. Once again the crucial causal nexus is creeping ever farther outwards, this time from the genotype into the surrounding cellular and extra-cellular bodily context. The final extent of this leakage should come as no surprise.

Consider the Mississippi alligator. These creatures lay their eggs in a nest of rotting vegetation which produces heat in varying quantities. Eggs that develop at lower temperatures (within some overall range) end up producing females, whilst those that develop at higher temperatures end up producing males. Eggs in a clutch will pass through the critical developmental window at various different temperatures, meaning that a mixture of females and males will be born. This environmental method of regulating sex ratio (a ratio which, for reasons of population-survival, needs to stay somewhere near 50:50 in the population) might seem a little hit and miss, but it works well enough (for more details, see Goodwin, 1994, p.38). The lesson is clear: environmental features are active causal factors in biological development, and thus the developmental systems that causally determine phenotypic traits, just like the behaviour-generating systems that produce on-line intelligence, can exhibit the most extreme, environment-involving variety of causal spread.[16]

So how does the representational theory of genes fare, once causal spread in developmental systems is acknowledged? Here are some views:

> We have often heard it said that genes contain the "information" that specifies a living being. This is wrong for two reasons. First, because it confuses the phenomenon of heredity with the mechanism of replication of certain cell components

---

[16]For additional evidence, see Lewontin's account of how the number of light-receptor cells in the compound eye of the *Drosophila* fly depends, in part, on the environmental temperature during development (Lewontin, 1983, pp.69-70).

(DNA), whose structure has great transgenerational stability. And second, because *when we say that DNA contains what is necessary to specify a living being, we divest these components ... of their interrelation with the rest of the network.* (Maturana & Varela, 1987, p.69, emphasis added)

[Selection] arises from cooperative-competitive interactions between specific (genetic) parametric constraints and intrinsic (generic) dynamics ... in such a view the gene in no way constitutes a program for development. Rather, it's a participant playing a parametric role in a self-organizing, synergetic process. (Kelso, 1995, p.183)

[Genes] do not specify macroscopic order. This order is inherent in the self-organizing system. (Keijzer, 1997, p.224)

For some theorists, then, the fact that developmental systems exhibit causal spread seems to undermine completely the representational understanding of genes. But is there really a mandate for this particular elimination of representational talk? We favour a less iconoclastic response to developmental causal spread, one which starts by recognizing a crucial distinction between two possible types of representational claim that might be made in the vicinity of genes.

It seems rather natural, in many ways, that a representational theory of genes should inspire the following thought: if one could find out the complete sequence of an organism's DNA, then, in principle, one would be able to use that information *alone* to 'compute' the adult organism, such that one would be able to predict, in every relevant detail, that adult's phenotypic form. Notice that this idea makes sense only if the ontogenetic development of the individual organism can reasonably be conceptualized as the unfolding of a comprehensive, predetermined plan that is stored in that organism's genes. DeLisi is just one of many, many theorists who openly embrace such a view:

The collection of chromosones in the fertilized egg constitutes the complete set of instructions for development, determining the timing and details of the formation of the heart, the central nervous system, the immune system, and every other organ and tissue required for life. (DeLisi, 1988, p.488)

This approach to development — call it *strong instructionism* — is anathema to theorists who embrace causal spread (e.g., Thelen and Smith, Goodwin). And, indeed, it seems that the existence of extensive causal spread in the developmental system does make strong instructionism difficult to defend. If phenotypic forms are, in many ways, the products of generic self-organizing processes in the cellular and extra-cellular body of the developing organism, in interaction with the environment, then knowing the entire sequence of that organism's DNA will simply be insufficient to predict phenotypic form. As the examples discussed earlier show, one will additionally need to know the relevant principles of biological (self-)organization, plus the values of the key environmental variables.

Fortunately, strong instructionism is not the only position available to the representationalist about genes. Indeed, it would be a grave mistake to confuse the general and relatively weak claim that genotypes are encodings with the more specific and much stronger claim that the genotype is a complete set of instructions for building the adult organism. Concrete evidence which is devastating for the latter claim might not be nearly so serious for the former, and one wouldn't want to throw out the weakly representational baby with the strongly instructionist bath-water. Of course, for the distinction between strongly instructionist and weakly representational theories to hold here, one needs to reach an understanding of what it means for genes

to be weakly representational, one that can cope with non-trivial causal spread in the developmental system. In what remains of this section we shall take the first steps towards such an understanding.

One might wonder about what it is exactly that genes do, within the dynamic context provided by Goodwinian morphogenetic fields, that makes a difference to the phenotypic structure of an organism? The answer, according to Goodwin, is that genes select or stabilize the developmental dynamics that, through self-organization, actualize one of the alternative generic forms that are possible for the organism in question. As Goodwin himself puts it, "During reproduction, each species produces gametes with genes defining parameters that specify what morphogenetic trajectory the zygote will follow" (Goodwin, 1994, p.102). In other words, genes determine parameters for the self-organizing systems that generate phenotypic forms.

The view of genes proposed by Goodwin seems (at least to us) to invite a kind of weak representational reading. Indeed, it seems not only permissible, but illuminating, to conceptualize the situation as one in which genes *encode* developmental parameters. For us, this is undoubtedly a step in the right direction, since the timely demise of the strongly instructionist developmental program does not undermine this weaker thesis. However, we have not as yet reached a fully satisfactory place to come to rest. First, notice that certain extra-genetic factors (e.g., non-genetic cellular machinery, environmental features) might, like genes, be treated as parameterizing developmental dynamical systems. These other factors would then co-determine, along with the relevant genes, exactly which possible trajectory of that system would finally be traversed by the developing organism. This suggests that if we want to preserve the intuition that genes are the only entities in the developmental system which enjoy representational status, then representation cannot be *just* a matter of parameter-setting; it has to be parameter-setting *plus something else*. More generally, what the appeal to parameter-encoding does is help us to understand the role that certain physical entities, as representations, might play in a system featuring large-scale causal spread. The task that remains is to pin down precisely what it is for these putatively representational entities to *encode* parameters, rather than just set them. The question 'What makes a physical state or process an encoding?' will be taken up in the next section.

