An Interactivist-Constructivist Approach to Intelligence: Self-Directed Anticipative Learning

W.D. Christensen & C.A. Hooker

Abstract This paper outlines an original interactivist-constructivist (I-C) approach to modeling intelligence and learning as a dynamical embodied form of adaptiveness and explores some applications of I-C to understanding the way cognitive learning is realised in the brain. Two key ideas for conceptualising intelligence within this framework are developed. These are: (i) intelligence is centrally concerned with the capacity for coherent, contextsensitive, self-directed management of interaction, (ii) the primary model for cognitive learning is anticipative skill construction. Self-directedness is a capacity for integrative process modulation which allows a system to 'steer' itself through its world by anticipatively matching its own viability requirements to interaction with its environment. Because the adaptive interaction processes required of intelligent systems are too complex for effective action to be prespecified (e.g. genetically) learning is an important component of intelligence. A model of self-directed anticipative learning (SDAL) is formulated based on interactive skill construction, and argued to constitute a central constructivist process involved in cognitive development. SDAL illuminates the capacity of intelligent learners to start with the vague, poorly defined, problems typically posed in realistic learning situations and progressively refine them, transforming them into problems with sufficient structure to guide the construction of a solution. Finally, some of the implications of I-C for modeling of the neuronal basis of intelligence and learning are explored; in particular, Quartz and Sejnowski's recent neural constructivism (NC) paradigm, enriched by Montague and Sejnowski's dopaminergic model of anticipative-predictive neural learning, is assessed as a promising, but incomplete, contribution to this approach. The paper concludes with a four-fold reflection on the divergence in cognitive modeling philosophy between the I-C and the traditional computational information processing (CIP) approaches.

1. Introduction

This paper outlines an original interactivist-constructivist (I-C) approach to modeling intelligence and learning, and explores some applications of I-C to understanding the way cognitive learning is implemented in the brain. I-C is a dynamical

W. D. Christensen, Department of Philosophy, University of Newcastle, Callaghan, NSW, Australia 2308. Email: plwdc@cc.newcastle.edu.au C. A. Hooker, Department of Philosophy, University of Newcastle, Callaghan, NSW, Australia 2308. Email: plcah@cc.newcastle.edu.au embodied approach which falls broadly within the recently resurgent school emphasising adaptive interaction rather than internal computation.[1] However, there is much controversy concerning just what, if anything, claims to a 'new non-cognitivist paradigm' might amount to (e.g. the commentaries on van Gelder 1998). Whilst a suite of new tools and a number of important issues for modeling intelligence are highlighted by dynamical and adaptive systems approaches, a consistent weakness is the absence of a *positive conception* of what intelligence and cognition are (cf. Keijzer and Bem 1996). This paper addresses the issue by initially outlining some principles for modeling the complex embedding relations involved in even simple adaptive behaviour, and then by developing two key ideas for conceptualising intelligence within the dynamically oriented I-C framework. These are: (i) intelligence is centrally concerned with the capacity for coherent, context-sensitive, self-directed management of interaction, and cognitive processes are the high order modulatory processes that produce this capacity, (ii) the primary model for cognitive learning is anticipative skill construction. The upshot of these features is that (iii) intelligence involves a sophisticated form of the root process of gradient tracking, with self-directed interaction arising as gradients become increasingly internally constructed and modified.

Living systems must so interact with their environment as to acquire, within their means, the resources they need for survival, and must so interact within themselves that these resources are used to regenerate themselves, including these capacities. In general there are many variant interactions available, both within living systems and in their environment, most maladaptive. Thus the basic problem intelligent systems have to solve is the coherent contextsensitive management of interaction processes with many degrees of freedom. Selfdirectedness is theorised as the capacity for high order integrative process modulation geared to solving this problem. Self-directed systems anticipate and evaluate interaction processes, generating and modifying action context-sensitively in order to achieve specific interaction goals (e.g. catching the gazelle) and maintain fundamental system parameters (e.g. nutrition). Learning is an important component of intelligence because of the complexity that intelligent systems must adaptively cope with. Instructional prespecification, such as through genetically determined phenotypic (including cognitive) 'modules', isn't an effective strategy for dealing with this type of complexity. Indeed, the adaptive problems that intelligent systems must solve are *typically* poorly or incompletely specified; the details must be filled in context-sensitively. Intelligent learners must start with vague, poorly defined, problems ('catch some food') and progressively refine them ('catch this gazelle'), transforming them into problems with sufficient structure to guide the construction of a solution ('run this way here').

These problems take the form of a requirement for skillful bodily interaction and hence skill construction is seen as a primary model for understanding learning processes. This represents a shift away from internal computation of a correct solution-algorithm as the core model of learning to the dynamics of generating and modifying bodily interaction processes. This framework brings to the fore what searching for algorithms suppresses, namely the ways in which the ongoing interaction process itself generates information for the interacting system which it can use to further modify subsequent interaction. Among the adaptive flexibilities this provides, it allows such systems to progressively construct improved anticipations of these processes, thus simultaneously improving their position as they move toward their goal *and* their capacity to get there. (A cheetah learns the dodging characteristics of its prey even while closing in on it.) This makes it possible for systems to flexibly refine and solve initially vague problems.

Such vague problems are ubiquitous for living creatures because their root problems are viability problems of which they have no understanding, yet which they must transform into specific sensori-motor problems within their performance capacities; hence it is the essence of survival that there should be processes that both guide action and improve the capacity to guide action while doing so. However the CIP approach subverts this perspective because, among other things, it deals in algorithms whose precision assumes explicitly defined problems to which they are eternal, optimal, solutions. Instructively, even when dual action/learning processes are broached, as in the illuminating work by Montague et al. discussed in Section 4 below, a CIP perspective leaves the cognitive character and significance of using action to shape the learning problem unclear (see Section 5 below).

Processes that both guide action and improve the capacity to guide action while doing so are here taken to be the root capacity for all intelligent systems. An account of selfdirected anticipative learning (SDAL) is formulated in these terms and argued to constitute a central constructivist process involved in both cognitive development and performance. SDAL involves the generation of high order anticipative structures that improve self-direction in ways that help complex organisms satisfy life regulatory processes.

After setting out these ideas in sections 2 and 3, section 4 explores some of the implications of the I-C approach to intelligence for modeling the neuronal basis of intelligence and learning. The neural constructivism (NC) paradigm of Quartz and Sejnowski (1997) is assessed as a promising, but incomplete, contribution to this type of approach. NC argues that empirical evidence concerning activity-directed plasticity in neuronal connectivity supports a constructivist model of learning, which in turn provides a direct link between neuronal and cognitive organisation. The great benefit of this research is that it demonstrates strong empirical grounds supporting a constructivist approach to intelligence and cognition. The primary deficits of NC are that, though constructivist, it retains a CIP formulation which contributes to an almost exclusive focus on internal brain processes and, partly as a result, it lacks an account of the way learning is shaped by interaction processes, and lacks an account of the type of high order modulatory processes required for SDAL. Other neural learning models involving attentional and motivational modulation of learning processes are examined as example cases which may provide a basis for remedying these problems, in particular a complementary error prediction model of Montague et al.

Our purpose in discussing neural models is both to explore the applicability of the I-C conception of intelligence developed in Section 3 and to reinforce the

relevance of its conceptual analyses by showing how they can usefully illuminate current empirical work. Readers approaching this paper from an empirical modeling background might prefer to first read Section 4 and then return to consider the conceptualisation offered in Sections 2 and 3.

The paper concludes with some broader philosophical reflections on the fundamental assumptions underlying the I-C adaptive approach to understanding intelligence, including their divergence from the CIP approach.

2. Core principles for an interactivist-constructivist model of intelligence

It is becoming increasingly common to note, as a significant theoretical issue for cognitive science, that brains are embedded in bodies and bodies are embedded in environments (e.g. van Gelder 1998). I-C shares this concern with the embeddedness of intelligence, and attempts to set it out in a systematic way by modeling intelligence and learning as a natural part of a more general, biologically based, account of adaptive systems and behaviour. By grounding intelligence and learning in an adaptive context the goal is to illuminate both the genesis of intelligence and its organisational characteristics as an embodied phenomenon. A biologically based approach to modeling intelligence turns out to produce a very different conception of what the central modeling issues are than that underlying (often implicitly) mainstream cognitive science models. Our first task is to make clear what these founding differences are. In this section we will outline the basic ingredients required for dynamically modeling adaptive systems, and we will contrast different ways of being adaptive, highlighting the general characteristics of the organisational mode involved in intelligence.

2.1. Modeling the problem holistically

The most fundamental and challenging aspect of modeling intelligence in an adaptive context lies in capturing the various kinds of holistic relations that are involved. We identify three kinds of holistic relations that are relevant: (i) the fundamental organisation of adaptive systems to achieve a sustained interactively based bodily integrity we call *autonomy*, (ii) the distribution of the dynamical organisational factors involved in adaptive interaction through the system-environment complex, and (iii) action coherency across multiple constraints and timescales. These holistic relations crisscross the boundaries drawn by standard models of adaptiveness and intelligence, requiring both that more of the system-environment context be included than is usually the case and that the conceptualisation of the processes involved be modified in important ways. As we will argue, incorporating these factors leads to a shift away from decontextualised conceptions of adaptive processes as locally modular unitary adaptations and/or optimising computations towards an integrated conception of adaptiveness as dynamic organised interaction in a complex environment satisfying many constraints simultaneously, over extended periods. We now discuss these three aspects in turn.

2.1.1. Autonomy. To understand intelligence biologically it is necessary to begin by focusing on the fundamental predicament of being alive. Living things are delicate and dissipative, features that might be expected to lead to their rapid disintegration. The secret to living lies in having a basic organisation that mitigates these perils for a time, which we call autonomy. Living things are delicate because their chemical bonds are of very low energy, easily disrupted by this world's physical forces and other biological agents; living systems need to constantly avoid such disruptions and repair themselves when they do occur. Living things are dissipative because they are constituted by far-from-equilibrium thermodynamic processes; they must constantly seek out sources of ordered free energy with which to replenish dissipated cellular structures and sustain the capacity for the processes that acquire these resources (e.g. foodsearch), and repair damage (e.g. reconstituting damaged tissue) or avoid damage (e.g. escaping a predator). Living creatures must do all this using only their own decidedly finite bodily capacities, and in the face of much ignorance about what these capacities really are and what the world is like. The basic problem of living then is: how to use one's own capacities to manage one's interaction with the world, and within oneself, so as to achieve these goals. Autonomous systems are those whose overall organisation is such as to do this. Autonomy is a subtle, enormously complex global requirement on the whole organism, for its constellation of processes must continually so interrelate as to regenerate the whole of itself.[2]

The significance of this for developing models of adaptiveness is that the basic normative constraint on adaptive processes is a global one; they must interrelate in globally organised patterns focused on the autonomy of the system. All of the more specific normative constraints on particular actions (e.g. avoid hunger, pain) derive from this global constraint. This contrasts with standard models which characterise normative constraints locally: selectionist adaptive models in terms of separate correspondences between individual traits and environmental features, and CIP models in terms of self-contained input-output optimisation problems.

2.1.2. Distributed dynamical organisation. Another important class of holistic relations involved in adaptiveness concern the distribution across system and environment of the dynamical organisational factors involved in the processes of adaptiveness. Standard models of adaptiveness and intelligence associate the generation of adaptive organisation with centralised classical control (Hooker etal. 1992), with control located either in the genes (selectionist models) or in the brain-mind (CIP models). However this centralised control picture is being challenged on both these fronts. In biology it is being increasingly recognised that gene action operates *within*, and acts as a modulator *on*, developmental processes that have their own rich dynamical organisational characteristics (e.g. Jablonka and Lamb 1995, Raff 1996). Indeed, it has been argued that the range of factors involved in generating the adaptive phenotype includes such a rich set of extra-genetic developmental and environmental resources that biological theory should recognise development as a holistic dynamical process rather than as a genetic control process (e.g. Oyama 1985, Griffiths and Grey 1994). Whether the evidence warrants an extreme holism in which there are no distinctive sub-processes within the full

developmental complex is debatable, but the evidence for distributed organisation is sufficiently strong to seriously undermine the gene-control approach.

Likewise, in cognitive science dynamical embodied approaches have challenged the centralised control picture of cognition. For instance, Brooks (1991) has argued that the centralised senseplan-act model of intelligence presents an architectural form that is inherently clumsy and slow, and has supported this argument by constructing highly functional robots that employ a distributed 'subsumption' architecture. Bickhard (1992) and Clark (1998) have pointed to the widespread role of 'scaffolding' in adaptive behaviour, in which the system uses organisation in the environment to shape interaction in ways that achieve the appropriate outcomes. Other researchers have drawn the conclusion that the factors involved in intelligent action are sufficiently distributed and rich in temporal structure that dynamical systems theory, rather than information processing, is the appropriate general theoretic framework for cognitive science (e.g. Beer 1995, Smithers 1995, Smith and Thelen 1993, van Gelder 1995, 1998). Again, it is debatable whether the extreme holism of dynamical systems theory is warranted, at least as the appropriate general framework for cognitive theory (see Section 5 below), nonetheless, the evidence that the processes of intelligent interaction involve extensive distributed dynamical organisation is strong.

The parallel between the debates in the two domains is striking, and suggests that in general centralised control models—attractive as a first approximation because they identify a single or small number of determining factors for adaptive organisation—are too simplistic. Though the form it will ultimately take is unclear, some form of holistic conceptual framework recognising distributed processes is required. The approach we are proposing offers a unified solution in the form of an account of adaptive organisation which bridges the biological and cognitive domains.

