Anticipation in Autonomous Systems: Foundations for a Theory of Embodied Agents

W.D. Christensen and C.A. Hooker
Department of Philosophy, University of Newcastle, Australia. F(ax)mail: +61249216928. P(hone)mail: +61249215186. S(nail)mail: Callaghan 2308, NSW, Australia. Email: plwdc@cc.newcastle.edu.au, plcah@cc.newcastle.edu.au.

Abstract

This paper outlines a theory of anticipation in autonomous systems. Our account of autonomous systems is designed to model the basic organisational form of life. Anticipation is an integral feature of the autonomy account, and is an important *foundational concept for an interactivist-constructivist* (I-C) theory of embodied intelligent agents. We present the basic conceptual framework of the I-C approach to intelligence, including an account of directed processes, normativity as process closure, and self-directedness as the basis of intelligence and learning.

Intelligence is understood as emerging through increasing self-directedness. Self-directed systems anticipate and evaluate their interaction flow, directively modifying the interaction process so as to achieve goals that regenerate or improve the system's autonomous closure conditions. Learning arises out of the drive to improve anticipation, which starts by being contextual, vague, and implicit, and becomes increasingly articulated and explicit as the system constructs anticipative models and goals for *managing* and *evaluating* interaction. Cognitive development occurs through self-directed anticipative learning (SDAL), in which a pushme-pullyou effect is generated as increasingly rich anticipation increases the directedness of learning by improving error localisation, context recognition and the construction of improved anticipation.

The paper concludes with an introduction to a general anticipatory conception of intentional agency, and a correlative critical appraisal of Rosen's pioneering analysis of anticipation.

Keywords: anticipation, autonomy, adaptive, self-directing, system

1 Introduction

This paper presents a view of agency as deeply embedded in the organisational characteristics of life, and of anticipation as a key feature both of living systems generally and of the adaptive specialisation involved in the evolution of intelligent agents in particular.[1] Living systems must interact with the environment to acquire the resources needed to regenerate themselves in the face of continual dissipation and damage. *Autonomy* is the possession of this interactive ability to dynamically generate system integrity. This is a global organisational constraint on living systems requiring that their processes all so interact that together they regenerate themselves. Autonomous systems are characteristically internally complex and must interact successfully with complex environments, so remaining viable requires maintaining process coherency across multiple constraints and timescales. Adaptiveness, therefore, is deeply bound up with the modes of organisation involved in achieving this complex and dynamic global process coherency. The paradigm examples of autonomous systems are uni- and multi- cellular organisms, but colonies, ecologies, business firms and cities also show related organisation and functions. They are in this key respect unlike other merely very complex organised systems, such as fires, swamps and computers. An autonomous system's interactions are measured against the benchmark of preserving autonomy, which thus constitutes the basic normative constraint on adaptive processes. All of the more specific normative constraints on particular actions (e.g. avoid hunger, pain) derive from this global constraint.

To anticipate is to act now in relation to some future state, event or process. Anticipation is thus an integral feature of autonomous systems because of their need to shape dynamical interaction with their environment in ways that achieve future outcomes that contribute to the system's integrity. (Note that shaping interaction includes all interaction-relevant system outputs: modifying the environment, e.g. by spinning a web, and modifying internal system state, e.g. to evaluate sensory signals differently, as well as generating behaviour, e.g. walking, to change the system-environment relationship, or licking a wound to heal oneself.) The interactive relationship between the present action performed and the future, autonomy-evaluated outcome required is the most basic form of anticipation, according to our account. The future outcomes required are those that sustain system autonomy and we shall call these the *closure conditions* of the processes. The hunger-hunt-catch-eat process, e.g., achieves its immediate closure with the processes that produce satiation and its deeper

[1] We have developed the framework of ideas in detail elsewhere – see Christensen and Hooker (1998, 1999a, b). Here the focus will be on presenting the basic conceptual structure as it relates to anticipation. Our work on order and organisation, and that on the root idea of autonomy, is drawn from collaborative research with John Collier and we wish to acknowledge his significant contribution.