Here is a second source of possible dissatisfaction with the present position. Ideally, one would want to say not merely that genes encode developmental parameters, but that, by encoding developmental parameters, genes encode for phenotypic traits. How might this extension be secured? The key move is suggested by the following remarks from Varela, Thompson, and Rosch:

> For many years biologists considered protein sequences as being instructions coded in the DNA. It is clear, however, that DNA triplets are capable of predictably specifying an aminoacid in a protein if and only if they are embedded in the cell's metabolism, that is in the thousands of enzymatic regulations in a complex chemical network. It is only because of the emergent regularities of such a network as a whole that we can bracket out this metabolic background and thus treat triplets as codes for aminoacids. (Varela et al., 1991, p.101)

The message here is that to make even the more limited claim that DNA encodes for amino acids during protein synthesis (a process which we discuss later in this paper), one has to assume certain regularities in the dynamic context in which gene-activities take place, that is, one has to treat the cell's metabolism as a stable ecological backdrop against which DNA operates. To make the stronger claim that genes encode for specific phenotypic traits, one has to extend this assumption of a stable ecological context to the self-organizing processes at work in the appropriate morphogenetic field, and to certain crucial factors in the organism's environment.

Thus any claim that a gene, or set of genes, encodes (weakly) for a specific phenotypic trait depends on the assumptions (a) that the principles of morphogenetic self-organization which are generic to the organism in question will operate as normal, and (b) that the expected values of the relevant variables in the developmental environment will obtain. This suggests that it might be legitimate to treat certain genes as coding for a particular phenotypic trait, just so long as the presence of those genes can be seen to yield — in a rich, normal ecological setting — that very phenotypic outcome.[17]

At a minimum, it now seems that it is possible to embrace all the important insights concerning dynamical complexity, self-organization, and environmental interaction that flow from a recognition of non-trivial causal spread in developmental systems, and yet to maintain a robust, empirically useful version of the basic idea that genes encode for phenotypic forms. This suggests that one might be able to turn the same trick for internal representations, as they figure in cognitive-scientific explanations of on-line intelligent behaviour, by working out the details of the idea that the way in which neural states do their representing is relevantly similar to the way in which genes code for phenotypic forms. From what we have seen so far (and we should stress that our story still has a few twists in store), the analogy would go like this: Genes do not fully specify the final phenotypic form. They operate within a normal developmental context of other causally active (biochemical and environmental) states and processes. However, that does not prevent genes from being representational in nature. By encoding certain parameters, genes stabilize the complex developmental dynamics that, in interaction with the environment, yield specific phenotypic forms. A genic account of internal representations would suggest that neural states which are implicated in the production of on-line intelligent behaviour do not fully specify external states of affairs or behaviours. Rather, they operate within an assumed, normal ecological context of other causally active (bodily and environmental) states and processes. However, that does not preclude our treating the neural states in question as representations. By encoding certain parameters, those states stabilize the complex neural and bodily dynamics that, in interaction with the environment, yield appropriate behavioural outcomes.[18]

The genic strategy enables us to defend the use of representational explanations, even when we are faced with the sort of non-trivial causal spread in which the extra factors and forces are responsible for much of the adaptive richness that might previously have been attributed to inner codes and programs. However, the tactic appears to incur a cost. For it seems to require an important shift at the very heart of cognitive-scientific explanation. The vast majority of cognitive scientists have taken their goal to be simply to identify the representational structures and computational (representation-manipulating) processes whose realization and activity in the nervous system explains intelligent behaviour, thought, and cognition. Thus cognitive science has, in general, concerned itself exclusively with strictly agent-internal, and ultimately neural phenomena. But the genic account seems to demand that cognitive-scientific investigations of the causal processes underlying on-line intelligent behaviour should factor-in more than merely neural states and processes. Moreover, notice that although, on the account of biological devel-

---

[17] The complex mixture of influences at work here is nicely captured in Luis Rocha's recent notion of 'emergent morphology' in which there "is not a linear mapping of coded messages to functional products, rather, messages encode dynamic structures that are left to their own devices as they self-organize" (Rocha, 1996, p.380).

[18] The idea that neural states contribute to the generation of behaviour by setting parameters for self-organizing dynamical systems is also at work in Kelso's notion of 'specific parametric influences' (Kelso, 1995) and in Keijzer's concept of 'internal control parameters' (Keijzer, 1997). However, both Kelso and Keijzer are essentially hostile to any representational understanding of the neural states concerned. In our view, the anti-representational leanings of these two theorists can be traced to a shared tendency to restrict the notion of a program to one which is strong-instructionist in nature. This move, which has the effect of squeezing to nothing the conceptual space in which other representational strategies might be developed, is particularly visible in the accounts which Kelso and Keijzer offer of how genes contribute to the generation of phenotypic traits (see the quotations reproduced earlier in this section; for the more detailed stories, see Kelso, 1995, p.183, and Keijzer, 1997, pp.218-26).

opment we favour, genes retain their status both as genuine encodings and as important causal factors in the developmental mechanisms underlying the building of phenotypes, they emerge as in no way causally privileged aspects of those mechanisms. If, in addition, it turned out that explanatory spread is indeed entailed by causal spread (a possibility which we raised earlier), then, from the fact of causal spread, one could immediately draw the further and more radical conclusion that whilst genes can rightly be said to encode for phenotypic traits, they should not, in any way, be given explanatory privilege in a proper understanding of biological development. So if neural states are, like genes, just one factor in a complex causal web whose combined action yields the target outcome, and if causal spread implies explanatory spread, then why should neural states enjoy any privileged explanatory role in cognitive-scientific thinking? Time, then, to reconsider the inference from causal spread to explanatory spread itself.