2.1.3. Action coherency across multiple constraints and timescales. If the processes of adaptiveness are dynamical they are, of course, temporally extended. However the autonomy account presented above adds to the very general concern with timing emphasised by advocates of dynamical models (cf. van Gelder and Port 1995) an additional and more specific set of issues. A creature's problem is to match the constraints of autonomy up with the organisation of the environment in the appropriate way across space and time through interactive relationships. In this setting the way the basic adaptive problem presents itself to creatures is as the requirement for organised whole activities. Such activities are complexes of processes across multiple timescales (hunting to the kill, defending a territorial range, raising cubs) that must satisfy multiple constraints both simultaneously and across time (avoiding injury whilst acquiring food, acquiring enough food to sustain both self and cubs, etc.). Thus, adaptive problems do not come as small, well defined, independently optimisable, units as textbook problems do. Generating adaptive action is therefore more like continuously modulating an extended process, rather than assembling complex activity from individually well-formed scripts. The human baby learns to crawl when a holistic functional movement emerges (lifting and repositioning limbs and torso whilst continuously adjusting center of balance and forward motion) which it

can modulate, not by computing and assembling discrete motor actions, such as calculating the next optimal incremental shift of each limb (Smith and Thelen 1993).

2.1.4. *The contrast with standard cognitive science.* In Section 5 we will use these issues as a basis for a critique of some of the standard assumptions within cognitive science. Here we note that incorporating these holistic relations results in a formulation of intelligence that is very different from traditional CIP accounts. Traditionally intelligence is conceived in terms of solving problems and modeled as providing formally correct solutions to many local, semantically interpretable, independently optimisable, formally well defined problems (e.g. Newell 1980a, b). The problem/solution domain lies wholly within the mind arena, the sensory and motor interfaces with the body being separate matters and the environment separate again. Instead here intelligence is understood to be a continuous management process that possesses the three kinds of holisms mentioned (the need to achieve autonomy, distributed organisation, and the need to produce functionally coherent activity complexes). Cognitive processes are embedded within an autonomous system involved in organised dynamic interaction processes in a structured environment and they operate by continuously modulating (rather than controlling) these interaction processes so as to be coherent across multiple constraints and timescales.

2.2. Adaptive management strategies

In this conception of richly embedded adaptive processes, process modulation plays a central role. Adaptive systems must shape an already dynamic interaction process by applying action which modifies the interaction flow in organised ways, rather like the way inserting a stick into a fast flowing stream modifies the pattern of the water flow. To understand such processes it is important to model the complex interrelations between the processes that generate action and the effects that result. We shall term the signals which a system uses to differentiate an appropriate context for performing action the system's explicit norm signals. For example, hunger signals differentiate blood sugar levels and act to initiate and focus foodsearch activity. Explicit norm signals provide information about appropriate action because they differentiate more and less discrepancy between some current system condition and a reference condition, the norm satisfaction state, which modulates subsequent performance.[3] For viability, these norm signals should reflect aspects of the system's autonomy conditions, in the way that hunger reflects nutrition levels. Process modulation always involves at least one or a few explicit norms since there must be some specific dynamical comparison basis which differentiates the context in which action is produced, even though many of the effects of the action may not be internally differentiated by the system.

Organisms typically possess an array of norm signals, many of which can be simultaneously relevant in a given context (cf. cheetah hunting). We shall refer to the full array of performance norm signals a system possesses as its *norm matrix*. These norms may often conflict, as when thirst or pain motivates the cessation of hunting while hunger motivates its continuance. The norm matrix thus establishes

a web of tensions which the system must continually balance by modifying its interaction processes so as to 'steer' itself along a path that provides sufficient satisfaction of all relevant performance norms. If the system is adaptively successful this dynamic modulation of activity will shape interaction in ways that satisfy the fundamental conditions of viability for the system.

A large part of understanding adaptiveness in this picture involves modeling the way the system manages, by modulating its actions, the interaction patterns that are generated. Characterising intelligence as a form of adaptiveness then becomes framed as characterising a particular type of *management strategy*. A management strategy is an organisational recipe for generating the interactional outcomes the system requires. It involves the interaction of the system's norm matrix, action generation processes (including anticipative modeling, see Section 3 below), and interaction dynamics. In order to characterise intelligence we shall be principally concerned with distinguishing *low order* and *high order* management strategies.

2.2.1. Low order management: the mosquito. Although adaptive action must satisfy complex holistic constraints, it is not necessary that the system explicitly recognise all or even a significant proportion of these constraints in its modulatory processes. A low order management strategy employs normative signals which explicitly differentiate only a narrow slice of the overall interaction process as it relates to the constraints the system faces. For instance, female mosquitos use heat tracking and chemotaxic processes, including flying up carbon-dioxide (CO_2) gradients, to home in on blood hosts (see Klowden 1995). Females must feed on blood prior to laying eggs because the blood meal supplies necessary protein for egg development; this is the overall criterion for achieving adaptive success (in contrast, males consume nectar). However the connections between CO₂ gradient tracking, blood hosts and egg production need not be recognised by the mosquito. CO₂ gradient tracking can be effectively achieved simply by sampling local CO₂ concentration and modifying flight in the direction of highest concentration. That this action generates an interaction process which allows the mosquito to find a blood host depends on a set of further relations, including the fact that blood hosts like mammals also emit characteristic chemical signals such as perspiration and carbon dioxide, and the concentration gradients of these chemicals can be used to locate the blood host. Thus, the mosquito's adaptive management strategy is low order because the information it uses to modulate its actions concerns only a very narrow aspect of the interaction process; most of the relations on which it depends are implicit.[4]

2.2.2. *High order management: the cheetah.* Because a low order management strategy uses only a few parameters to direct action its success depends on simple regularity in the environment. However if the interaction processes the system must engage in for adaptive success possess many degrees of freedom, and are therefore complex and variable, a low order strategy will be inadequate. The system must differentiate more aspects of the interaction process in order to produce effective action. The action's success still depends on regularity in the environment, though now it will be more complex and subtle. Cheetah hunts, e.g., are rapid,

uncertain and dangerous; they must respect a range of practical norms, such as avoiding debilitating physical damage during a chase (like breaking a leg) and during the kill (e.g. from a kick), avoiding exhaustion before a kill is assured and avoiding a kill in insecure circumstances (e.g. where the food might be taken by others). In consequence, a cheetah must select an appropriate type of animal to target (such as a gazelle), an appropriate context (young or small isolated animals are preferred, in conditions permitting both stalking and rapid, safe chasing), and must attack using an effective technique (using cover, fluid movement and observation of the prey's attention for effective stalking, initiating the chase from a sufficiently close distance, rapid killing by crushing the throat of the prey, etc.) Despite the great speed and agility of cheetahs only half of their hunts are successful (see Eaton 1974).

To manage the complexity, variability and danger of their hunting cheetahs possess far more complex sensori-cognitive-motor processes than the internal processes of mosquitos, and this provides them with the type of high order modulatory capacity required to continuously integrate the many factors involved in producing effective action. Indeed, cheetahs show a complex interplay between interaction dynamics, internal affective norm processing and action generation. In particular, they are able to evaluate their own performance and use information from interaction to improve performance. Gaining the skills required for successful hunting requires extensive learning. As cubs, cheetahs spend a great deal of time learning hunting skills by playing with siblings, chasing lizards, and so forth. The mother facilitates this process by bringing small live prey, such as a hare, back to the cubs, allowing them to practice chasing and killing techniques. As the cubs begin to mature they accompany the mother on hunts and observe the real process first-hand. Even so actual hunting experience is required before proficiency is achieved; many juveniles, for instance, make the mistake of initiating the chase from too great a distance. The hunting capacity of a mature cheetah is thus a complex product of an extended history of mutual shaping between internally generated action and the success and failure of the ensuing interaction processes.

Sophisticated action requires a high order modulation capacity and on our account, to be given in Section 3, *degree* of intelligence corresponds roughly with degree of capacity for high order integrative process modulation. Cheetah hunting is a paradigm animal example of intelligent, intentional action, and its sophisticated context-sensitivity stems from the cheetah's capacity for high order integrative process modulation. Because process modulation generally is geared to serving the requirements of autonomy, as it becomes increasingly normatively elaborated it yields an increasingly integrated sense of 'self' acting in this way. According to our account, this is the essence of intelligence. This is a more complex framework for understanding intelligence and learning than the standard CIP framework, but we show in Section 3 how this context provides important insights for understanding the dynamics of learning processes. As we emphasised at the outset, adopting an adaptive systems perspective motivates a shift in the basic model of learning from internal formal computation to anticipative managerial skill construction.

3. What is intelligence? Self-directedness and self-directed anticipative learning

3.1 Self-directedness

Within this basic theoretical context we can now begin to outline an I-C account of intelligence. Although the term intelligence is often used in a fairly loose way, at times as virtually equivalent to adaptiveness (not least by non-cognitivist dynamics and robotics-oriented researchers), there are important distinctions that need to be drawn. The characteristic which most strongly suggests intelligence is the capacity for fluid, adaptable, context-sensitive (in some cases 'insightful') action. We are more likely to call the actions of a chimpanzee building a social network of allies intelligent than we are a spider building a web, even though they are both striking instances of complex adaptive behaviour. Likewise, to return to the earlier examples, a cheetah's hunting ability seems more intelligent than a mosquito's procedure for finding blood hosts, even if the mosquito were to prove equally (or more) successful.

In this section we develop an account of self-directedness as the basis for the type of context-sensitive adaptability shown by cheetahs and chimpanzees, but not mosquitos and (mostly?) not spiders.[5] Self-directedness is based on a constructive capacity for high order integrative process modulation geared to managing interaction processes with many degrees of freedom. *Self-directed systems anticipate and evaluate the interaction process and modulate system action accordingly*, thus generating action context-sensitively in order to achieve specific interaction goals (e.g. run smoothly, catch the gazelle) and thereby satisfy system norms (e.g. freedom from injury and adequate nutrition levels). We now discuss the major features of self-directedness—action modulation, anticipation, evaluation, and constructive gradient tracking—then show how they combine to form a complex integrated capacity for context-sensitive action.

3.1.1. Action modulation: generating the right kind of extended interaction sequences. As argued in Section 2, the general problem for adaptive systems is to produce environmental feedback that supports autonomy, and this requires coordinating internal to external conditions. The solution is to generate actions that form coherent whole sequences which yield the required outcomes. Appropriately shaping action to generate these coherent sequences is the system's management problem, and in Section 2 we contrasted low and high order strategies for solving the problem. As we noted, the high order strategy of a cheetah involves integrating many more parameters in producing action than does the low order strategy of a mosquito. However the difference does not simply lie in the number of parameters integrated, it also concerns the *type* of parameters involved. A low order strategy employs localised parameters, where the localisation is characteristically spatial, temporal, and functional (e.g. momentary spatially local variations in CO₂ concentration producing local flight settings). This localisation effectively means that the system only manages the interaction process with respect to temporally and spatially small scale, functionally limited aspects of the overall interaction process. To put it another way, the management horizon of a low order strategy only incorporates a

small slice of the overall interaction pattern. Higher order strategies, on the other hand, open up the management horizon to incorporate larger scale features of the interaction pattern.[6] The hunger-satiation signals of a cheetah, e.g., are temporally non-local in the sense that they measure variation over time of blood sugar levels and indicate roughly how soon another meal should be acquired. They are also functionally non-local in the sense that they measure a global system condition, degree of starvation, and affect a plethora of functions in an organised way. Thus, not only are more parameters integrated in action generation, but the parameters themselves are frequently non-local in important ways.

Self-directed systems generate the right kinds of interaction sequences through higher order process modulation (cf. Thelen 1995). In this way they effectively expand the management horizon to include more of the interaction process, thereby adopting a more holistic approach to modulating action. This is, of course, a matter of degree rather than a sharp division. Simple self-directed systems only expand the horizon to a limited degree. The bumblee example we discuss below provides an instance where there is a small, but nonetheless highly significant, expansion of the horizon of management as compared with mosquitos. Increases in self-directedness are marked by a progressive extension and enrichment of the management horizon, such that highly self-directed systems manage interaction with respect to *many* features of interaction from the local to the global, in time, space and functioning. The capacities for anticipation and evaluation which we now discuss are tools for achieving this type of management expansion.

3.1.2. Anticipation: how will/should the interaction go? In Section 2 we pointed out that increasing the degree of high order management a system employs increases the capacity for context-sensitive action, but that it also increases the *need* for high order management by opening up degrees of freedom in the interaction process (involving complexity in both the system and environment). This means that the interaction flows will typically display a high degree of temporal variability because variations in any of multiple factors may produce highly divergent interaction pathways, many of which will not be adaptive for the system. E.g., to pick out just one relevant factor, if the gazelle detects the cheetah too early it may escape and the cheetah will go hungry. Faced with this kind of problem adaptive systems must be selectively sensitive to temporal patterns in interaction. Specifically, they must *anticipate* the interaction flow by predictively modulating action so that action generation in a particular context is coordinated with reward outcomes in that context. Thus, increased temporal variability in interaction introduces the need for anticipative action management, which should combine predictive and normative expectancies about how the interaction flow will and should go.

Constructing such anticipation involves extracting one or more parameters from interaction and using the parameter values to modulate action. A simple example has been shown to occur in bee foraging. (We examine a model of the neurological basis of this behaviour by Montague et al. in Section 4.) Real (1991) has demonstrated in experiments with bumblebees that the bees can modify their foraging behaviour to selectively land on flowers whose colour reliably predicts

nectar reward. Placed in an environment of artificial blue and yellow flowers, where blue flowers contained a constant 2 microlitres of nectar, yellow flowers contained 6 microlitres and the remaining yellow flowers contained no nectar, 85% of the bumblebees visits were to blue flowers. In other words, the bees were able to anticipatively associate flower colour with reliable reward and use this to shape their behaviour. This capacity for anticipation marks a significant increase in sophistication over the capacity of mosquitos, whose actions are largely stereotypical reactions. Bumblebee foraging, in contrast, is context-sensitively shaped through predictive reward learning.