closure with the processes that produce nutrition. Viability requires that the system achieve an integrated overall adaptive 'process flow' of this kind. It is the fundamental normative requirement for an autonomous system since it is the sine qua non of its continuing identity as that autonomous system. Thus all autonomous systems rely on an internal capacity for *directed interaction*: anticipative modulation of system action so as to shape the system-environment interaction process in ways that will achieve the closure conditions for autonomy, and the system process organisation which provides this capacity is its *directive organisation*. The problem of understanding adaptive behaviour is one of modelling the way in which the system directive organisation interacts with environmental order and system constraint to satisfy system autonomy. Autonomy, adaptiveness and anticipation are thus closely integrated, and we may speak of AAA systems, and the AAA-ness of life.

Our approach to intelligent agency is to characterise it as a particular type of adaptive mode of organisation: one focussed on high order interaction management. The most elementary kinds of anticipation involve simple signal-action-outcome relations. Mosquito's, for example, engage in fixed action pattern-like responses to carbon-dioxide concentration (they orient flight in the direction of highest concentration), and this allows them to find blood hosts (the required outcome). However, by increasing its ability to anticipate and evaluate the interaction process a system may encompass multiple features of a context and/or multiple action possibilities in its performance. This in turn can improve the system's adaptability. Cheetah's anticipate the influence of many kinds of variables on hunting, such as type of prey, cover available, etc., and this allows them to act fluidly and appropriately in a complex and changing context. We call this type of ability self-directedness because it involves modifying performance to suit the context, thereby achieving the system's goals despite the variation of important factors. Self-directedness thus involves high order process management to achieve the normative outcomes the system requires (this normative perspective is the 'self' of self-directed) whilst varying many specific aspects of performance. On our account it is the primary capacity involved in the development of intelligence. Self-directedness involves many inter-related factors, in this paper we shall explore them with a focus on the role of anticipation.

The centrality of directed interaction marks the essential difference in orientation between autonomy and autopoiesis (Varela et al. 1974, Maturana 1981). Both concern open systems and their regeneration or 'self-production'. But for autopoiesis the operative paradigm is one of an internally closed set of interaction processes, e.g. a system that can manufacture all its own distinctive components within itself. Here imports and exports of matter and energy may be dynamically essential but do not participate in defining process organisational closure (see also Mingers 1995). By contrast, for autonomy the paradigm is the system that actively, directively constructs and/or compensates for external dependencies, and constantly changes itself as it manages its interactions to respond adaptively to its environment. Here the organisation of process closures essentially includes the interacting aspects of the environment, but this is compatible with the internal locus of active, directive construction characterising such systems.

(*Technical aside.* Since we speak in this paper of order and organisation, we briefly characterise their technical meanings here. The root notion of *order* is that derived from algorithmic complexity theory: the orderedness of a pattern is the inverse of the length of its shortest, most compressed, complete description. Redundancy or correlation orders are determined by the minimal number of elements in which a redundancy can be detected. When we speak of *high order* features we refer to features characterised by high order correlation relations, relatively independently of whether they concern highly ordered features. Organisation is a particular kind of ordering involving relatively high order relations. Gases are disordered and hence unorganised but regular crystals are highly ordered though very simply organised because their global ordering relation is highly redundant. By contrast (roughly) machines and living things are organised because their parts are relatively unique and each plays distinctive and essential roles in the whole. That is, an *organised* system displays a non-redundant global ordering relation of relatively high order - though for this reason organised systems are less highly ordered than are crystals. A system's *organisational depth* is measured by the degree of nesting of sub-ordering relations within its global ordering relation (cf. cells within organs within bodies within communities). In these senses living systems are deeply organised, and have many very high order constraints, processes etc. The global constraint to autonomy is one of these. On the principled dynamical characterisation of organisation see Collier and Hooker 1999 and references.)