5   Representation, Explanation and Non-Communicative Content

We have scouted the idea that it is legitimate to view inner states or processes as at least weakly representational just so long as they yield, in a rich, normal ecological setting, the kind of outcome appropriate to the contentful gloss. In actual fact this is an extreme position which cannot be correct. To see why, consider the case of a trained-up connectionist network that solves a certain problem. And assume that the solution, as normally understood, involves the use of some distributed representations at the hidden unit layer. In line with the genic analogy, we might now argue that a single weighted connection (chosen from the many that cooperatively generate the hidden layer encoding) constitutes a specification (in the weak, genic sense) of how to solve the discrimination problem. For the action of that weighted connection, *in the usual ecological context provided by the other weighted connections*, is clearly sufficient to determine a solution to the problem. But this (to be blunt) is silly: the single weighted connection is not usefully seen as any kind of specification of how to solve the larger problem. What this shows, we think, is that the appeal to ecological context cannot be allowed to run wild, as it seems to do in the case of the lonesome weight.

Here is another problem. We seem to have suggested both that neural, bodily and environmental factors may often be on an explanatory par as regards the causal determination of intelligent behaviour, *and* that various inner states may properly be seen as (weakly) representational, that is, as sets of coded instructions whose action contributes to solving the problem. But there is at least a prima facie tension between these two claims. If the inner states really constitute *instructions*, why isn't that enough to warrant the ascription of a special explanatory role? The recipe is a special element in the cake-baking process, even though we all agree that the presence of an operative oven is an essential contributing factor. If, instead, we argue that the other factors are instruction sets too (thus restoring the parity) we effectively trivialize our notion of genic representation. Our own story, it seems, demands that the inner states often play a special role (that of instructions) in determining behaviour.

The correct thing to say, we think, is that causal co-determination does not imply explanatory parity. The truth behind the suggestions of parity is that the other factors matter and, moreover, that they can matter in ways that we have traditionally thought only the representational factors mattered. What is needed now is an understanding of the ways in which the representational and non-representational contributions may differ (as differ they must, on pain of triviality).

One tempting, but ultimately insufficient move is to appeal directly to a kind of teleological asymmetry.[19] Once again, genes provide our model. DNA structures have often been shaped

---

[19]What mileage there is in this idea is, we think, pretty much exhausted in Clark (1998). The present treat-

and molded, by evolution, *precisely so as* to 'tip the balance' in favour of certain adaptively desirable outcomes. The same is true of neural states, only here the shaping and molding might be achieved not only by evolution, but also by learning. Genes or neural states might thus constitute instructions insofar as they (unlike much of the ecological backdrop) have been selected *in order that* such-and-such a response or behaviour should occur. The cake recipe, ditto, exists precisely in order that cakes ensue, whereas the oven and kinetic energy, we may assume, do not. Or consider another analogy. A glider pilot, in plotting a course from A to B, may well rely on the presence of certain kinds of local structure and 'helping hands' to achieve her goal. She may, for example, factor in the reliable presence of certain updrafts and thermals. Behavioural success thus depends on a combination of her plan and these assumed factors and forces. There is clearly a causal-spread-style, multi-factor story to be told. But for all that, the pilot remains *the pilot*. She is a special element in the causal web insofar as her actions are selected so as to reach the goal — even though, in so doing, various other factors and forces, whose shapes were not thus selected, are co-opted into service.

The trouble is, this story still does not separate the truly representational wheat from the contentless chaff. Consider a robot arm that must reach out and grip a part. The multi-factor story hereabouts might include reference to (a) the force of gravity, (b) a spring-like assembly in the arm and (c) a set of instructions for reaching (in that ecological context). The appeal to teleology distinguishes the force of gravity from the rest. But it leaves the spring-like assembly and the putative instruction set on a par. For both may have been selected precisely so as to enable skilled reaching.

The most we can say, perhaps, is that the brain and central nervous system constitute an especially potent site of teleological plasticity, since they are resources in which minor and cheap-to-implement changes can make a major difference to behaviour: such, in general, is the virtue of programmable software over gross mechanical solutions. But even if it is allowed that the brain (and central nervous system) thus constitute something like 'primary loci of adaptive plasticity', it still remains unclear why we should depict the neural states as representational. A push in the back may, after all, be a cheap and easy-to-implement means of tipping an ecological balance in favour of a certain outcome (a cliff-face plummeting scenario, for example). But the push is not thereby revealed as a context-bound instruction to plummet. It is just an applied force that tips the scales in a certain (downwards) direction. The frog at the bottom of the discursive beer glass is thus revealed. When is an element of an extended causal nexus an internally represented instruction rather than a mere force selected so as to tip a delicate ecological balance in an adaptively valuable direction?[20]

The trouble, at root, is that cognitive science, and philosophy, still lack a decent definition of internal representation. This deficit, as Robert Cummins nicely points out, is often obscured by the fact that there are several theories on offer whose job, as Cummins puts it, is to "pin the meaning on the symbol" (Cummins, 1996, p.66). Such theories try to show why a given symbol should be associated with one meaning rather that another, but they don't tell us why, or when, we should buy into the business of symbols and meanings at all. They don't tell us, in short, what internal representation *is*.