Anticipative process modulation is a widespread phenomenon, and is central to fluid intelligence. Further examples include the following: Catching a ball by observing its flight and predicting its future trajectory, and moving one's hand to intersect with this trajectory. A cheetah using visual information about prey alertness and available cover, combined with expectancies about the prey's speed and agility, to judge the appropriate point at which to initiate the chase. A detective (whom we shall call Sleuth for later reference) using evidence from a murder scene to construct a profile of the murderer which creates expectations about the murderer's behaviour, and in turn using the profile to direct the investigation.

3.1.3. Evaluation: how did the interaction go? Self-directedness also involves evaluation of the success of interaction. The norm matrices discussed in Section 2 provide the means to achieve this. An organism globally evaluates its interactions with respect to whether they result in it maintaining coherence within its norm matrix (putting it roughly, staying pain-free, well fed and happy). Within this general constraint, each of the norm signals a system possesses provides characteristic information about the relations between interaction and the system's autonomy closure conditions, such as the way that satiation reflects starvation avoidance (though, just as in the case of satiation, the reflection will generally be imperfect). Norm signals range from relatively low order (localised) evaluators to higher order (more holistic) evaluators. Proprioceptive signals such as stretch and pressure sensing used to modulate motor tasks like grasping provide examples of relatively low order success/error signals, while signals such as generalised discomfort and euphoria are high order signals indicating conditions of malfunction or success without being activity-specific (that is why they are high order). The norm matrix of a system thus provides it with an array of 'steering' information for remaining adaptive and, as we will discuss in Section 3.2 below, the organisational characteristics of the array are important for understanding the system's adaptive characteristics, including what types of learning processes it may be capable of.

Anticipation itself can form a very important aspect of evaluation because a system can evaluate whether its anticipations are correct, as measured by whether they yield successful action completion. This provides a means for the system to generate new information, including both predictive and normative information, about the conditions of successful interaction.[7]

3.1.4. Constructive gradient tacking: learning to improve performance. Anticipation and evaluation are each important for context-sensitively shaping action, and in combination over extended interaction sequences they facilitate constructive gradient tracking processes. A simple gradient tracking process, of the sort displayed by mosquitos, involves no modification to the tracking process, e.g. the mosquito flight orientation process itself never changes in response to variations in CO₂ concentration, only the flight direction. In contrast, a constructive gradient tracking process involves having changes induced in the system by the unfolding interaction that in turn change the way it subsequently interacts. Because of this constructive gradient tracking is a fundamental aspect of learning processes. When hunting, e.g., a cheetah develops anticipations about the hunting process, about the speed and agility of the prey and the like, anticipations which in turn modify the way the cheetah subsequently conducts the next moves in the hunt. In turn this modified behaviour is evaluated, generating a new set of anticipations that further modify the cheetah's hunting behaviour, and so on. This is a much more powerful type of gradient tracking process in which *performance capacity*, as well as current state, tracks the situation. And what is tracked is no longer a simple environmental gradient, like a CO₂ concentration gradient, but a combination of system interaction processes and environmental organisation, as evaluated by system norms. Thus, cheetahs track something like 'effective hunting', specified as a relationship between injuryfree movement effort and prey character (kind, size, health, etc.) terrain style ecological location, as evaluated against injury risk hunger urgency satiation potential ecological risk.[8] In this way constructive gradient tracking provides a means for a system to increase its self-directedness by enriching its anticipative and evaluative capacity.

3.1.5. The integrated self-directed agent. Anticipation, evaluation and constructive gradient tracking combine to allow a system to achieve fluid context-sensitive coordination of whole action sequences. The self-directed system is able to improve its context-sensitivity-its sensitivity to the environment and to its normative requirements—by replacing lower-order (stereotypical) management with higher-order integrative management, in which action is constantly re-shaped both by system norms and by information derived from interaction. This provides the basic form of learning capacity in constructive gradient tracking processes: improvements in performance capacity track organisation in the environment (such as ecological relations like prey behaviour). The mosquito can be said to learn its next flight orientation through environmental interaction, in an attenuated sense of that term, but the advent of constructive gradient tracking marks a distinctive increase in learning capacity. Moreover, in coordinating its autonomy constraints with the environment this kind of management generates an increasingly rich normative perspective, yielding an increasingly integrated sense of 'self' geared to serving the requirements of autonomy. This is the 'self' proper of self-directedness. The system that anticipatively steers itself through its environment to satisfy its own norms, learning to improve its performance as it goes, displays a distinctively intentional selfhood.[9]

Our general hypothesis is that intelligence, as a distinct form of adaptiveness, evolved as increasing complexity of organism and interaction necessitated increasingly sophisticated forms of self-directed high order integrative process management. Degrees of intelligence thus correspond with degrees of self-directedness. Mosquitos are directed but not self-directed systems: whilst they do engage in a complex dance as actions and environmental organisation combine to produce shaped interaction processes that achieve adaptive closure conditions, they do not learn to modify their interaction processes to improve performance. Cheetahs, on the other hand, are powerful learners in just this sense. Whether their interaction strategies work or not is of course not fully under their control, but it is much more so than it is for mosquitos. The detective Sleuth investigating a murder scene is more self-directed again, being able to modify the investigation process much more extensively (e.g. switching the whole crime profile from a business crime to a crime of passion, with attendant changes to investigative direction and methods), and on very finely discriminated information (e.g. on the basis of a distinctive cigarette butt).

3.2. Self-directed anticipative learning

As we have seen, learning plays an important role in self-directedness. Self-directed contextsensitive action is achieved by allowing performance to be learned. The learning problem for self-directed organisms is to translate open-ended high order viability requirements (e.g. obtain satiation) into specific sensori-motor problems (e.g. catch this gazelle) by constructing a set of effective interaction strategies (such as hunting techniques). Thus, constructivism is the second plank (with interactivism) of the I-C approach to modelling intelligence. In this respect, note that instructional prespecification is not an effective strategy for managing complex processes with many degrees of freedom because the potential process complexions rapidly overwhelm the instruction set. For this reason constructivism is a much more plausible approach to intelligence than nativism, and traditional artificial intelligence 'programming' models of cognition will be poor approximations to the kind of capacities that we are suggesting are central to intelligence.

We shall now outline an account of self-directed anticipative learning (SDAL) as an important constructivist mechanism involved in cognitive development. There is a spectrum of learning processes ranging in power from the virtually null case of momentary acquisition of information (e.g. local direction of maximum CO_2 concentration), to cases such as the bumblebee in which performance is modified by learning but learning capacity is not modified, through to cases in which the ability to learn is itself improved by the learning process, such as the way improvements in a cheetah's hunting technique allow it to learn more about a prey's behaviour, which in turn improves its capacity to modify its hunting. Likewise, but even more powerfully, as Sleuth acquires evidence in the murder investigation the profile of the murder and surrounding events improves, which in turn improves the ability to find and recognise new evidence. SDAL processes are of this last type: they involve a virtuous cycle of interactions in which progressive self-modification improves both performance and learning ability as a function of interaction. It is because of this

progressive character that we believe SDAL is a central feature of the development of high order cognition.

Earlier we noted that anticipation and evaluation can operate in tandem over extended interaction processes to produce learning in the form of constructive gradient tracking processes. The nature of the learning process depends a great deal on the organisation of the norm matrix and the way the system forms anticipations. SDAL requires significant asynchrony and a spread of low and high order norms. Asynchrony is required because some norms must be held relatively constant against other norms and interaction processes in order to serve as a directive scaffold shaping learning. Thus, hunger is an ongoing feature of a cheetah's learning process, and it is relative to this evaluative signal that it refines its hunting technique. Similarly, pain signals allow an animal to modify and refine its motor skills so as to avoid those kinds of actions which result in damage.[10]

Because high order norms leave open the details of which particular conditions best satisfy them, these conditions can be learned. Not only does this provide a powerfully permissive learning framework—any exploratory activity is acceptable that is efficacious in generating learning (see e.g. Section 4 below)—it also opens up the possibility of establishing refined lower-order norms served by learned signals to more specifically guide the achievement of their satisfaction. As a result of developing its prey-chase experience, a young cheetah may learn to creep close enough to prey (new norm) in order to chase safely and effectively. In this way the embedding autonomy-supporting closure conditions (e.g. damage control) that are the ultimate, if initially implicit, rationale for behaviour are increasingly enfolded into operational closure conditions (e.g. specific chasing styles) and made more explicitly accessible to organism self-direction through the learned signals which indicate them (e.g. visual perception of terrain combined with kineasthetic perception of muscular strain).

However, both low and high order normative signals have limitations in the kind of information they can provide. Low order signals may specifically indicate that a process has failed to achieve an operational closure condition, without indicating the underlying reasons for this failure, especially higher order organisational reasons. Sudden loss of pressure sensation may, e.g., indicate that the motor action used to grasp a glass was unsuccessful, but not indicate why. Conversely, a high order signal like generalised discomfort may indicate that a process has resulted in damage or distortion to the system (e.g. poor running technique, or poor diet), without indicating the specific features of the process that were at fault. In either case these signals can be sufficient to facilitate the retention or elimination of operational strategies, thereby permitting learning in the manner just described, however this learning process is in itself only weakly self-directed because a system furnished with it alone will have none but cognitively rudimentary means to improve its interaction processes (copying, blind trial and error). The young cheetah's hunger may provide incentive to learn improved hunting technique, but does not in itself indicate what needs improvement, for that trial and error and/or copying mother may be all that is available. Of course, these techniques, though cognitively unsophisticated, also remain our own basic techniques in the face of sufficient

ignorance, and unavoidably must do so, but we can do better by building on their past use.

SDAL processes gain increased power through the construction of anticipatory models of the interaction process. These anticipations generate new information by modifying interaction, and modification and enrichment of the anticipations results in improved ability to localise success and error, thereby improving learning capacity. For instance, as our young cheetah constructs its chase anticipative model through practice and perhaps copying mother, and becomes better able to anticipate the speed and movement of its prey during a chase and to evaluate that against its own capacity, it might learn that to be successful it must initiate the chase from a closer distance, or choose a constant swerving policy in the face of more erratic prey swerving, in order to give the prey less time to react and prolong the pursuit to the point where the cheetah becomes exhausted. In this case the anticipative model, combined with normative feedback (e.g. about exhaustion, as well as hunger) allows the cheetah to localise the defect(s) in its hunting technique. In turn, this refines the ability of the cheetah to form and evaluate anticipations about the hunting process (and, as noted earlier, to form new lower order norms).

This improvement may happen both within a particular problem solving episode and across episodes. Within a single investigation, e.g., Sleuth's investigation begins 'mechanically', following a set of general investigation techniques that are applied in all similar situations, but becomes increasingly self-directed as context information emerges in the form of clues and an articulated suspect profile. The changing suspect profile changes the search procedure, thereby changing the information generated by the investigation, and at the same time it changes Sleuth's expectations about what information is useful, and what investigation strategies are likely to be successful. In turn, the new information may change the suspect profile (e.g. when some norm is violated or the clue trail evaporates). Beyond the single investigation, practice in applying investigative techniques and opportunistically elaborating crime profiles leads to learning about higher order features of crime patterns and clue types. Over many investigations, Sleuth's general investigative, including profiling, techniques improve.[11]

Rich anticipative models and normative matrices thus play a critical role in making the learning process powerfully self-directed, because between them they provide the system an appropriate array of internally accessible signals which it may use to direct its own behaviour and internal processes. It is able to construct a correspondingly rich gradient to track which is sensitively correlated with its autonomy-relevant environmental context. When this capacity is effective, as when Sleuth's suspect profile changes the investigation in ways that lead to new clues, further improving the profile, the result is a virtuous self-modifying interaction cycle in which initial learning improves the system's learning ability, leading to a progressive increase in the system's anticipative depth. This is the essence of solving a divergent problem, where the correct problem definition, solution criteria and method all come clear at the same time as does the solution.

In this manner the I-C approach provides a natural and integrated approach to intelligence. As systems get more complex, and their interaction processes gain many degrees of freedom, they face a process management problem. They must integrate multiple normative constraints using divergent information to produce coherent interaction. The solution is increasingly high order process modulation capacity that generates increasingly holistic, self-directed process management. As these high order self-directed processes become increasingly sophisticated they incorporate anticipative learning processes of an increasingly powerful and cognitive nature. We believe that a theory of this type must form the central plank of an account of the emergence of high order cognition, in particular of human cognition. We shall now turn to examining some implications of this interactivist-constructivist approach to intelligence for modeling learning processes in the brain. Our ultimate aim is to use these theoretical considerations to develop testable empirical predications about at least some features of neural learning processes.

4. Modelling constructive learning in the brain

For brains to participate in the kinds of constructive learning processes characterised above they must possess certain features. Articulating the nature of these features provides one way of providing I-C with some empirical "bite": a way of evaluating the biological plausibility of the proposed learning models and of developing a concrete basis for comparison with competing theories in cognitive science. In this section we will take some initial steps towards this end. Our focus will be the Neural Constructivism (NC) paradigm recently put forward by Quartz and Sejnowski (1997). Quartz and Sejnowski's work is significant because it provides strong empirical evidence supporting a constructivist interpretation of cognitive development, and presents a neural learning mechanism that may play an important role in constructive learning. It thus serves as a powerful argument for cognitive constructivism. However it also possesses weaknesses which our analysis can illuminate: although the NC paradigm is constructivist it is not particularly interactivist, and partly as a result does not model the kinds of interaction processes or high order process modulation involved in self-directedness and SDAL. In consequence it currently fails to capture the construction of high order anticipation, and hence to capture a central aspect of cognitive development and intelligent capacity; but this evidently need not be so and we urge its further development.