2 Dynamical anticipation

All living systems are autonomous, including bacteria, plants, and so on, but only some are intelligent; these systems have, we believe, specialised their directive organisation to enhance their capacity for anticipative modulation of interaction. While these features are the ingredients from which intelligence is formed, in most adaptive systems they occur in elementary forms. There are three major aspects determining a system's anticipative modulation power: the width of its anticipatory time window, the degree of articulation of the autonomy-related norms which it can use, and the order of the system and system-environment relationships that it can effectively modulate. Since these aspects or dimensions are themselves multi-faceted, and our approach to cognition is a multi-dimensional integrated one: there is no single 'mark of the mental', instead there are complexes of adaptive capacities that combine in various ways to form adaptive strategies; actual intelligent capacities in nature can be expected to be found in many different locations throughout this multi-dimensional space according as self-organisational processes (genetic, epi-genetic) and dynamical constraints (historical, selection, organisational and basic nomological) permit. We will now discuss the basic form of anticipation and some of the ways it can be elaborated to produce intelligence.

Anticipative actions are future oriented and have a natural time-window determined by the characteristic time-scales of the interaction processes and the autonomous closure conditions of the system in which they are embedded. As such even an elementary directed process in which a signal I initiates an action **a** involves simple dynamical anticipation of the form: 'Performing action **a** *now* (in response to the occurrence of signal I) will *in future* generate the closure conditions for **a** (say system condition of type **x**, within time-window t_w)'. A hunger signal (I), e.g., initiates hunting action (**a**) of a kind which anticipates satiation within the time to prey dispersal and/or hunting incapacitation (t_w). This example also makes it clear that anticipation in this sense is not basically linguistic in form, but rather has a non-propositional dynamical nature. In simple directed processes this anticipation is implicit in the process organisation, measured only by the health, and ultimately life or death, of the system, but in more complex directed processes at least some components of the dynamical anticipation can be enriched and made more explicit, and this is central to the emergence of intelligence.

A simple but fundamentally important form of dynamical anticipation involves distal perception and mobility. As Smithers (1995) points out, the presence of distal perception processes, e.g. vision, in mobile systems such as organisms and robots allows these systems to in a sense 'see' into the future, inasmuch as forward-looking perception provides information concerning environmental conditions with which the system will very shortly interact, thereby expanding the system's 'interactive present'. Thus, realised through its modulatory effects on the system's motor and other processes, distal perception functions as a means for the system to project anticipatively into the future. Memory processes, on the other hand, provide a means to extend the interaction time-window into the past, allowing the system's interaction history to have a modulatory influence on its current state. Memory can also facilitate dynamical anticipation by generating expectancies concerning regular relationships in the system's interaction with the environment. This kind of learned expectancy can be realised in very simple conditioning processes such as the desensitization of a reflex.

More complex memory processes can facilitate more detailed forms of dynamical anticipation, as in the case of dynamical emulation. In many organisms neuronal systems involved in motor activity learn to emulate aspects of the dynamics of motor tasks such as reaching and grasping. These emulators are then able to supply context-appropriate directive signals more rapidly than is possible with sensory feedback loops. This process (also ubiquitous in control engineering) provides smooth and effective anticipative motor activity (see, e.g., Grush 1997). To illustrate the power of this form of dynamical anticipation, consider catching a ball. The most effective way to catch a fast moving ball is to anticipate the ball's spatio-temporal trajectory and move so as to intersect it. Simply moving towards the current location of the ball will likely defeat the aim since by the time your hand gets there the ball will have moved on.

Emulation processes can range from relatively contextual and immediate motor signals to relatively more 'offline' imagination processes that can operate in the absence of overt behaviour. Imagination greatly enhances the capacity for dynamical anticipation by allowing the system to partially decouple its directive processes from the immediate context, permitting offline rehearsal and exploration of interactive possibility. The latter is particularly important, since opening up the capacity for modal anticipation permits high order cognitive processes such as resolution of competing goals and planning ability.

To sum up, increases in dynamical anticipation capacity enrich and expand the system's time-window for directed interaction, simultaneously reducing local context-dependency and improving context-sensitivity by allowing the system to shape its actions over longer timescales and with respect to more detailed, in some cases modal, information concerning the flow of the interaction process. As will be discussed below, these capacities are important for strong forms of self-directedness.

3 Normative evaluation: Measuring the success of performance

The elaboration of anticipation is a key part of the development of intelligence, but another equally important aspect is the normative evaluation of performance. Normative evaluation allows a system to identify sources of success and error in its interaction, and to modify its actions when things do not go well. Indeed, evaluation plays an integral role in the construction of anticipations and, in complex cases, of constructing goals.