At first glance, of course, the answer is obvious. An internal representation is an inner state whose job is to stand for some other (usually external) state of affairs. This answer works

---

ment aims to distinguish the representational contribution in a more secure way, and to present a more positive alternative to the non-representationalist party line.

[20]This genuine difficulty in maintaining a distinction between internal representations and other forces is what appears to be driving Webb's claim (mentioned earlier) that certain variables in her cricket-robot, variables which correspond to sensory or motor signals, should not be thought of as representations, because their role in connecting sensors to motors is, in effect, equivalent to the role of the gears in connecting the motors to the wheels.

well enough for public representational schemes such as language. Here, the speakers of the language use the symbols to stand for other things. The inner economy, however, is different in one crucial respect: it must operate by means of the direct causal properties of the states and processes. There is no room for inner homunculi that literally understand such-and-such a state as standing-in for such-and-such a state of affairs. Instead, what happens happens, with no process of literal interpretation intervening.

What is needed, then, is a way of understanding inner states as stand-ins that does not commit a kind of homuncular fallacy. This is where a certain notion of strong (i.e. non-genic) representation seems to answer the call. The idea[21] is that the requisite notion of standing-in is provided by the capacity of certain states and processes to figure in episodes of environmentally de-coupled 'off-line' problem-solving. And what this requires, in turn, is that the states and processes constitute a kind of model of the target domain, viz an inner arena in which the unfolding of states marches in step with the way things unfold in the wider world. The notion of strong internal representation is thus the notion of an item or items that figure in an inner model capable of supporting a kind of vicarious problem-solving: one in which moves can be tried out before committing the real flesh and blood to action (see, e.g., Campbell (1974), and Dennett's (1995) notion of a 'Popperian Agent'). The adaptive role of 'standing in for' can thus be cashed out in terms that do not invoke an inner understander, but which make clear contact with the primary (external, public) usage nonetheless.

Genic representation, however, is meant to do duty in a variety of cases in which the capacity for off-line use is not present, viz cases in which the inner resources are delicately keyed so as to make maximal use of bodily and environmental factors and forces so as to promote ongoing adaptive success (e.g. the Franceschini maps). There are, fortunately, several features of the strong representational scenario that can still carry over to such cases, including

I   the presence of kind of arbitrariness in which what matters is not the shape or form of the individual representations themselves, but rather their role as content-bearers;

II   the fact that the (putative) representations, although not actually understood (by the subsystems concerned) *as* content-bearers, are nonetheless consumed by other subsystems in a way which makes sense only if we depict the subsystems as requiring the kinds of information the content-ascribing glosses highlight;

III the presence of an entire system of states whose combinatorics and/or dynamics are somehow appropriate to the target domain.

These three conditions seem to play, in addition, a clear role in many arguments that depict DNA as a kind of code, as will become clear if we pause for a moment to think about the details of protein synthesis.[22]

The first stage of protein synthesis is known as *transcription*, during which the organism's DNA acts as a template for the manufacture of molecules of *messenger RNA (mRNA)*. The second stage is known as *translation*, during which the mRNA molecules produced by transcription determine the manufacture of different proteins. Molecules of mRNA are divided into triplets of nucleotide molecules known as *codons*, and every instance of a particular mRNA codon, as transcribed from its DNA template, is believed to result in an instance of the same amino acid being added to an emerging protein. However, there is nothing in current biological knowledge to suggest a convincing physical/chemical reason why the mappings could not have been set up

---

[21]This is a condensed version of an argument developed at much greater length by Clark and Grush (forthcoming).

[22]For useful introductions to protein synthesis, see, e.g., Hodson (1992) and Maynard Smith (1995).

differently (Crick, 1968). Thus it seems that the mappings from particular nucleotide triplets to particular amino acids are, in the sense we require, *arbitrary* (condition I). In the biological literature, the standard formulation of this idea is to say that the genetic code is thought to be arbitrary. However, from our perspective, that is not the way to put the point. As will become clear later, we understand arbitrariness to be a necessary condition on there being a code *at all*. Therefore, if the mapping in question were not arbitrary, there would be no pressure to think of the system in question as one of encodings. Oversimplifying somewhat, the right question is not, "is the genetic code arbitrary?", but rather, "is there a genetic code?".

Next consider the machinery of translation. Located in the cytoplasm of the cell are protein-manufacturing-sites called *ribosomes*, and molecules of another sort of RNA: *transfer RNA (tRNA)*. Molecules of tRNA are single nucleotide triplets attached to single amino acids. What happens during translation is that an mRNA molecule becomes attached to a ribosome, and then passes through it, one codon at a time. When a new codon moves into place, the ribosome (through trial and error) locates a molecule of tRNA that, according to certain rules (the so-called base-pairing rules), features a particular nucleotide triplet. The ribosome then strips off the amino acid from the other end of the tRNA molecule, and adds it to the protein which it is building. Stripped of its amino acid, the tRNA molecule floats off into the cytoplasm, to be 'recharged' with 'the right' amino acid. It seems that the machinery of ribosomes and tRNA is most illuminatingly understood as a (distributed) *consumer* sub-system that *decodes* the *information* carried by the mRNA (condition II). (That is why, of course, the process is called *translation*.)

Finally, it seems clear that there is a *systematic relationship* between the the sequence of codons that makes up an individual mRNA molecule, and the resulting sequences of amino acids that make up the proteins generated from that mRNA molecule (condition III).

Features I to III thus form, we believe, a coherent complex that pivots on the idea of arbitrary intervening states or processes whose systemic role is to act as the bearers of specific contents. The three features (arbitrariness, content-based consumption, and systematicity) seem nicely applicable to the genic cases, and offer at least a hint of some ways in which even weak internal representation involves more than the mere presence of a contributing force.