One of the principle requirements of a constructivist brain is extensive activitydependent plasticity. Brains must be capable of rich self-organisation in response to (and partly *generating*) the changing interaction context. Quartz and Sejnowski have compiled extensive evidence for such neural activity-dependent plasticity and used it as the cornerstone for a 'neural constructivist' programmatic framework for cognitive neuroscience that emphasises the self-structuring aspects of neural organisation. The central theme of the NC paradigm is that learning guides brain development. They argue that the cognitive features of the brain are "built from the dynamic interaction between neural growth mechanisms and environmentally derived neural activity" (1997, p.537). Because cognitive activity and brain development function as mutual constraints on one another, the result is "nonstationary" learning—learning that changes the learning architecture, which in turn changes the learning process, thus changing learning, and so on. Because the learning architecture is not fixed, cognition cannot be considered an inherent manifestation of brain capacity; rather, Quartz and Sejnowski claim, the mind and brain essentially co-develop as cognition is *constructed* from the learning process. The principal developmental mechanism they point to as responsible for this constructivist learning process is activity-dependent neuronal growth.

The general model of NC learning presented by Quartz and Sejnowski is (roughly) as follows: The learning system starts as a simple network, i.e. one with relatively few connections (as well as inappropriate connection weights) in relation to what would be needed for the objectively accurate representation of input information. This latter is the functional task of the network and those network functions which achieve it are the network's target functions. The learning task is to somehow pass from the network's initial functions, which will be far from the target functions because of the network's initial condition, to the target functions. However improvement is tractable because the system can relatively easily find its best approximation to the target function without being swamped by an unmanageably large number of possible states, as would be the case if the network was initially large. In addition to conventional weight modification, the system has a learning algorithm that adds structure to the network according to some performance criteria. As the network grows its representational capacity increases, and hence its bias decreases, and this increase in representational capacity is guided by the problem domain itself. Essentially the environment[12] acts as a scaffold for network development. Moreover, because the increase in network complexity and organisation is progressive, and guided by increases in performance competency, the system can find its way to a relatively unbiased representation of its target function when that would be computationally intractable if the representational space was fixed.

In support of the NC paradigm as a model of human cognitive development, Quartz and Sejnowski review an extensive body of empirical evidence that brain development exhibits significant activity-dependent directedness. Here the evidence is briefly summarized. Contrary to maturational and selectionist models of brain development, empirical studies show that synaptic density, and axonal and dendritic arborization all increase during development. Quartz and Sejnowski argue that the most cognitively significant measure is dendritic arborization. Dendrites are the primary receptive surface of the neuron, and the integration of synaptic activity depends on the geometry of the dendritic branches. In addition, dendritic change has localised effects, which means that change can be accumulated progressively—an important property for cumulative learning. Quartz and Sejnowski highlight the following developmental characteristics of dendritic growth as evidence for the NC model:

- dendritic length increases dramatically during development, and its greatest period of growth corresponds to the periods of intense cognitive development;
- dendritic structure is heavily activity-dependent;

- it is the degree of correlation in the afferent activity rather than simply the presence of activity that underlies dendritic organization. Dendritic segments detect correlated activity and grow preferentially in such regions;
- hebbian learning can occur in local dendrite areas;
- local stimulation can induce dendritic branching.

Quartz and Sejnowski claim that one of the primary mechanisms responsible for this pattern of dendritic activity is likely to be a Hebbian volume learning process based on nitric oxide (NO) diffusion. NO is a membrane permeable gas whose synthesis is induced by NMDA receptor activation-since NMDA receptors are Hebbian filters (they detect pre-post coincidence), NO production obeys Hebbian rules. NO rapidly diffuses into surrounding tissue, and is known to act as a retrograde signal in the induction of long-term potentiation; NO also affects local blood supply. Quartz and Sejnowski speculate that, in addition, NO concentration acts to regulate dendritic branching. This would allow the probability of branching or retraction to be proportional to the activity of nearby synapses over time, effectively translating the associative conditions for synaptic weight change to those responsible for connection modification. The result is that neurons can sample their local region for correlated afferent activity, using this activity to direct connection modification (pp.549-550). Thus, the development of neuronal organisation is the result of feedback between occurrent synaptic activity and neuronal plasticity mechanisms such that correlated activity is amplified, not only through Hebbian synaptic mechanisms, but also through Hebbian connection modification mechanisms which induce new neural connections from correlated upstream sources.

The picture of brain development which Quartz and Sejnowski claim follows from these considerations is this:

[T]he human brain's development is a prolonged period in which environmental structure shapes the brain activity that in turn builds the circuits underlying thought. In place of pre-wired modules, patterned activity builds up increasingly complex circuits, with areas staging their development. Cortical areas further away from the sensory periphery wait in anticipation of increasingly complex patterns of activity resulting from development in lower areas. As this development proceeds, areas of the brain become increasingly specialized for particular functions, reflecting a cascade of environmental shaping. Some brain circuits close to the sensory periphery, such as in our early visual system, are in place by six months of age; but those in language areas, further away from the sensory periphery, do not begin to complete their development until the eighth year of life.[13] (p.550)

This represents a shift away from the tendency of conventional cognitive science towards modelling learning processes in terms of self-contained algorithmic solutions, of which the assumption of a fixed learning architecture is just one manifestation and, insofar, it resonates with the I-C perspective. The empirical evidence for nonstationarity in brain development is strong, and does indeed open a new and

exciting window onto the relationship between mind and brain—a window which is strongly constructivist in perspective.

From the standpoint of the I-C framework for modelling intelligence, however, NC has flaws which ultimately stem from the retention of classical cognitive science modelling assumptions. The most significant problematic assumptions are as follows:

- 1. A focus on internal brain processes which suppresses the role of extra-neural interaction as an extended process in learning. The learning problem is idealised as matching a pattern of input activity, and learning is treated atomistically as a collection of separate, abstracted processes of conformation to the pattern. The model ignores the fact that real interaction is a temporally and spatially extended process, and that much incoming neuronal activity will be feedback and feedforward from prior neural activity via the body and environment.
- 2. *No endogenous evaluative constraints, no self-directedness.* All information directing neuronal development is externally sourced. Constructive learning is conceived as environmentally guided neural circuit building. There is no self-directedness, no anticipative action modulation.

These assumptions are interlinked, and complemented by a tendency to regard neural development as a process of pattern-fitting to environmentally derived information, which neural structure comes to 'represent', with representation becoming a blanket term for cognitive significance. It is not the notion of representation per se that is at issue, but the model chosen for it, namely one focused on environmentally directed implantation (a literal 'in-forming') that excludes the role of organism directed interaction and construction processes.[14] By contrast, the I-C approach highlights the multiple errors created by three interlocking biases: suppressing the role of interaction in learning, idealising away the temporality of the learning problem, and ignoring the active role of neurons in generating feedback which directs the learning process. For instance, it is only when interaction based feedback is included in the picture that the significance of normative signals becomes apparent (Section 2 and 3.2). Such signals are the basis of an organism's ability to evaluate its behaviour, and therefore play a critical role in producing learning which is self-directed.

Indeed, there is an absence of endogenous evaluative constraints on learning in the NC paradigm. The only significant constraint on neuronal development that is picked out by the NC model is correlated afferent activity, which is assumed to be derived from environmental organisation. Apart from the staged layer architecture, internal constraints are essentially local. There is no obvious reason to expect that such a learning process will be self-directed or recognisably cognitive. In Section 3 it was suggested that dynamical asynchrony is a critical factor in the capacity for self-organisational richness which underwrites learning capacity (note 11 and text), and that evaluative signals are probably an important source of such asynchrony in brains. Essentially such signals provide an additional scaffolding structure for neuronal self-organisation—a structure which classifies conjoint environment/behaviour contexts. This scaffolding makes the learning process self-directed, because the

system has an internal measure for appropriate activity. Learning is in effect jointly scaffolded by the organisation of environmental interaction and endogenous evaluative measures of the success of the interaction.

The role of such constraints in learning gives us additional reason to be interested in embodiment. Not only will the system's physical characteristics be an important structuring factor in interaction, but bodily factors such as damage, starvation, etc. will provide a major source of evaluative constraints which direct learning. In reconciling the modulation of interaction with its endogenous normative constraints, the system needs to integrate diverse sources of information, an inherently higher ordered problem than mere passive pattern-conformation, requiring construction of high order regulators, e.g. through use of anticipative models. Thus, for the reasons provided in Sections 2 and 3, adding these additional constraints to the learning process does give us a reason to expect that learning will become increasingly self-directed, and recognisably cognitive.

The SDAL model of learning provides a process account of how progressive increases in high order anticipative directedness can occur. It highlights the fact that increasing directed interaction competency actually improves the system's ability to discriminate environmental information, providing a basis for further improvement in learning. No such process is suggested by the NC model, which implicitly assumes that the complexity of input information lies outside the shaping capacity of the system, and indeed the NC model directs attention away from the possibility of such a process by ignoring output. The interaction of mind and brain in cognitive development, highlighted by NC, is only part of the constructivist story—mind, brain, body, and environment are intimately intertwined. An adequate cognitive neuroscience must include all of these factors in its theoretical framework.

In this connection the neural mechanism central to the bee foraging case described in Section 3 provides an illuminating extension of NC. Montague et al. (1995) develop a model of the bee learning process found by Real 1991. The neural architecture used in the model is based on a neuron with widespread projections within the bee brain whose activity carries information concerning the



Fig. 1. (Based on figure in Montague et al. 1995)

reward value of nectar. Bee foraging is modeled using visually guided flight in a simulated three dimensional environment. The model is designed to show how neuromodulatory effects can bias actions and regulate synaptic plasticity to generate a form of predictive Hebbian learning. In the model the simulated bee, which possessed a cone shaped field of view, moved around a three dimensional arena whose floor consisted of blue and yellow squares. The architecture of the model is as follows (see Fig. 1).

P is a linear unit that receives sensory information from **B**, **N**, and **Y** concerning the percentage of blue, yellow and neutral input in the visual field, weighted by W_B, W_N, and W_Y, and reward information from R. P responds transiently to time-averaged changes in input activity. Its output represents an ongoing comparison of net previous activity and the sum of current reward and sensory activity. **P** serves to label changes in sensory input as 'better than expected' when change in output is positive (d(t)>0), or 'worse than expected' when change in output is negative $(d(t) \le 0)$. **P** output models changes in neuromodulator release by modifying weights W_B and W_Y . (In the models discussed here W_N was held constant.) As the model bee moved above the colour field changes in the activity of its sensory neurons occurred. When there was no reward signal, **P** output biased actions by determining whether the bee continued on its present direction or tumbled randomly. When the bee collided with a coloured square it received a reward according to the volume of nectar, generating activity in r(t). This caused an adjustment to the weights W_B and W_Y. Thus, **P** output continuously guided actions but regulated learning only during the receipt of reward. With respect to our discussion of self-directedness, P output is a higher order parameter with which the model bee modulates subsequent interactive activity and it provides the model bee with an elementary self-directing capacity. With this architecture the model bee provided a reasonable simulation of real bee foraging, learning to preferentially visit the flowers whose colour was a reliable predictor of nectar reward (73%-85%, as compared with 85% for real bees).

As Montague et al. put it, the model shows, "how the learned weights can be used to choose appropriate actions and how the resulting action choices influence the learning. In addition, the actions taken by the model are also influenced by the structure of the simulated environment in which it moved." (1995, p.728) The noteworthy feature of this model from our perspective is that it incorporates interaction, evaluation and anticipation into the learning process, making it a prime candidate to serve as a model of the type of neural architecture that subserves self-directedness.

This hypothesis is strengthened by research showing that the mesencephalic dopamine system in mammals performs an analogous functional role. Dopaminergic subcortical nuclei in the midbrain and basal forebrain project diffuse ascending axons which innervate large regions of the cortical mantle, and play a role in many aspects of development, learning and behavioural control, including the regulation of attentional and motivational states. Montague et al. (1996) shows how this system could serve to provide information about reward to cortical regions in the same way as **P** does in the bee model (see also Montague and Sejnowski 1994). In their interpretation activity in the cortex anticipates the future receipt of reward. Fluctuations in the activity levels of neurons in the dopaminergic system provide a measure of error in these anticipations, and this error signal delivers neuromodulators to cortical and subcortical regions that influence synaptic plasticity and thereby act to modify the anticipations of those areas. Montague et al. are able use the theory to make empirically supported predictions about human decision making behaviour in simple tasks.

There is no reason why the activity-dependent connection growth mechanisms Quartz and Sejnowski base the NC paradigm on cannot be complemented with the type of anticipative reward mechanism involved in the bee and mammalian dopaminergic systems. Indeed, Sejnowski is involved in all of the research we have discussed. Our argument is that *only* when interaction, evaluation and anticipation are conjointly incorporated into the learning picture can distinctively cognitive features, such as self-directedness and the constructive learning processes involved in SDAL, be characterised. The omission of these factors is a significant flaw in the current formulation of the NC paradigm. If these additional features are added, we believe, a highly promising neuronally based constructivist account of cognitive learning is possible.

Combining the Hebbian connection growth mechanism with the anticipative-reward learning mechanism provides an intriguing insight into some of the neural characteristics required for powerful SDAL processes of the kind humans are capable of, and how these capacities may grade up from very simple forms of self-directedness in which learning capacity is limited. Simpler forms of self-directedness involve anticipative reward modulation of activity, and this can be achieved with the type of simple modulatory architecture found in bees. Stronger forms of SDAL require extensive plasticity and the capacity to progressively acquire pattern recognition and motor skill capacity. A large cortex featuring activity-dependent connection growth may provide the basis for this. Continuous increases in self-directedness, generating increasingly powerful constructive learning capacities, could therefore be achieved by continuously adding cortical capacity to the basic sub-cortically based anticipative reward architecture.