For a system to normatively evaluate its performance it must have some means of measuring whether its directed processes actually achieve closure. To this end it requires modulatory signals that act as proxies for closure. Organisms typically possess an array of affective and aversive signals that serve this function: pleasure, pain, hunger, etc. Such signals can be more or less action specific – satiation, for example, is specific to food consumption (indicating success), whereas happiness is a less action-specific evaluative signal (it might be induced by many different kinds of activities). We shall refer to relatively action-specific norms as operational norms (ONs), and to those that relate to more general functional conditions as integrative norms (INs). In a cheetah's hunting-feeding process cycle there are two basic ON signals – hunger and satiation. Hunger measures departure from the blood sugar closure level, thus triggering the hunting-feeding process cycle, whilst satiation measures the restoration of that closure level, terminating the process. ONs thus provide normative evaluation of the operational success of interaction (*this activity* is going right/wrong) – for instance, if hunting is unsuccessful hunger will drive the cheetah to keep trying. INs, such as generalised discomfort or pleasure, are related to more general system health conditions (*this overall aspect* is going right/wrong) and form a complex of higher order normative evaluations. In combination ONs and INs form a multiordered matrix that allows a system to direct and evaluate the interaction process with respect to its affective value for the system.

A subtle but important feature of evaluative signals is that they allow a system to learn about what produces success and error. If a cheetah cub hurts itself by falling when trying to climb a tree it will in future attempts anticipate this error and be more cautious about achieving secure grip and balance. This process can allow the discovery of implicit closure conditions (here grip and balance) and an enfolding of these conditions into explicit anticipative action modulation. Thus, cheetah cubs do not initially know that careful stalking is an important condition of success for hunting, only learning this through imitation of the mother and failures when the prey detects them too soon. In this way the skills of stalking become a focus of cheetah learning about effective hunting. The cheetah is always trying to modulate its interaction flow so as to balance its degrees of satisfaction of its autonomy norms; in these examples the cheetah learns to elaborate its normative matrix, both by inserting new operational norms (grip/balance, close stalking) and by adjusting the relations among them in the hunting process, e.g. by inhibiting the hunger-driven desire to chase until these new goals are met. (But note that these relations among norms only ever appear as aspects of the modulation of interaction flows, the normative matrix is not an abstract entity but an aspect of dynamic modulation processes.) This picture outlines a general model of learning processes in organisms.

4 Self-directedness

Anticipation and evaluation are the basic ingredients of directed interaction. The development of intelligence involves an elaboration of these capacities to increase the *self-*directedness of the system. Self-directedness is the ability to flexibly achieve goals in a complex variable context. The discussion will now turn to the way in which directedness varies in degree from simple directed interaction to sophisticated self-directed interaction.

Achieving context-sensitive performance in a complex variable context hinges on the ability to anticipate what actions will be successful in the particular context the system finds itself in, and adapting the performance to suit. Thus, increasing self-directedness involves improving the richness of the system's interaction window. The various aspects of the interaction window, including its modalities, scope, predictiveness, evaluative power, and the extent to which it includes learning, all affect the ways this improvement can occur and the ultimate nature of the self-directedness of the system. An organism that walks on land and navigates visually, for instance, will have an interaction window with quite different characteristics to one that swims and uses sonar. Intuitively, the effectiveness of a system's interaction window depends on the system possessing a suitable repertoire of interactive skills for satisfying its functional requirements in its environment, and on it being able to anticipate the context sufficiently well to produce the actions that will result in the appropriate outcomes for the context. For example a cheetah needs a repertoire of hunting skills, and it must finely tune these skills to suit the context, such as choosing catchable prey, assessing sufficiently safe chase landscapes, using available cover to stalk the prey and getting sufficiently close that it will not be able to escape whilst not so close that the prey will detect the cheetah before it attacks.