That said, the notion of content-based consumption still looms uncomfortably large in the story we have told. Uncomfortably, in that content-based consumption is one of those notions that seems clear and obvious until you try to cash it out. It is easy, of course, to cash it out when you are dealing with agents (like ourselves) who use systems of conventional symbols in which, for example, the word 'dog' is clearly playing a role best understood by reference to content rather than form. Other words, like 'cat', would clearly have done just as well. But, as we noted earlier, it is crucial in this debate that we not confuse the operation of internal representations with that of conventional, communicative codes (such as English or Spanish). For (as we all know) internal representations, unlike communicative ones, must secure their effects *without first being understood*.[23] As Cummins puts it:

> A system exploits an internal representation in the way a lock exploits a key: by being causally affected by its structure. Anything less austere than this will undermine the explanatory interest of internal representations altogether. (Cummins, 1996, p.102)

In particular, Cummins cautions against being misled by the form of speech 'X represents Y to Z' since "this makes representation look like a form of communication" (op cit. 104). Arguments that insist that such-and-such an internal state cannot be a representation since it

---

[23]We all know this, but it is *terribly* easy to pay mere lip service to the idea, especially when we think about the role of content in cognitive-scientific explanation.

doesn't represent anything to the system itself (see, e.g., Beer, 1995; Brooks & Stein, 1993; Harvey, 1992) seem to trade on this very confusion. For example, Brooks and Stein (1993, p.2) caution that the firing of a particular neuron, even if it is non-accidentally associated with, say, the presence of red objects, is not thereby revealed as a representation of redness *for* the system itself. Instead, it represents redness to us, the external observers. Now this notion of representation clearly trades on the idea of a state that presents itself *to* an understanding system *as* having a certain content. Such a notion is substantially stronger than the standard cognitive-scientific use. A better way of carving the cake, in line with Cummins' suggestion, might thus be to say that although the firing of the neurons does not literally communicate anything to any understanding sub-system, the firing may still be functioning as an internal information-bearer and hence as a form of (non-communicative) representation. What is missing is any active appreciation of the content: but the content may be bringing about its effects nevertheless.

Much recent anti-representationalist sentiment is, we agree, rooted in this (mistaken) tendency to equate internal representations with elements in some kind of internal yet still basically communicative economy. But the mistake is not quite as straightforward as Cummins makes it appear. For as soon as we abandon the close analogy with conventional communicative codes, the distinction between a mere internal force and an internal representation does indeed threaten to evaporate. If the inner state *really* brings about its effects the way a key opens a lock then why not conceive the inner state simply as a force (one among many, in fact) that contributes to the bringing about of an effect? For example, once we recognize that internal representations cannot be depicted in essentially communicative terms, the most obvious difference between an instruction and an applied force simply fades away. For an instruction, we would like to say, can be *disobeyed*, whereas an applied force cannot. (It may sometimes fail to achieve its purpose — the push may fail to send the victim over the cliff — but this surely does not amount to the victim's disobeying some enemy's command to fall!) This gap between the understood content and the indicated action is central to the ordinary (communication-based) notion of an encoded instruction: but it cannot, we think, be preserved in our understanding of any genuinely explanatory appeal to an inner code.

The hope, then, is that despite these very real differences, the triple appeal to arbitrariness, consumption and systematicity may provide a robust and workable means of distinguishing the cases (force versus coded instruction), and one that works without committing the communicative fallacy. But even here we must proceed with caution. An element in a communicative language system is arbitrary because its meaning is fixed by convention: the very same item could easily have been used to mean something else. Similarly, coded instructions or representations look like items awaiting active understanding by another sub-agency. How can we unpack these ideas in a way that circumvents the tacit appeal to communicative role?

Consider a robot built by Christopher Longuet-Higgins and described and discussed by Johnson-Laird (1983) and, more recently, by Keijzer (1997). The robot avoids falling off the edge of tables, but it seems to have no table edge detector. Instead, the main wheels drive two small wheels that hold a piece of sandpaper under the base plate. Things are set up so that the small wheels, relative to the sandpaper, echo the location of the real robot relative to the tabletop. Whenever a small wheel hits a ridge at the edge of the sandpaper, a circuit is closed and the robot turns. As a result, the small wheels and sandpaper model the big robot and tabletop. By ceding control to the inner model, the big robot avoids falling off the edge. Commenting on this set-up, Johnson-Laird writes that:

> The small wheels and the piece of sandpaper are not intrinsically a model but they become a model in the robot because of their function as an arbitrarily selected symbolic notation that is used to register the position of the robot in the world. (1983, p. 404, quoted by Keijzer (1997))

The idea we wish to draw out of this is that the inner elements are here counted as arbitrary because the relevant equivalence class of inner systems (the set of different systems that could play the same adaptive role) is fixed not by the first order physical properties of the components (the thickness and weight of the sandpaper etc.) but by their capacity, when appropriately organized, to act as the bearers of a certain (specifiable) body of information. These states bear that information insofar as they support a body of internal dynamics that is fit to track or model a specifiable set of external states of affairs. The inner states thus allow a certain kind of information (one that is not directly transduced by the robot's sensors) to guide the robot's behaviour. The absence of an additional inner element that understands the inner states *as* representations of the proximity of edges is unimportant. For the states do indeed carry that information, and they enable the information to be effective in guiding the behaviour of the larger system. The equivalence class of inner sub-systems that could play the same role includes anything that can bear that information and make it effective within the larger system.