5. Philosophical reflections on models and methods in cognitive science

In this section we conclude our account with some reflections concerning the deeper implications of adopting a thoroughgoing I-C approach in cognitive science and philosophy. In particular, we reflect briefly on the fundamental ways a biologically grounded I-C approach to modelling and method diverges from that of the currently dominant CIP paradigm in respect of four interrelated groups of modelling assumptions. These were summarised early in Section 2 and have been commented on in various places as the I-C model was developed; in keeping with the I-C conception of making performance norms more explicit, here we briefly focus on them more explicitly.

5.1. The modelling problem: holism versus modularity

Recall that the basic adaptive problem all creatures must solve is how to coordinate their internal autonomy needs to their external environment, in ways that lie within

their capacities. This problem presents itself to creatures episodically as a requirement for organised whole activities (e.g. hunting to the kill). In response to complex variable versions of the problem (hunt *this* gazelle *here*, etc.) some creatures self-organisingly construct higher order regulation to manage these interaction flows as shaped wholes, e.g. by extracting one or a few parameters which regulate that shape (cf. learning to crawl). Moreover, the appropriate kinds of higher order management endow creatures with forms of self-direction which in turn provide the organisational foundation for intelligence. In this conception, we noted, an intelligent system is thrice globally holistic in character: (a) the underlying autonomy constraint is a whole system constraint, (b) the required actions are whole integrated flow sequences of autonomy significance, and (c) the required coordinating management organisation is a holistic feature of whole-system (embodied mind)-environment interaction. As against this methodological and substantive holism or non-modularity, the standard approaches to cognitive science attempt to simplify the situation by eliminating aspects and/or components of the models.

The problems begin at the broadest level where standard conceptions of adaptive modelling suppress process relationships that are important to a systematic understanding of intelligence. Selection theoretic models of adaptiveness, e.g., while useful for modeling the outcomes of populational adaptation processes, abstract away from the dynamical interaction and developmental processes of the individual organism that underlie those outcomes and hence are not very useful for capturing the dynamical embodiment factors involved in intelligence (see Christensen and Hooker 1989c for discussion). A selectionist explanation of even mosquito gradient tracking behaviour (it was reproductively advantageous) already fails to shed any light on mosquito interaction dynamics and internal process organisation; it can do no better for cheetahs, rather the explanatory lacunae loom still larger. Likewise, CIP models of cognition characteristically focus on hypothesising a special subset of internal formal problem solving processes and treat the detailed dynamical features of embodiment and interaction as outside the purview of the model. The greater sophistication of the internal modulatory processes of cheetahs deservedly command more attention than that of mosquitos, but this does not make them better candidates for internalist cognitive modelling. Cheetah interaction dynamics and internal action-generating processes are intimately connected in a single dynamical system and cannot sensibly be understood separately. In short, the standard adaptive models now available are deficient, and crucially so where it comes to understanding the roots of intelligence. Selectionist models treat the system interior and most of the interactive dynamics as a black box, internalist AI-style models treat everything else but the 'mental' part of the system interior as a black box; neither is alone satisfactory, nor in combination do they remove each other's deficits. In contrast with these approaches to adaptiveness, I-C incorporates dynamical interaction and developmental processes within an autonomy-based general theory of adaptive systems, and shows how modelling these factors directly provides a natural framework for a theory of intelligence.

In this context, classical internalism, according to which cognition is characterisable as a distinctive form of process 'inside the head', is worth further comment.

In classical artificial intelligence (AI) internalism takes the form of an obvious and extreme modularism. Cognitive processes are characterised as symbol string inputs ordered under logical syntax and operated on by logical transformation functions to produce symbol string outputs. Intelligence is treated as confined to the middle of a triad of modules: (I) sensory reception and symbolic encryption as planner input, (T) intelligent central planner as logical input-output transformer, (O) motor encryption of planner output and motor effector as behavioural output. This divides up the overall adaptive capacities of an intelligent system into a group of functionally distinct capacities corresponding respectively to interaction capacity produced by the body of the system and its sensors and motor effectors, and control capacity produced by a central planning module (CPM, = T). Then the capacities of the CPM are further modularised into the capacities to execute formal algorithmic solutions to self-contained formal problems. Connectionism is, or can be, far less modular, especially in respect to the latter CPM modularity, however standard connectionism retains the general internalist assumption and its I-T-O expression.[15]

I-T-O modularity encapsulates a set of powerful modelling assumptions for cognitive sciences. The very distinctiveness of the mental operations separates mind from bodily dynamics, however materialist its proponents may consider themselves; all evolution carries on dynamically until the distinctive mental states/entities appear, whence the distinctive mental processes arise and for the first time intelligence becomes possible. Psychology becomes an independent discipline sui generis. (Methodological solipsism is a strong version of internalism.) Of course, cognitive processes must have some important relation to the 'outside', but this is finessed as representation, where the relations to the 'outside' are collapsed into some kind of referring relation and the focus is on the internally characterised representational contents. The referring relation 'skims over' all the detail of process organisation and interactive dynamics to magically directly connect internal mental entities to the 'outside' world. Once the general character of this relation is specified the details are suppressed and attention focused on the 'inner workings'. By contrast, under an I-C approach intelligence is conceived in terms of dynamical organisational processes that grade up by degrees from very elementary dynamical capacities, such as the mosquito's carbon dioxide gradient tracking, to more properly intelligent ones.

Needless to say, the I-C approach is also radically at odds with the modularising of the internal structure of T (the CPM of the AI version). These 'micro modular' modelling assumptions in cognitive science are closely related to the general modelling strategies of abstraction and functional decomposition. In attempting to model a complex system it assists tractability to abstract a general pattern of organisation thought to correspond to the 'essential' features of the system. This general pattern may ignore many aspects of the system deemed to be peripheral. Conversely, in attempting to understand global properties of the system it can be helpful to functionally decompose the system both by functional order (where complex functional capacities are decomposed into simpler, more basic sub-functions) and by task (where more global functional organisation is decomposable into structurally localised capacities of specific material components of the system).[16] Abstraction and functional decomposition are indispensable tools for effective research in biology and

cognitive science because they make complexity more manageable (see Bechtel and Richardson 1993, Wimsatt 1987). Nevertheless, they should be recognised *as* methodological strategies which must be sensitive to the actual empirical details.

In particular, it is reasonable to expect that non modular process organisation will often occur in evolved organisms. Natural biological and cognitive systems have been constructed from the ground up, and as a result they display complex process interdependencies and functional multiplexing that can be highly counterintuitive, particularly from a perspective conditioned by human top-down engineering practices.[17] In these circumstances systems may not (and in many important cases won't) possess the kind of general functional organisation that an analytical abstraction is inclined to posit. Likewise, they may not possess the kind of modular organisation which functional decomposition presupposes as its first approximation, and which we humans are still inclined to build structurally into our engineering designs to preserve analytic simplicity and division of labour. Thus, any particular modularity assumptions must be subjected to careful scrutiny, and abandoned if it is found that the system process organisation does not in fact respect the hypothesised partitioning. In particular, whether or not it proves empirically supported the I-C organisational analysis of intelligent capacities is a demonstration that formal decompositions, no matter how compelling they may seem within a formal context, cannot thereby be assumed to apply modelling real world systems, empirical support for their underlying structural assumptions must be demonstrated. It is a mistake to buttress one's preferred modularities as metaphysical principles in the manner of Fodor 1992.

Our point is not that modularity doesn't occur; it clearly does (e.g. cells, organs, organisms, etc.), and is just as clearly highly important. Rather, the point is that much of what is found is what may be termed 'soft' modularity. Soft modularity involves partial partitioning of structure and function combined with complex process interdependencies. Such modules as occur are in fact generated and sustained by processes; interactions across very complex pathways and at multiple timescales are often involved in these processes.[18] Simplifications made for modelling tractability must be recognised as partial approximations that can omit or distort relationships which are important in a broader perspective. The interconnections across diverse structural and functional aspects of the system may be just as important as more localised or functionally specialised features. Moreover, because of this interconnectedness a too narrow focus on very specific empirical data can easily lose the more global process relations which are just as important. This means that there is an important role for qualitative modelling at the meso and macro level that attempts to capture these relations, and that qualitative modelling and more concrete modelling and empirical investigation strategies should be pursued in close communication with one another.

This type of highly interconnected 'soft modular' organisation is precisely what we have argued occurs in intelligence. Interaction dynamics and the internal processes that generate and evaluate action are intimately interwoven, such that illuminating models must capture all of these factors simultaneously. Our account of self-directedness is an attempt to develop an integrated model of intelligence which captures at least the broad shape of the relations. The need for integrative models only becomes more pressing as cognition becomes more sophisticated; learning

processes must actively integrate diverse constraints to produce coherent efficacious performance. SDAL utilises interaction to generate anticipations, and uses the information generated by the way interaction is modified by these anticipations to further modify learning. It is a thoroughly integrated process, and its power derives from this.

5.2. The modelling problem: 'frames' versus information

The managerial modularity of CIP leads naturally to a modular formulation of cognitive function: it is assumed that action can always be analysed into a collection of *preexisting well*defined problems posed within the T module to which the system individually computes optimal, typically algorithmic, solutions, which are provided as additional information to the T module (to transduce into motor output). It is in the nature of optimal algorithmic solutions that they are complete and so self-contained within their particular problem definition or space. A 'frame' for each such problem module, or class of closely related problem modules, i.e. a set of constraints which together provide the broad structure of possibilities in the situation, is thus presupposed. (Cf. e.g. the Task and Problem spaces of Newell 1980b.) The assumption is applied universally; even the bumblebee model discussed earlier introduces a bee problem frame, that of the so-called two-arm bandit problem.[19] The formal optimality analysis then utilises the constraints the frame identifies. Notoriously, this way of modelling intelligent systems leads to fundamental problems, perhaps the most important being the 'frame problem': how can systems feasibly manage the storage and context-sensitive choice of problem frames themselves? Because of the variety, complexity, and metamorphosis of frames, the frame problem has proven notoriously intractable within CIP.

Indeed, the CIP conception of the basic problem/optimisation module obscures the following facts. First, the adaptive problems faced by a system are frequently poorly defined. Adaptive systems must often transform vague problems into more specific ones, as do cheetahs faced with hunger when they learn how to acquire food by developing an effective hunting technique. To do so they must uncover a great deal of implicit information about adaptive interaction and enfold it into explicit anticipative process modulation. A focus on algorithms prevents an understanding of how it is possible to learn things that haven't already been clearly conceptualised, but doing this is central to solving life problems. How is conceptual progress possible? This is a basic cognitive problem, for which CIP approaches have no easy answer. From our perspective the answer must be that it is only possible by doing more that is relevant than can be initially explicitly known, but having processes that will render it explicit in a relevant way. Understanding how the resolution of initially vague problems is possible is central to understanding how intelligence is organised and how it is linked to creativity.

Second, every adaptive process is embedded in a complex matrix of norms, making often conflicting demands, and hence the problem for the organism is something like maintaining sufficient coherency within the matrix, over sufficient time. Some degree of dehydration is permissible while in opportunistic pursuit of food, but before too long it must be given dominance until reduced; injury can be briefly risked where a kill or protection of cubs is at issue, but only cautiously, or in extremity; and so on. This means that the 'problem' may not be locally well-defined in a way that sensibly allows treatment as an optimisation. Success may be a matter of continually balancing a complex web of tensions. (In terms of traditional rationality theory, this can roughly be thought of as a multi-dimensional, dynamic satisficing process, but one which cannot in principle be transformed into an optimisation process under additional constraints.) Moreover the 'problem' itself is typically nonstationary for constructivist learners: they construct further norms from their initial norms and anticipative modelling and elaborate anticipative modelling from norms and experience, and these elaborations alter, respectively, the criteria for success and the operative constraints.)

Third, the basic form of gradient tracking is that of an iterative process in real time in which the actual course of action is not fixed in advance but depends on the feedback received as interaction proceeds. This means that, while the anticipative goal of the action sequence is some autonomy-relevant outcome, the iterative, feedback-dependent nature of the process renders that achievement in principle inaccessible to a global optimisation formulation.[20]

For all these reasons algorithmic optimisation makes little sense as a focus for modelling intelligence. Rather, it is the constructive transformation of the problem itself that involves the real learning, the place where real intelligence is displayed. Following rules can be 'mechanised' (a computer can do it), creating appropriate rules requires intelligence. In particular, finding an optimal solution with well worked out rules to a well worked out problem is merely a technical prowess; but when faced with an open-ended, under-specified, multi-tensioned, issue of how to organise oneself on many timescales across sufficient time and in relation to a dynamic environment, transforming that issue into a nested set of complex integrated routines that maintain viability is quintessentially an exercise in intelligence.

There is thus a natural link between interactivism and constructivism. Constructivism claims that cognitive capacity progressively emerges as the system uses mutual shaping between interaction and internal processes to generate increasing differentiation and adaptive competence in its interactive capacity.[21] The key intuition is that the mature form taken by cognitive processes is deeply shaped by the interaction of system and environment. In contrast, nativist approaches assume that the primary basis of cognitive ability is an internally specified cognitive architecture, for which development is merely a maturation process. The evidence for extensive activity-dependent plasticity in the brain strengthens the case for constructivism, but also reinforces the importance of interaction, as our criticisms of Quartz and Sejnowski's NC paradigm have suggested. Furthermore the complexity of the adaptive problems intelligent systems must solve also strengthens the case for constructivism. Transforming vague problems into more specific, solvable ones requires the use of high-order modulation coupled with environment-sensitive construction of more specific interaction processes (Section 2).