Four examples will be used to illustrate the way in which systems can become increasingly self-directed – Climber, Catcher, Sleuth, and science. Climber has a gradient tracking procedure that allows it to find the tops of mountains simply by measuring local height variations with its foot and stepping in the direction of greatest elevation. This simple directed process relies almost entirely on the structure of the environment in the generation of overall performance. Climber has little operational plasticity and no evaluative capacity, and so is unable to modify its behaviour to suit variable circumstances. Thus, although Climber is capable of directed interaction, it is not self-directed. Catcher, on the other hand, uses perception of a ball's flight and internal emulation of trajectories to project the trajectory of the ball and uses this anticipated trajectory to guide running, reaching and grasping movements. This makes Catcher self-directed because it uses information from its interaction with its environment to context sensitively modify its further interaction in ways that lead to the achievement of closure (catching the ball).

Self-directedness involves a constructive aspect – the system uses information from interaction to modify what it does. By modifying itself in increasingly sophisticated anticipatory ways this feature allows the system to exploit more complex environmental regularities than can simpler directed systems such as Climber. Self-directed systems are able to 'track' generalised environmental gradients – not just simple spatial gradients but high order gradients in time, food quality, survival riskiness, social or epistemic interest, and so on.

They may also indirectly increase their self-directedness, as well as their interactive effectiveness, by constructing such gradients. An important way systems shape their interactions is through using features in their environment as cues to action, i.e. as scaffolds (Bickhard 1992, Clark 1997). This may range from simple physical or chemical markers (cf. respectively termite mound construction, ant food trails) to sophisticated human cultural dress cues and computer systems. If a system constructs such environmental cues whilst they in turn also induce changes in its performance ability, then that scaffolding forms part of its self-directedness. Consider the way *writing* affected law-making by providing stable records that could serve as a focus for shaping, and being shaped by, social decision making. (If the cues only affect the expression of actions but not their form, as is the case in termite mound-building, then although they may act as a powerful organising force they do not contribute to self-directedness.)

As an illustration of a constructive gradient-tracking process consider Sleuth, who uses clues from a murder to build a profile of the suspect and then uses this profile to further refine the direction and methods of the investigation. The investigation is not a simple gradient tracking procedure because Sleuth doesn't simply trace a spatial gradient as Climber does. Instead, there is an interplay between the discovery of clues, the construction of a suspect profile and subsequent modification of the investigation strategy (method modification). Sleuth must use initial evidence to develop a profile of the murder and the suspect before the investigation process can find the subtle patterns that lead to the murderer. Moreover, it is this self-directing aspect of Sleuth's investigation process that makes it powerful. Science illustrates more powerful self-directed capacity again, inasmuch as it is able to change both its general methods, including experimental procedure and theory construction, and its high order goals, by enriching its epistemic values, such as truth, with goals such as consistency, controllability, intelligibility, informativeness, etc.[2] Catcher, Sleuth and science respectively utilise increasingly powerful anticipative models (AMs), emulations of the course of interaction, which enrich their time window and hence their anticipative capacity.

In general, increases in self-directedness involve constructive interaction processes that generate more powerful AMs along with elaborated and integrated evaluative norms (ONs and INs), all of which the system can then use as directive constraints to guide and improve performance. Self-directedness grades in strength depending on the depth of anticipative and evaluative directedness and the plasticity available for process modification. The stronger the self-directedness of the process the more context-sensitive it can be. In the case of Sleuth the investigation procedure is context-sensitively self-directing whilst the overall goal (catch the suspect) and general methodology (profiling techniques) remain fixed. In the case of science high order methods and norms can be context-sensitively modified in addition to lower order investigation and there are powerful institutional processes that amplify such capacities. In this respect a noteworthy feature of self-directedness is the capacity for learning, and the discussion now turns to this issue.

---

[2] See Christensen and Hooker (1999a) for an initial interactivist-constructivist theory of science and critical evaluation of related models of evolutionary epistemology.