The notion of an inner consumer or content-sensitive sub-system can now be glossed the same way. What matters is that the putative consumer process (the circuit closer in the robot above) is targeted on a certain inner state not because of the state's intrinsic physical properties, but because of the information that it happens to bear. The counterfactuals (which fix the equivalence classes) make it plain: had nature exploited some other inner state as an indicator of the table edge, then the consumer process would have been targeted on it instead (cf. Millikan, 1989).

Internal representations, we conclude, are best seen as inner states or processes whose functional role is to bear certain specifiable contents. And this functional role, in turn, is characterized by (i) the way we are forced to invoke information or content in order to specify the *equivalence classes* of inner organizations that could support the same kind of environmental activity and co-ordination (cf. Pylyshyn, 1986), (ii) the presence of *consumer* subsystems that exploit the inner states for their content, and (iii) the presence of some kind of *combinatoric structure* in the representational states. Whether this last condition is necessary or merely typical we leave open. The presence of combinatoric structure in the inner states contributes, we think, to the appearance of a full inner *code* as against an isolated inner sign or symbol. But notice that the inner processing of the Longuet-Higgins robot (like that of the Franceschini robot) does not rely on classical combinatorics. Instead, it uses a kind of structured inner model to support a variety of content-related encodings. It is probably this latter capacity that matters more than the presence of re-combinable symbols per se.

The three features (arbitrariness, consumption-for-content, and broad systematicity) thus fit together like this: Arbitrariness (the fact that any suitable information-bearer could play the same adaptive role) is what forces us to depict certain states or processes in contentful terms; content-based consumption emerges as a kind of 'inner de-coding' in which some specific information-bearer is enabled to guide behaviour; and systematicity is the economical icing on the informational cake, allowing structurally related inner states to guide different-but-related behaviours.

What, finally, of that lonesome weight? Nothing we have said so far shows how to stop the kind of runaway appeal to ecological context that might allow us to, for example, treat a single weight in a connectionist network as encoding a context-assuming algorithm for solving a larger problem.

A quick response would be to say that, given the constraints developed in the previous section, the lone weight could not count as an internal representation at all (and hence, by extension, could not count as the vehicle of any kind of content). This quick response, however, evades the true thrust of the objection, which concerns the acceptability of a certain practice: the practice of treating certain contributing factors and forces as mere background against which

an instruction set has been selected to act. Thus we may imagine taking a single line from a 30 line LISP program and arguing, in a similar vein, that the lonesome line is in fact a perfectly good algorithm or instruction set for solving the larger problem, albeit one that works only against the powerful backdrop of the other 29 lines of code! What stops the notion of genic encodings degenerating into this kind of triviality?

The answer, we think, is nothing but good sense and explanatory hygiene. One important factor is that the story should isolate a plausible locus of co-ordinated encoding activity. That is to say, it should not focus on a fragment of a larger inner system that was itself the locus of selective activity. In the case of the lonesome weight, the adaptive processes that fixed the value of the lone weight also, and as part of the same adaptive episode, fixed the other weights and connections. In the case of the single line of LISP code, the process that yielded the target line also, and as part of the same adaptive episode, generated the other 29. Under such circumstances it is unacceptably arbitrary, and explanatorily unsatisfactory, to focus on the single weight or line of code.

A second important factor concerns the visible contribution of the target element to adaptive success. Thus suppose we encounter a behaviour that is rich and flexible and yet seems guided by a surprisingly minimal kind of inner instruction set. All the richness and flexibility, let's assume, results from the operation of other factors and forces (such as the spring-like action of the muscles and a benign operating environment). In this kind of case there would be little value in treating the inner contribution as a kind of context-exploiting algorithm for producing the behaviour. For *all* the richness and flexibility of the behaviour is, by hypothesis, here rooted in the operation of the other factors: and it is the richness and flexibility, surely, that the appeal to internal representations and instructions is meant to explain. And to be fair, much of the work of Kelso, Thelen and Smith, and others seems to us to have this flavour. It often involves, for example, the use of rich inbuilt synergies (fixed couplings between systemic elements) whose presence permits very simple neural 'commands' to generate complex and flexible behaviours. The effect, as Kelso (1995, p.38) notes, is like taking a 4 wheeled vehicle and coupling the wheels so as to allow control via a single steering motion: a much simpler solution than issuing a whole set of subtly varying commands to control each wheel. The effect is also, Kelso suggests, like:

> An army general saying 'take hill eight' with the many subordinate layers of the military (subsystems) carrying out the executive command. (Kelso, 1995, p.38, reporting an image due to Peter Greene)

Now this image, it is true, can also be deployed within the more traditional representationalist camps, so as to suggest a kind of stack of information-processing homunculi (see, e.g., Dennett, 1978). But Kelso's point is that the subordinate layers need not be relying on information-processing at all. Instead, they may be exploiting inbuilt synergies, local context, internal self-organization and the whole gamut of possibilities described in section 2 above. (Goodwin's model of genetic control is exactly parallel in this respect.)

The general's command is, at best, a kind of limiting case of an encoded instruction set. It does not really specify a solution to the problem. It is more like a request that a solution be found and implemented! To depict it as a good specification of a solution *in context* (in the context, that is, of all the factors and forces that do the real work) is misleading to say the least. Nor does a single line of LISP code specify a solution in the context of the other 29. By contrast, suppose the full 30 lines of code specify a whole sequence of moves that will solve a problem just so long as there are certain spring-like couplings between muscles, sensor physics etc.. Here, it seems explanatorily acceptable to depict those additional characteristics as the effective backdrop against which the encoded solution was selected to operate.