To sum up. On the one hand, the presuppositions within which the frame problem arises are mis-posed. As the discussion of optimisation shows, to suppose that living successfully can be analysed in terms of optimisable solutions to selfcontained, well defined problems is itself to mis-frame the task facing real intelligent systems. And that in turn mis-directs attention in looking for the underlying organised processes that might sustain viable behaviour and learning. On the other hand, the kind of learning intelligent creatures actually and most appropriately engage in when facing their life tasks is of a sort that has an intrinsically context-sensitive structure to it which is much more suited to providing smooth regulation of transitions among situations than is frame transition. Higher order regulation to manage the coordinated flows of interaction as shaped wholes, e.g. by extracting one or a few parameters which regulate that shape, also specifies possibilities/constraints for the situation and in this sense replaces frames; but it provides the capacity to modulate interaction coherently across multiple processes and timescales through suitable interlocking parameter regulation.

5.3. The modelling problem: modelling semantics

Underlying and supporting the modular problem formulation that gives rise to the frame problem is a complementary pair of crippling semantic assumptions. These are worth exposing to critical assessment since they are widely, if often tacitly, adopted, and doing so will also serve to highlight the different character of our own approach to semantics.

The first assumption is implicit in the presentation of the bumblebee modelling of Montague etal. 1995. On the one hand, it is clear from their modelled constitution that all that the bees learn, and can learn, is to extract probabilistic correlations between flower colour and nectar presence. On the other hand, bees are taken to be solving a formal decision-theoretic two-arm bandit problem, a problem posed and solved in the already-meaningful symbolism of costs and benefits to an agent. How is the slide from the one formulation to the other made? Well, it is assumed that the bees have a problem to solve and that the solution will be arrived at computationally, whence the use of the formal decision-theoretic model. But this last moves slides across a crucial ambiguity in the notion of computational modelling. A computation is an allowed syntax-respecting transform in some syntactically specified formal system or 'language'. So it is important to specify the syntax within which some state change is a computation. It is distinctive of CIP that its computations are in the syntax within which both the problem and the action solution (both input and output) are expressed as *cognitive* issues for the *agent* in question. That is why its internal algorithms are so easily read as symbol manipulations acting on symbolic representations. However, there is an ambiguity between this sense of computation and typical models in the scientific literature which describe themselves as computational, intending instead only the claim that the models are quantitative and hence can be modelled on a digital computer. This is computation in numerical syntax and is not at issue here; indeed, clearly both CIP and many non-cognitivist models, including van Gelder's dynamical models, can be captured in numerical computational models.[22] The shift from correlation extraction to solving the two-arm bandit problem is then tantamount to the shift from the latter to the former sense of computation, and this is then taken to license the use of problem-defining

terms that are already semantically interpreted in terms of agent goals, costs and benefits. By thus borrowing from an assumed well understood semantics the issue of modelling semantics is finessed, suppressing all questions concerning the basis in interactive process organisation for attributing agency conceptions.

This approach leaves open still the semantics of those internal states which model environmental interaction conditions and this is typically settled by the complementary adoption of a representational, referential semantics for them (second assumption). In this conception the semantic significance of a cognitive system's interaction states is what they refer to, in the linguistic sense of reference, and this is to be identified with the environmental situation which gave rise to them (see note 15). Here the reference relation leaps over all the idiosyncratic interactive complexity through which a creature relates to its environment to 'directly' connect inner state with outer condition. Once again, by thus borrowing from an assumed well understood semantics the issue of modelling semantics is finessed, this time suppressing all questions concerning the interactive basis for attributing representational conceptions.

Taken together, these assumptions make semantic modelling 'very easy'. One has only to find some causal correlations between environmental features and creature behaviour (e.g. flower colour and bee feeding behaviour) to postulate internal representations and then construct an abstract problem model in these terms (e.g. the two-arm bandit decision problem) which is empirically predictive for the representing behaviours, to then be able to read off the significance of inner processes and states from the model's agency terms (e.g. that bees are security oriented in their nectar gathering decisions). Moreover this procedure is very general, the correlation-based reference relation can bridge between the world and the internal condition of both the most complex and the simplest creatures, and similarly for the use of abstract problem models.

Unfortunately, adopted thus simplistically, this is altogether too powerful a modelling procedure - it provides a cognitive-semantic 'just so' story for any nearly system whatever. Even Braitenberg's simple photo-taxic vehicle can be modelled as solving the 'path orientation decision problem', making decisions in terms of representations of its current orientation to a light source and its goal of arriving at the source.[23] But here all the directionality resides in the spatial arrangement of the body of the vehicle in relation to the inverse-square radial organisation of the electromagnetic field emanating from the light source. (Between them they literally frame the problem, as all body-environments do.) There are no representations or decisions, and so nothing to interpret semantically. More generally, just as the selectionist evolutionary 'just so' stories that form its biological counterpart cannot be transformed into powerful explanatory insight (cf. the modularity discussion above) so too do these semantic strategies in themselves similarly fail to provide explanatory insight - as the Breitenberg vehicle example illustrates. Indeed, direct reference in the required sense is dynamically inaccessible to creatures (note 15) and the attempt to characterise it causally has proven a notorious quagmire, while well-defined, optimally soluble problems are also largely similarly inaccessible to creatures (cf. preceding discussion).[24]

In the I-C approach, by contrast, semantics is a more complex multidimensional, context sensitive, phenomenon. Norms are determined by the autonomy conditions of the system and semantic content is determined by norm-referenced differentiated modulation of downstream effect (note 15). An electromagnetic radiation field has no significance for a Breitenberg vehicle, though it moves systematically with respect to it (above), a CO₂ stream has (roughly) the significance of 'now is a time to orient and fly' because the mosquito's directive organisation includes at least an implicit flight orientation norm as reference (Section 2). On the other hand, because of its greater organisational complexity, the sight of a gazelle has (roughly) the significance of 'now is a good time to determine hunting priority in relation to the likely risks and rewards of this opportunity and the state of various other activity requirements competing with hunger for attention'. Thus semantics must be modelled in a way that is sensitive to the organisation of the system, and cannot be generically ascribed. A semantic analysis provides, not an easy 'just so' story, but must be constructed from a detailed account of the dynamical organisation of the system; it cannot be provided until that account is available and has all of the empirical content of that account.

This is not to say that CIP modelling is never useful. To the contrary, aptly constructed CIP models may prove informative at least where external behavioural modelling is a useful first step (e.g. the bee foraging model above) and of course wherever there is reason to attribute real internal symbolic computation.[25] The point is that particular computational models and their semantic interpretation must be constructed in ways that are sensitive to the interaction dynamics and normative constraints of embodied adaptive systems. Ignoring the isolation of process from context involved in CIP modelling, and the approximations involved in its abstractions, is what causes difficulty.

5.4. The modelling problem: dynamics and meso modelling

While the preceding discussion demonstrates the importance of a dynamically grounded holistic approach within an adaptive framework, dynamical modelling in itself will not capture the process organisational considerations I-C introduces. Dynamical embodiment research has tended to focus exclusively on the study of emergent dynamical patterns, their critical bifurcation points and control parameters, and the like using as the fundamental framework the dynamical modelling of differential equations (d.e.'s) as fields on differential manifolds, e.g. on phase space.[26] This is a very important new tool in the armoury of cognitive science and its significance has been rightly emphasised. However, this type of model also possesses limitations which mean that it can only be a part, not the whole, of the conceptual framework for cognitive science. In particular, it does not explicitly describe the physical organisation of the system—a chemical clock and a pendulum, for instance, may be modeled as equivalent dynamical oscillators. Only the global dynamical outcome is specified, not the organised processes which produced it and which, according to the I-C account, are crucial for understanding its cognitive nature and significance.[27]

In particular, we believe that our analysis makes clear why it is that the challenge of introducing high order, self-steering modulatory capacities must eventually be faced. The recent work of Brooks, a passionate advocate of the dynamical embodied modelling approach (see Brooks 1991), perhaps provides a first hint of this. Brooks' early work focused on the development of simple robots called Creatures designed to mimic insect-level performance in much the way illustrated in the mosquito example. However, in a dramatic shift of research focus Brooks' switched from attempting to model insect-level intelligence with Creatures to modelling human intelligence by building a humanoid robot capable of interacting in a humanlike fashion. This new research was called the Cog project (see Brooks and Stein 1993). A humanoid robot, though, possesses far greater sensori-motor complexity than a Creature, and this increase in complexity raises a new set of design issues. Creatures are designed with simple sensors, limited behaviour repertoires, and few goals. Because their modes of adaptiveness are so simple it is safe to allow their behaviour to be largely situation-driven. A humanoid robot, on the other hand, has much more complex sensory information and many more possibilities for action, so it cannot avoid the issue of somehow selecting actions appropriate to its situation. For example:

Suppose the humanoid robot is trying to carry out some manipulation task and is foveating on its hand and the object with which it is interacting. But, then suppose some object moves in its peripheral vision. Should it saccade to the motion to determine what it is? Under some circumstances this would be the appropriate behavior, for instance when the humanoid is just fooling around and is not highly motivated by the task at hand. But when it is engaged in active play with a person, and there is a lot of background activity going on in the room, this would be entirely inappropriate. If it kept saccading to everything moving in the room it would not be able to engage the person sufficiently, who no doubt would find the robot's behavior distracting and even annoying. (Brooks 1997, p.298)

The robot must be able to orchestrate its many low-level processes to produce coherent high level context-sensitive behaviour. In order to achieve this Brooks argues that the robot must have some form of motivation which provides it with preferences over courses of action. In effect, motivation here can be understood as a form of high order process modulation. For us the implication is clear, a robot like Cog must be self-directed if it is to function effectively.[28] However to adequately capture this phenomenon we need to model the processes at this intermediate (meso) scale. There is, then, no way to avoid all three of (relatively) micro, meso and macro modelling if intelligence is to be adequately captured. Micro models are essential to specify the basic dynamical interactions which set the broadest constraints on what is possible and thence to provide the basic dynamical modelling framework for investigating collective phenomena. Macro models are essential to specify the global features of cognitive significance, and capture those emergent dynamical features which have no relevant finer analysis, and to relate them both to the bulk of the salient empirical data. However, complex nonlinear systems frequently have process structure at many scales, and hence meso models are required to capture the process structure

intermediate between the micro and macro scales. In particular, on our account it is meso scale processes, like the high order modulatory processes involved in self-directedness, that are central to understanding intelligence.

6. Conclusion

The paradigms that we use shape the questions we ask. We have presented an interactivistconstructivist (I-C) approach to theorising intelligence as a particular type of adaptive capacity, one that emphasises the smooth, high order integration of internal with external interaction processes to sustain coherent living conditions. We propose that the complexity of this management task leads to the emergence of a self-directing or self-steering capacity capable of exercising high order regulation to coordinate, through the modulation of these interactive processes, the joint satisfaction of multiple performance norms. Learning and intelligence emerge as sophisticated forms of self-directedness where the normative and anticipative information made available through interaction is used to improve both the current performance and the capacity for such performances, creating a virtuous interactive cycle that permits the solution of initially vague problems. This directs our attention to the search for such integrative capacities in both nervous systems and robots, design questions of a quite different kind than conventional CIP cognitive modelling encourages. Nonetheless, preliminary investigation leads us to believe that this perspective can be fruitfully fused with contemporary empirical work in neuroscience (and robotics) and, we would hope, with studies in the evolution of intelligent organisation (cf. Miklos et al. 1994). Thus we have a new set of questions to pose and a new set of organisational ideas to proffer whose investigation we hope will make a contribution to the long awaited emergence of an integrated theory of cognition from the disciplines of neuroethology, psychology, cognitive robotics and philosophy.

Notes

- [1] For characteristic works in this approach see, e.g., Beer 1995, Bickhard and Terveen 1995, Brooks 1991, Clark 1997, Etxeberria 1998, Hendriks-Janson 1996, Lakoff and Johnson 1999, Smith and Thelen 1993, Smithers 1995, van Gelder 1995, 1998. See further note 27. This approach often labels itself noncognitivist or non-representationalist to make the point that, in contrast with the currently dominant computationalist information processing (CIP) approach (see Section 5), it does not begin with cognitively interpretable elements but with more basic dynamical processes, though it is ultimately just as concerned to illuminate and explain standard cognitive phenomena.
- [2] There are, e.g., several thousand biochemical interactions going on simultaneously in a single cell and between them these must continuously regenerate the whole cell. Note however that autonomy is not to be identified with structural reproduction of all the parts (as autopoiesis is defined), since in adaptive creatures organised structural change must occur as they develop and adapt. Rather it is the regeneration of the organised interactive processes of a continuously viable whole system that is key. For further discussion see Christensen and Hooker 1998c, 1999a.
- [3] Or, if you wish, selects among possible subsequent performances. Explicitness is thus with respect to what *the system itself* can differentiate with its internal dynamical processes. See Bickhard 1993 and Bickhard and Terveen 1995 for extensive arguments for taking a 'system-level' perspective in understanding representation and epistemics. The explicitness of norms grades up

continuously from a-normative dynamical interaction to dynamically distinguished standing reference conditions. The degree to which a dynamical state or condition is an explicit reference condition or norm is roughly determined by the relative longevity and system-wide impact of the constraints it imposes. We do not assume that such norm indicating conditions are symbolic or involve consciousness.