6   Conclusions: a Spectrum and a Puzzle

Causal spread obtains when some phenomenon of interest turns out to depend, in unexpected ways, upon factors external to the system previously/intuitively thought responsible. Thus, we encounter causal spread when we learn that the sex of the offspring of the Mississippi alligator is determined not by direct genetics but by the ambient temperature of the rotting vegetation in which they lay their eggs. Of course, it is allowed that in *all* cases there will be some crucial work done by external factors and forces. In the case of, for example, ordinary genetic sex determination, there may be external influences (such as diet and the Ph. balance of the vagina in humans). No algorithm, as an anonymous referee usefully points out, ever works totally independently of environmental conditions (see also Clark, 1998). In most cases, however, we can usefully think of the environment as providing simply the conditions for the successful application of the algorithm (e.g. a reaching algorithm that works only given Earth gravity). Sometimes, however, the *kind* of work done by additional factors and forces turns out to be precisely the kind of work previously assigned to a more isolated, typically inner system (recall, once more, the Mississippi alligator). In the cases where causal spread threatens to reconfigure ordinary cognitive scientific explanations of adaptive intelligence, one means by which we may distinguish such non-trivial spreading concerns the role of external factors and forces in supporting the kind of rich, flexible, context-appropriate behaviours that normally invite a story about detailed inner programs, sets of rules, connectionist pattern-matching etc.. To be sure, there will always be room to argue about specific cases. Maybe not everyone expected all the work of binding (see section 3) to be achieved by linking sites in internal memory. And the rough criteria we have suggested do make the identification of non-trivial causal spread depend somewhat on the classes of models and explanations that previously held sway. But we think the central idea is clear enough to work with, and we are in no doubt but that something like the argument from causal spread lies behind many recent treatments of situated and embodied cognition.

If we are on the right track, however, not every causal-spread-engendered reconfiguration of our cognitive-scientific understanding signals a tolling of the bell for inner representations. The notion of genic representation, we think, shows how a representational explanatory strategy can not only operate in the midst of non-trivial causal spread, but also contribute, in important ways, to our understanding of the phenomena under investigation. The notion, it should by now be clear, is not really needed at either end of an explanatory spectrum. At the most minimal (lonesome weight) end, the appeal to genic encoding is unhelpful as it obscures the true sources of flexibility and intelligent response. At the more maximal (full LISP program) end, the standard notion of a relatively full internal instruction set seems to apply. The home of genic representation lies in the presently uncharted, but potentially vast middle ground: the very ground that seems to be increasingly mandated in attempts to understand the role of real genes in the generation and explanation of biological form. In the cognitive case, the nature and extent of this middle territory is still up for grabs. But the general lie of the landscape is clear: it is characterized by achievements that still rest fairly heavily on the use of internal information-bearing resources, but that rely on a variety of additional factors and forces to take up the substantial slack left by the information-based story. One other point bears repeating. It is that in taking up this slack, the other factors and forces must reveal themselves as the unexpected root of a good deal of the *kind* of flexibility and subtlety normally associated with representation-based control.

Such a vision bequeathes a puzzle. Should we say that in such cases what we learn is that much of the flexibility and richness normally attributed to *cognitive* processes turns out to be grounded in non-cognitive factors, thus effectively identifying the notion of a cognitive process with that of a representation-guided one? Or should we expand the notion of the cognitive to

include a variety of non-representational elements?[24] Restricting the notion of cognition to that which is representationally mediated preserves long-standing intuitions about the link between cognition and something more like reflective reason. But various non-representational factors and forces seem to play a major role in enabling the kinds of fluid, context-sensitive response that distinguish biological intelligence from mere computational simulcra (such as expert systems). The very idea of the cognitive domain is, we think, somewhat shakily poised between these two extremes. Our own view is that neither option is really mandated by our current understanding of the notion of a cognitive process, and that the choice will thus be dictated largely by fashion, chance and rhetoric. So, for today at least, we choose to remain agnostic on this issue.

More concretely, we hope to have shown both the urgency and the difficulty of balancing representation-based understanding against the spectre of causal spread. We are convinced that the bulk of daily on-line problem-solving is indeed characterized by a variety of complex and deeply interactive exchanges between brain, body, and world. Doing justice to this delicate ecology without abandoning the explanatory gains of four decades of cognitive-scientific research is, we believe, a vital part of the current effort to re-position the mind more intimately in the web of body and world.

Department of Experimental Psychology
South Parks Road
University of Oxford
Oxford
OX1 3UD
UK
E-Mail: michaelw@psy.ox.ac.uk

Philosophy-Neuroscience-Psychology Research Program
Philosophy Department
Washington University
One Brookings Drive
Campus Box 1073
St.Louis
Mo 63130
USA
E-Mail: andy@twinearth.wustl.edu

Acknowledgements

---

[24]For the former view, see Crane (1995, especially p.191). For the latter, see van Gelder and Port (1995).

References

Agre, P. (1988). *The Dynamic Structure of Everyday Life.* Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20:4.*

Beer, R. D. (1995). Computational and dynamical languages for autonomous agents. In (Port & van Gelder, 1995), 121-47.

Boden, M. A. (Ed.). (1996). *The Philosophy of Artificial Life.* Oxford University Press, Oxford.

Brooks, R. A. (1991). Intelligence without reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pp. 569–95 San Mateo, California. Morgan Kauffman.

Brooks, R. A., & Stein, L. A. (1993). Building brains for bodies. Artificial Intelligence Lab Report 1439, MIT.

Campbell, D. (1974). Evolutionary epistemology. In Schilpp, P. (Ed.), *The Philosophy of Karl Popper*, pp. 413–63. Open Court, Le Salle, IL.

Clark, A. (1997a). *Being There: Putting Brain, Body, and World Together Again.* MIT Press / Bradford Books, Cambridge, Mass. and London, England.

Clark, A. (1997b). The dynamical challenge. *Cognitive Science, 21:4*, 461–81.