- This formulation leaves open the nature of the mosquito's internal process organisation through which [4] orientation is achieved. The mosquito may have at least one explicit norm for its gradient tracking process in the form of an internal process which constructs a signal measuring the divergence between actual spatial orientation and the direction of maximal local CO₂ concentration, using the signal to reduce divergence by modifying flight orientation.. This would be a very rudimentary norm because its operation is momentary, spatially local and limited to flight behaviour. However it need not be the case that there actually is an internal signal in the mosquito which performs this function: flight orientation may work through the type of simple contralateral sensori-motor connection that Braitenberg's light-tracking vehicle uses (see note 24 and text). If this is the case the spatial difference in CO₂ concentration is not directly integrated by a signal within the mosquito, but is instead integrated by the whole structure of the mosquito body: sensors simply connect directly to separate motor systems, so that activation differences between sensors generate laterally different motor outputs which are finally integrated as an orienting torque through the relative rigidity of the mosquito's body. (It must maintain some appropriate spatial relation between the motor systems and between them and the body as a whole.) There is thus, on our account, a real organisational difference between having an explicit norm and not having one. Whether mosquitos have them for their blood-acquiring activity is an open issue, though there is neuronal evidence that, as with bumblebees (Section 4), they in fact have several operative norms (see Klowden 1995). Our contention that intelligent organisms have many such norm signals available is based on the sophistication already found in insects, and qualitative arguments concerning the need for intelligent organisms to integrate many factors in producing appropriate action, and on neuronal evidence of at least one major plausible supportive architecture (the mesencephalic dopaminergic architecture, discussed in Section 4). Equally, we recognise that the mosquito may have more than one operative norm governing flight, while still responding stereotypically, and/or it may possess dynamical integration of behaviour over time that goes beyond stereotypical reaction (cf. Section 3). Experience with detailed modeling of even apparently very simple real and artificial systems quickly reveals their surprising dynamical complexity (e.g. Beer 1995). We use the mosquito, and the cheetah (see below) and later the detective, effectively as model systems to develop the distinctions we consider important to understanding the process organisation underlying intelligence; while we have tried to remain empirically reasonable in what we do attribute to these creatures, we attribute no more to them than is necessary for our purpose. (In this we parallel the treatment of the bumblebee to be discussed below.) Were it shown that any of these were still more *cognitively* complex that our attributions warrant (they are certainly more dynamically complex), that in itself would only alter our cognitive classification of them, not our analysis of cognitive organisation which is the focus here.
- [5] We shall shortly discuss an empirically-based neural bee foraging model which assigns to bees an elementary self-directive capacity, so we have no desire to preclude other arthropods such as spiders from showing similar rudimentary capacities, and we leave the issue as an open empirical one, though we hope that the conception of self-directedness developed here will lead to a sharper, experimentally accessible characterisation of such capacities. We are indebted to Naor 1993 and to David Naor for stimulating discussion on generalised gradient tracking as intelligent heuristic policy.
- [6] Cf. Smithers' concept of expanding the interactive present (Smithers 1995).
- [7] As the mosquito illustrates vividly, for primitive creatures there is no separation of anticipatory and normative signals or conditions, rather both are simply implicit in the overall organisation of the interactive dynamics. Separation of anticipatory and normative aspects comes by degree and is a feature of more sophisticated systems like the cheetah. If, e.g., there are explicit reward/punishment norm signals (such as pain) available to the system then it can distinguish this

information from sensory-derived anticipatory information. But even for creatures like cheetahs all signals will carry both kinds of information; perception of a gazelle carries both its physical character and its salience as hunger satiator, and pain itself can not only indicate malfunction but provide useful kineasthetic and proprioceptive information about bodily disposition. The anticipatory and normative aspects emerge only in terms of how the system uses the signals to modulate its activity; they are differentiated by the system, not inherent in the signal. To the extent that a system uses a signal to modulate the time structure of its interactive dynamics in relation to the environment it treats the signal as carrying anticipatory information and to the extent that it uses a signal to modulate the generation of action in relation to bodily conditions it treats the signal as carrying normative information. In the case of humans the young baby does not immediately show strong anticipatory/evaluative differentiation, rather, that only emerges through subsequent development; in the adult, anticipatory modeling of interaction is sufficiently well differentiated from evaluation, at least at the conscious level and for normal functioning, that it is reflected in a separation of descriptive from prescriptive, fact from value. It because anticipation has both predictive and normative aspects that we are able to learn about values as well as facts.

- [8] Ecological risk concerns the likelihood while chasing of being attacked by others, e.g. lions or elephants, or of one's cubs being attacked, the risk of not instead satisfying other needs, like thirst, in those circumstances, and so on, and ecological location includes the physical features relevant for such risks: presence of other species, location of water and cubs, etc.
- [9] Thus intentionality, like intelligence, is measured by, and derives from, self-directedness. They are thus understood to be distinct, yet intimately interrelated, aspects of the same directive process organisation, in much more richly articulated ways than CIP, with its I-T-O modularism, can naturally provide.
- [10] The severe incapacitation which results from absence of such signals can be seen in leprosy. Further to our text remarks on the importance of asynchrony, we note here that the dynamics of all the processes we discuss depends on asynchronous interaction; some elements of the system's structure and organisation must be relatively invariant compared to others in order to serve as directing constraints on them. For example, that the human skeleton in general changes on a longer timescale than that of behavioural processes is crucial to its providing constraints which precisely enable coherent, predictable behaviour like running and throwing. In particular, the dynamics of self-directing processes depends on asynchronous interaction between high and low order processes; overall goals and organisation must remain sufficiently constant for specific component activities to have coherence and point. If Sleuth's overall conception of the case changes more rapidly than, say, the investigation process, the latter will be in danger of irrelevance or, worse, of destroying evidence or cover. Purely computational theories have no natural place for timing (van Gelder and Port 1995) and these crucial features of real embodied intelligent processes tend to be suppressed.
- [11] Science is a self-directing process that is more powerful again, being able to anticipatively change its general methods, including experimental procedure and theory construction, and its high-order goals by enriching its primary epistemic norm, truth, with supporting surrogate norms (consistency, controllability, intelligibility, informativeness, etc.). The equivalent of Sleuth's general profile construction is explicit in proto-theories (e.g. of space-time, measurement, and statistical inference) and supported by the rich generalised construction tools of logic and mathematics. One of the chief uses of the scientist's detailed, powerful, theoretically structured anticipative model for an experimental procedure, combined with the rich spectrum of epistemic performance norms on experiments (such as reliable reproducibility), is precisely to specify the loci and nature of potential errors in order to anticipatively check for, and correct for, them. Another is the complementary capacity for increasingly refined evaluation of which experimental procedures can be reasonably entrenched in standard laboratory practice and effectively propagated throughout the relevant discipline(s). See further Christensen and Hooker 1999b, Hooker 1995.
- [12] Or at least, the input data, which is as much of the environment as appears in Quartz and Sejnowski's information processing model of learning. As we will discuss below, and should be clear from the previous discussion, this is an inadequate way to model the influence of the environment on learning.

40 W. D. CHRISTENSEN & C. A. HOOKER

- [13] This general picture of neural development shows interesting resemblances to that proposed by Cunningham 1972, whose pioneering ideas seem to have been largely ignored. Cunningham proposed a model of simple Hebb-like growth mechanisms for neural connections and argued that one could extract a resulting temporal sequence to neural organisation whose functional capacities mirrored Piaget's developmental stages. For some discussion of Cunningham's model and its contrast to conventional control models see Hooker 1996, pp.192-4. The question posed there: 'How can a system be (self-)organizing so as to improve control capacity?' is here answered functionally with SDAL and neurologically with the Cunningham/Quartz-Sejnowski neural capacity for parametric self-modification which makes neural organization sensitive to collective variables. This constructive approach in fact stands in a tradition that goes back at least to Pask 1960, 1981; see also Cariani 1992, 1993 for discussion, sympathetic to, and complementary to, that provided here.
- [14] Insofar as 'represents' is used loosely to mean simply 'organised in relation to', there is some sense to the way in which Quartz and Sejnowski use the term and the matter could be left there were it not for the fact that it has a particular connotation within the standard CIP paradigm that makes its use here highly misleading. In that paradigm what an internal condition represents is the external condition that stimulates it (and was the chief cause of its formation), what, in the standard language-like reading, the representing symbol array refers to. This is upstream or reverse causality representationalism. It is notoriously inaccessible to real creatures because they have no natural dynamical access to what sends them their environmental stimuli, only to what they do with stimuli once received. Of course they can respond by attempting to further interact with the sender/sending condition, but this is itself a directed, causally downstream response to the stimulus, not a reaching upstream. Their downstream response is also the only thing they can subsequently modulate in relation to the further cycle of stimulation that results, not the signal origin per se. Thus we are led instead to downstream representationalism, where the significance of a signal is what the system does in response to it and any internal representation of its origin must be a sophisticated construction that emerges from this activity, not that defines its significance. This provides system-accessible forms of natural significance and representation, ones directly connectible to system norms and anticipative models through self-directing processes in the manner discussed in note 15. The difference is not trivial. Upstream representationalism concentrates attention on the encoding of correct information as the most critical aspect of cognition; the learning problem and the learning process are viewed as quite distinct things, and learning can be idealised as passive pattern conformation. In contrast, a downstream modulation model of cognitive activity concentrates attention on output, where the most important measure of success is appropriateness of behaviour in the circumstances. (See also Bickhard 1993, Bickhard and Terveen 1995 for further discussion.) If downstream process modulation is substituted for representation as the 'general currency of the brain' (cf. Quartz and Sejnowski 1997 p.538), then the insight that learning involves organisational change which reflects environmental regularities is preserved, without prejudging the nature of the adaptive relation. Furthermore, attention is directed towards a more dynamic account of learning.
- [15] Connectionism relaxes the extreme modularism of AI by allowing distributed processing architectures that yield markedly less dynamical modularity because of the resulting global character of the relevant connectionist net states. More significantly, connectionist nets introduce a degree of plasticity (in the node weights). The resulting interaction between a connectionist network's dynamical characteristics and its inputs (including, for supervised learning, any error-correcting inputs) breaks down the sharp distinction between processing architecture and input content, allowing the functionality of a connectionist net to emerge from this interaction rather than be predetermined by a fixed representational space, as is the case with AI. An interesting consequence is that the cognitive process boundaries become 'smeared', inasmuch as a functional connectionist net cannot be characterised independently of its environment in the same way a Turing machine can. Despite this interactivist clue, which really undermines I-T-O modularity, connectionist methodology and rhetoric unquestioningly retains the general internalist assumption that cognition is basically a form of problem solving input-output processing that can be given a largely intrinsic characterisation.

- [16] See Cummins 1984, and Hooker 1989, Section IV.2.3 for an overview of Cummins' difficult but valuable analyses in a dynamical systems setting.
- [17] It is often remarked that biological organisation is 'messy' and ad hoc, littered with 'Rube Goldberg mechanisms' which result from the inherent vagaries of the evolutionary process. There is an element of truth to this, but it also reflects a human prejudice for top-down engineering which does grave injustice to the power and subtle elegance that can also be found in biological systems. Current human engineering is hopelessly crude and simplistic by comparison. When Lockheed-Martin can design stealth bombers with the fault tolerance, self-maintenance and 'smart' capacities of a barn owl, and manufacture and fuel them using resources as cheap and readily available as mice, it will be really on to something.
- [18] See, e.g., examples in the articles on complexity in *Science*, April 1999, and Collier and Hooker 1999 for an account of some relevant aspects of organisation in biological adaptive systems. Cf. Karmiloff-Smith 1992.
- [19] That is, where the bee must choose between a strategy which yields a small but reliable reward and a strategy which yields a larger but less reliable reward, given that the net outcomes for either strategy are equal. Real 1991, who posed the problem, is very clear about the formal computational, decision theoretic problem formulation to be used to model living agents.
- [20] At least to the extent that the process is in large part implicit, which every real process that can support learning deeply is. There is, clearly, no well defined global optimisation problem posed by living; and if there were, no creature could have access to it (since it concerns both knowledge of self and of future no creature has in advance). However it is true that once anticipative models have been constructed for kinds of goal-achieving processes, e.g. for gazelle hunting, these processes can thereafter be improved against their explicit norms. Although they must remain context-sensitive in application (the hunt must be for *this* gazelle in *this* situation), it might be thought that the improvement process can best be modelled by some higher-order optimising process. Perhaps on occasion it can; e.g. humans certainly find it useful to use optimising models in situations that are sufficiently well understood theoretically. But even here it often proves gravely misleading whenever the real world situation outruns the modular modelling constraints (as it notoriously does, e.g., in economic and social action contexts). For the reasons set out in (1) and (2) we see neither the necessity for, nor much support for, taking this approach as *basic* and leave its reasonable application to further study.
- [21] It is reasonable to expect that essentially the same constructive processes will account for both early development and later learning, though the latter will occur in a more richly elaborated constraint framework. In this century the pioneer and doyen of constructivism has been Piaget, who expressed the position succinctly in his own framework:

.... knowledge arises neither from a self-conscious subject, nor from objects already constituted (from the point of view of the subject) it arises from interactions that take place mid-way between the two but, by reason of their complete undifferentiation, if there is at the start neither a subject in the epistemological sense of the word, nor objects conceived as such, nor invariant intermediaries, the initial problem of knowledge will therefore be the construction of such intermediaries (Piaget 1972, pp. 19-20).

Quoted also at Hooker 1995, p.229, for which see for further discussion. This constructivist tradition was developed in potentially fruitful ways by Piaget's contemporary Vygotsky (1986) (even while Piaget was being seduced into an internalist formalist psychology of 'stages'), by Rummelhart and McClelland (1986) and Bickhard (e.g. Bickhard 1993, Bickhard and Terveen 1995, Bickhard and Campbell 1996). Quartz and Sejnowski provide a welcome neurophysiological contribution to this tradition, following Cunningham (1972), see note 14.

- [22] Though we add the qualifications that mathematical science exhibits many non-computable functions and many non-linear n-body processes that are in principle not tractably modelable.
- [23] See Braitenberg 1984. His photo-taxic vehicle consists just of a rectangular frame supported by a wheel at each corner laterally aligned in pairs in the usual way, and equipped with two

photoelectric sensors attached to the 'front' corners and two independent electric drive motors on the 'rear' wheels, with sensors connected to motors contralaterally. Then the inverse-square difference in light intensity at the receptors from a localised distant light source drives the rear motors differentially so that the vehicle automatically turns towards the light, the contralateral sensori-motor connection ensuring that the 'far' motor is driven harder by the 'near' photoreceptor. Thus photo-taxis is effected by just contralateral wiring connecting a-directional sensor output with a-directional motor response; overall direction is only created because they are held in suitable (here rigid) mutual spatial relations by the vehicle body.

- [24] Of course, as systems further interact with their environment on this basis they will come to associate and differentiate environmental feedback signals and respond to them as clusters and in this way construct an interactive representation of a causing environmental situation as a norm-evaluated flow of actions and feedbacks (i.e. of what possibilities for interaction and rewarding feedback there are) in something like the manner Piaget describes (note 22). But this is a sophisticated construction, not a founding semantic condition. (The human infant takes more than a year to achieve a rich object constancy, and the peculiarities of quantum interactions quickly rob even experienced adults of it.) Similarly, as both norm matrices and anticipative models become increasingly elaborated and integrated, there will emerge a semantics of internal modulatory states which, at least in part, is increasingly classical in structure; but again this will be a sophisticated construction, not a founding semantic condition, and much, likely most, of a human's effective interactive organisation will not be included in this specialised domain. See further note 26 and Bickhard 1993, Bickhard and Terveen 1995, Christensen and Hooker 1998a, b, Hooker 1992. We are indebted to Bickhard for valuable instruction on this matter.
- [25] Possibly, e.g., for linguistic activity, but we retain an open mind about this until the process organisation underlying linguistic competence is unraveled (rather than just 'just so' modelled) - cf. note 25. Conversely, we do not claim that there are no norms, anticipative models or semantics to be found simply because what emerges from dynamical modelling is dynamically characterised (as Beer, commentary on van Gelder 1998, suggests). Rather, these are to be discovered by careful empirical organisational analysis. Whence we do not preclude discovering the need to attribute symbolic processing to some activities.
- [26] This orientation is well illustrated in Smith and Thelen 1993, Smithers 1995 and van Gelder 1995, 1998, exemplified by van Gelder's holistic dynamical d.e. model of the Watt-steam-governor-and-steam-engine as paradigm for intelligent control (van Gelder 1995), the studies in Smith and Thelen of the emergence of crawling in infants as dynamical bifurcation, and the attempt by Smithers to characterise autonomy in terms of the differential morphology of interaction fields. For further general discussion see van Gelder and Port 1995, van Gelder 1998 and Christensen 1999, chapter 1.
- [27] It is always possible to capture the internal organisation of the system by modelling it as a system of coupled component subsystems. However there is then no principled basis for: (a) specifying when such organisation should be explicitly modeled, or (b) individuating the system in a principled way. In consequence this modeling procedure is by itself too weak to capture the distinctions crucial to delineating the cognitively relevant differences among systems, at least of the sort the I-C approach regards as central. For another aspect of this issue see note 26.
- [28] Note, however, that the issue is not distinctive to humans (or humanoid robots), but of fundamental importance to all forms of intelligence, which is why we chose to illustrate it with the cheetah example.

References

- BECHTEL, W. & RICHARDSON, R.C. (1993) Discovering complexity: decomposition and localization as strategies in scientific research. Princeton: Princeton University Press.
- BEER, R.D. (1995) Computational and dynamical languages for autonomous agents. In PORT and VAN GELDER 1995.

- BICKHARD, M.H. (1992) Scaffolding and self scaffolding: central aspects of development. In L. T. WINEGAR and J. VALSINER (eds) *Children's development within social contexts: research and methodology*. Hillsdale, N.J.: Erlbaum.
- BICKHARD, M.H. (1993) Representational content in humans and machines. *Journal of Experimental and Theoretical Artifical Intelligence*, **5**, 285-333.
- BICKHARD, M.H. & TERVEEN, L. (1995) Foundational issues in artificial intelligence and cognitive science—Impasse and solution. Amsterdam: Elsevier Scientific.
- BICKHARD, M.H. & CAMPBELL, R.L. (1996) Topologies of learning and development. New Ideas in Psychology, 14(2), 111-156.
- BRAITENBERG, V. (1984) Vehicles: experiments in synthetic psychology. Cambridge, MA: MIT Press.
- BROOKS, R.A. (1991) Intelligence without representation. Artificial Intelligence 47, 139-159.
- BROOKS, R.A. (1997) From earwigs to humans. Robotics and Autonomous Systems 20: 291-304.
- BROOKS, R.A. & STEIN, L. (1993) Building brains for bodies. *MIT AI Lab Memo #1439*, August 1993.CARIANI, P. (1992) Some epistemological implications of devices which construct their own sensors and effectors. In F. VARELA and P. BOURGINE (eds) *Towards a practice of autonomous systems*.
 - Cambridge, Mass.: MIT Press.
- CARIANI, P. (1993) To evolve an ear. Systems research 10(3), 19-33.
- CHRISTENSEN, W.D. (1999) An interactivist-constructivist theory of intelligent embodied agents. PhD thesis, University of Newcastle.
- CHRISTENSEN, W.D. & HOOKER, C.A. (1998a) Towards a new science of the mind: wide content and the metaphysics of organisational properties in non-linear dynamical models. Symposium on Mind. Special Issue, *Mind and Language*, 13 #1, 97-108.
- CHRISTENSEN, W.D., & HOOKER, C.A. (1998b) Symposium: Paul M. Churchland, The engine of reason, the seat of the soul: A philosophical journey into the brain. *Philosophy and Phenomenological Research*, **LVIII** (4), 871-878.
- CHRISTENSEN, W.D., & HOOKER, C.A. (1998c) From cell to scientist: toward an organisational theory of life and mind. In J. BIGELOW (ed.) *Our cultural heritage*. Australian Academy of Humanities, University House, Canberra.
- CHRISTENSEN, W.D. & HOOKER, C.A. (1999a) Organised interactive construction: the nature of autonomy and the emergence of intelligence. To appear in A. ETXEBERRIA and A. MORENO (eds.) *Communication & Cognition*, Special edition on autonomy.
- CHRISTENSEN, W.D. & HOOKER, C.A. (1999b) The organisation of knowledge: beyond Campbell's evolutionary epistemology. *Philosophy of Science*, **66** (Proceedings, PSA 1998), S237-249.
- CLARK, A. (1997) Being there: putting brain, body, and world together again. Cambridge, Mass.: MIT Press.
- COLLIER, J.D. & HOOKER, C.A. (1999) Complexly organised dynamical systems. *Open systems and information dynamics*, 36, 1-62.
- CUMMINS, R. (1984) Functional Analysis. In E. Sober (ed.), *Conceptual issues in evolutionary biology*. MIT Press, Cambridge.
- CUNNINGHAM M. (1972) Intelligence: its origin and development. New York, Academic Press.
- EATON, R. L. (1974) The cheetah; the biology, ecology, and behavior of an endangered species. New York: Van Nostrand Reinhold Co.
- EXTEBERRIA, A. (1998) Embodiment of natural and artificial agents. In G. VAN DER VIJVER, S. N. SALTHE and M. DELPOS (eds.) *Evolutionary systems*. Dordrecht: Reidel.
- FODOR, J. (1992) Precis of The modularity of mind. Behavioral and Brain Sciences 8, 1-42.
- GRIFFITHS, P.E. & GRAY, R.D. (1994) Developmental systems and evolutionary explanation. *Journal of Philosophy* **XCI**(6): 277-304.
- HENDRIKS-JANSON, H. (1996) Catching ourselves in the act: situated activity, interactive emergence, evolution and human thought. Cambridge, MA: MIT Press.
- HOOKER, C.A. (1989) Evolutionary epistemology and naturalist realism. Part IV of K. HAHLWEG and C.A. HOOKER Evolutionary epistemology and philosophy of science. In K. HAHLWEG and C.A. HOOKER (eds.), *Issues in evolutionary epistemology*. Albany N.Y.: State University of New York Press.
- HOOKER, C.A. (1992) Physical intelligibility, projection, objectivity and completeness: the divergent ideals of Bohr and Einstein. *British Journal for the Philosophy of Science*. **42**, 491-511.

HOOKER, C. A. (1995) Reason, regulation and realism. Albany: SUNY Press.

- HOOKER, C. A. (1996), 'Toward a naturalised cognitive science', in R. KITCHENER and W. O'DONOHUE (eds.) (1996), *Psychology and Philosophy*. London: Sage.
- HOOKER, C.A., PENFOLD, H.B. & EVANS, R.J. (1992) Towards a theory of cognition under a new control paradigm. *Topoi* 11: 71-88.
- JABLONKA, E. & LAMB, M.J. (1995) *Epigenetic inheritance and evolution: the Lamarckian dimension*. Oxford University Press, Oxford.
- KARMILOFF-SMITH, A. (1992) Beyond modularity: A developmental perspective on cognitive science. MIT Press.
- KEIJZER, F.A. and BEM, S. (1996) Behavioural systems interpreted as autonomous agents and as coupled dynamical systems: a criticism. *Philosophical Psychology* **9**(3): 323-346.
- KLOWDEN, M. J. (1995) Blood, sex, and the mosquito: control mechanisms of mosquito blood-feeding behavior. *BioScience* 45, 326-31.
- LAKOFF, G. & JOHNSON, M. (1999) *Philosophy in the flesh, the embodied mind and its challenge to Western thought.* New York: Basic Books.
- MIKLOS, G.L., CAMPBELL, K.S.W. & KANKEL, D.R. (1994) The rapid emergence of bio-electronic novelty, neuronal architectures, and organismal performance. In R.J. GREENSPAN and C.P. KYRIACOU (eds) *Flexibility and constraint in behavioral systems*. John Wiley & Sons Ltd.
- MONTAGUE, P.R., SEJNOWSKI, T.J. (1994) The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms. *Learning & Memory*, **1**: 1-33.
- MONTAGUE, P.R., DAYAN, P., PERSON, C. & SEJNOWSKI, T.J. (1995) Bee foraging in uncertain environments using predictive hebbian learning. *Nature* **377**, 725-728.
- MONTAGUE, P.R., DAYAN, P. and SENOWSKI, T.J. (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience* **16**, 1936-1947.
- NAOR, D. (1993) Studies in rationality and public policy, Part I: heuristic policies and rational response, part II: the United States Strategic Air Command Basing Study. Presented to the *International Conference on Non-Formal Reason*, Newcastle, Australia.
- NEWELL, A. (1980a) Physical symbol systems. Cognitive Science, 4, 135-183.
- NEWELL, A. (1980b) Reasoning, problem solving, and decision processes: the problem space as a fundamental category. In R. NICKERSON (ed.) *Attention and performance VIII*. Hillsdale, NJ: Erlbaum.
- OYAMA, S. (1985) The ontogeny of information. New York: Cambridge University Press.
- PASK, G. (1960) The natural history of networks. In M.C. YOVITS and S. CAMERON (eds) *Self-organizing* systems. New York: Pergamon.
- PASK, G. (1981) Organisational closure of potentially conscious systems. In M. ZELENY (ed.) Autopoiesis: A theory of living organisation, New York: North Holland.
- PIAGET, J. (1972) The principles of genetic epistemology (trans. W. MAYS). London: Routledge & Kegan Paul.
- PORT, R.T. & VAN GELDER, T. (eds) (1995) *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.
- QUARTZ, S.R. & SEJNOWSKI, T.J. (1997) The neural basis of cognitive development: a constructivist manifesto. *Behavioural and Brain Sciences* **20**(4): 537-596.
- RAFF, R.A. (1996) *The shape of life: genes, development, and the evolution of animal form.* Chicago: University of Chicago Press.
- REAL, L.A. (1991) Animal choice behavior and the evolution of cognitive architecture. Science 253, 980-986.
- RUMELHART, D. E. & MCCLELLAND, J. L. (eds.) (1986), *Parallel distributed processing*, Vol. 1. Cambridge, Mass.: MIT Press.
- SMITH, L. V. & THELEN, E. (eds) (1993) A dynamics systems approach to development: applications. Cambridge, Mass., Bradford/MIT Press.
- SMITHERS, T. (1995) Are autonomous agents information processing systems? In L. STEELS and R.A. BROOKS (eds.) *The artificial life route to artificial intelligence: building situated embodied agents.* Lawrence Erlbaum Associates.
- THELEN, E. (1995) Time-scale dynamics and the development of an embodied cognition. In PORT & VAN GELDER 1995.

VAN GELDER, T. (1995) What might cognition be if not computation? The Journal of Philosophy, XCI(7), 345-81.

VAN GELDER, T. (1998) The dynamical hypothesis in cognitive science. Behavioral and Brain Sciences

21(5), 615-627. VAN GELDER, T. & R. PORT (1995) Its about time: an overview of the dynamical approach to cognition. In PORT & VAN GELDER 1995.

VYGOTSKY, L.S. (1986), Thought and language. (Cambridge, Mass.: MIT Press.)

WIMSATT, W. (1987) False models as means to truer theories. In M.H. NITECKI & A. HOFFMAN (eds.), Neutral models in biology. Oxford: Oxford University Press.