Clark, A. (1998). Twisted tales: Causal complexity and cognitive scientific explanation. *Minds and Machines, 8*, 79–99.

Clark, A., & Grush, R. Towards a cognitive robotics. Forthcoming, in *Adaptive Behavior.*

Clark, A., & Toribio, J. (1994). Doing without representing. *Synthese, 101*, 401–31.

Clark, A., & Wheeler, M. (1998). Bringing representation back to life. In (Pfeifer, Blumberg, Meyer, & Wilson, 1998), 3-12.

Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.). (1994). *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, Cambridge, Mass. MIT Press / Bradford Books.

Cliff, D., & Noble, J. (1997). Knowledge-based vision and simple visual machines. *Philosophical Transactions of the Royal Society: Biological sciences, 352 (1358)*, 1165–75.

Crane, T. (1995). *The Mechanical Mind.* Penguin, London.

Crick, F. (1968). The origin of the genetic code. *Journal of Molecular Biology, 38*, 367–379.

Cummins, R. (1996). *Representations, Targets and Attitudes.* MIT Press, Cambridge, Mass.

DeLisi, C. (1988). The human genome project. *American Scientist, 76*, 488–93.

Dennett, D. (1978). *Brainstorms.* Harvester Press, Brighton.

Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* Penguin, London.

Franceschini, N., Pichon, J. M., & Blanes, C. (1992). From insect vision to robot vision. *Philosophical Transactions of the Royal Society, series B, 337*, 283–94.

Globus, G. (1992). Towards a non-computational cognitive neuroscience. *Journal of Cognitive Neuroscience, 4*, 299–310.

Goodwin, B. (1994). *How the Leopard Changed its Spots: the Evolution of Complexity*. Phoenix, London.

Harvey, I. (1992). Untimed and misrepresented: Connectionism and the computer metaphor. Cognitive science research paper 245, University of Sussex. Reprinted in *AISB Quarterly*, 96, pp.20-7, 1996.

Hodson, A. (1992). *Essential Genetics*. Bloomsbury, London.

Hurley, S. *Consciousness in Action*. Forthcoming, Harvard University Press.

Husbands, P., Harvey, I., & Cliff, D. (1995). Circle in the round: State space attractors for evolved sighted robots. *Robotics and Autonomous Systems, 15*, 83–106.

Husbands, P., & Meyer, J.-A. (Eds.). (1998). *Evolutionary Robotics: Proceedings of the First European Workshop, EvoRobot98*, Vol. 1468 of *Lecture Notes in Computer Science*, Berlin. Springer.

Johnson-Laird, P. (1983). *Mental Models*. Cambridge University Press, Cambridge.

Keijzer, F. (1997). *The Generation of Behavior: On the Function of Representation in Organism-Environment Dynamics*. Ph.D. thesis, University of Leiden.

Kelso, J. A. S. (1995). *Dynamic Patterns*. MIT Press / Bradford Books, Cambridge, Mass. and London, England.

Lewontin, R. (1983). The organism as the subject and object of evolution. *Scientia, 118*, 63–82.

Mataric, M. (1991). Navigating with a rat brain: a neurobiologically inspired model for robot spatial representation. In (Meyer & Wilson, 1991), 169-75.

Maturana, H. R., & Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. New Science Library, Boston.

Maynard Smith, J. (1995). *The Theory of Evolution*. Cambridge University Press, Cambridge. Canto Edition, First Published 1958, Second Edition 1966, Third Edition 1975.

Meyer, J.-A., & Wilson, S. (Eds.). (1991). *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*, Cambridge, Mass. MIT Press / Bradford Books.

Millikan, R. (1989). Biosemantics. *Journal of Philosophy, 86:6*, 281–97.

Pfeifer, R., Blumberg, B., Meyer, J.-A., & Wilson, S. (Eds.). (1998). *From Animals to Animats 5: the Fifth International Conference on Simulation of Adaptive Behavior*, Cambridge, Mass. MIT Press / Bradford Books.

Port, R., & van Gelder, T. (Eds.). (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press / Bradford Books, Cambridge, Mass.

Pylyshyn, Z. (1986). *Computation and Cognition*. MIT Press, Cambridge, Mass.

Rocha, L. (1996). Eigenbehavior and symbols. *Systems Research, 13:3*, 371–8.

Smithers, T. (1995). Are autonomous agents information processing systems?. In Steels, L., & Brooks, R. (Eds.), *The Artificial Life Route to Artificial Intelligence*, pp. 123–62. Lawrence Erlbaum, Hillsdale.

Thelen, E., & Smith, L. B. (1993). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, Mass.

van Gelder, T. The dynamical hypothesis in cognitive science. Forthcoming in *Behavioral and Brain Sciences*.

van Gelder, T. (1995). What might cognition be if not computation?. *Journal of Philosophy*, *XCII*(7), 345–81.

van Gelder, T., & Port, R. (1995). It's about time: An overview of the dynamical approach to cognition. In (Port & van Gelder, 1995), 1-43.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, Mass. and London, England.

Webb, B. (1994). Robotic experiments in cricket phonotaxis. In (Cliff, Husbands, Meyer, & Wilson, 1994), 45-54.

Webb, B. (1996). A cricket robot. *Scientific American, 275:6*, 62–7.

Wheeler, M. (1994). From activation to activity: Representation, computation, and the dynamics of neural network control systems. *Artificial Intelligence and Simulation of Behaviour Quarterly, 87*, 36–42.

Wheeler, M. (1996). From robots to Rothko: the bringing forth of worlds. In (Boden, 1996), 209-36.

Wheeler, M. (1998). Explaining the evolved: Homunculi, modules, and internal representation. In (Husbands & Meyer, 1998, pp.87-107).