5 Self-directed anticipative learning: Constructing cognitive ability

Learning is an important component of self-directedness because powerful forms of context sensitivity require more information than a system might practicably possess before it has encountered its environment. Learning can have a variety of forms ranging from simple conditioning to the kinds of sophisticated constructive learning illustrated in Sleuth and science. The kind of learning processes that Sleuth and science display are themselves strongly shaped by the construction and modification of anticipation, and for this reason we term them *self-directed anticipative learning* (SDAL) processes. Because SDAL involves synergistic improvement in both interaction and anticipative competence it plays a fundamental role in the acquisition of cognitive ability.[3]

In an SDAL process the system learns about the nature of the problem as it tries to solve it. Cheetahs are not born knowing how to hunt gazelles, they have to acquire the necessary skills through practice. Likewise, Sleuth does not initially know who the murderer is or even what types of clues may reveal this – discovering these things is part of the investigation process itself. SDAL works through a positive feedback loop in which interaction generates information that improves the system's anticipation and thereby modifies the system's interaction processes, generating yet more refined information, and so on. As its anticipations improve the system gets better at recognising the conditions under which success and error occur, improving its ability to localise sources of success and error to particular aspects of the interaction process (cf. cheetah cubs learning to grip when climbing and stalk when hunting). In this way the solution, the specific method for achieving it, and in some cases the proper formulation of the goal itself, are all progressively acquired.

As discussed above, generating an AM improves the ability of the system to interact successfully. Thus, Catcher anticipates the future position of the ball, and in a similar but more powerful way a cheetah learns to anticipate the characteristic capacities of prey animals – e.g. their ability to detect the predator in varying conditions such as open grassland versus wooded area, during the day versus at night, and the prey's dynamical characteristics such as running speed and patterns of evasion, etc. Sleuth may learn that the murder fits the profile of a serial killer. The anticipatively modified interaction process is then evaluated by the system's existing ON/IN matrix. If the prey is too alert or fleet the cheetah remains hungry, if it is too large and dangerous the cheetah may be injured. The AM assists in the evaluation process as well by associating success/error events with interaction conditions, thus helping to localise them. As it is better able to anticipate the

[3] A detailed discussion of the organisational features of SDAL is provided in Christensen and Hooker (1998, 1999b), the present discussion is confined to a qualitative outline of SDAL.

dynamics of the chase, e.g., the cheetah learns an appropriate stalking distance for that kind of prey. Sleuth might discover that blue-haired grandmas can be deceptive or, through the arrest of an innocent person, that the forensic techniques employed were flawed. The result of this anticipative evaluation is further modification to the AM-guided interaction process.

The problem for the system is to use its AM to maintain or improve coherence between the interaction process and the ONs and INs that direct and evaluate the interaction process. That is, the system needs to learn to anticipatively interact with the environment in ways that tend to trigger success ONs and INs and avoid triggering failure ONs and INs. Maintaining or improving this interaction/evaluation coherence helps ensure that the change to the system's directive processes respects the closure constraints of the system and the nature of the environment. In some cases this results in adaptive change that respects the system's implicit closure constraints without making them explicit, as when a cheetah modifies its hunting technique to reinforce successful strategies without any explicit understanding of the reason for the improved success.

In some cases involving the construction of new goals, however, SDAL can enfold aspects of the system's implicit closure constraints into explicit process modulation capacity. Above we discussed examples in which a cheetah learns to focus on grip while climbing and on stalking as an important aspect of hunting. Another more sophisticated example is a young tennis player who employs a coach to improve her technique. The coach may observe that the player loses too many points at the net because of poor approach shots; the coach may then have the player practice her approach shot technique and teach her to only approach the net after a high quality approach shot. In this situation, a closure condition for effective net play is a good approach shot, but before the intervention of the coach the player was unaware of this condition. The coach, however, creates new goals for the player (hit good approach shots, go to the net after a good approach shot) that make the previously implicit closure condition explicit. Another example of the same type of phenomenon occurs when a person learns to select food on the basis of nutritional value as well as taste and capacity to satisfy hunger.

SDAL can also result in increased learning capacity in situations where the AM modifies the interaction process in a way that further improves the AM. If Sleuth's suspect profile is reasonably accurate it can allow Sleuth to recognise new kinds of clues, and thereby further improve the suspect profile. The richer the system's anticipative/normative structure is the more directed its learning can be, and the more potential there is that learning will improve the system's capacity to form AMs. For instance, an experimental scientist's AM concerns theories of what the instruments and system are doing, of proper system state preparation procedures, proper data collection and processing procedures, and so on; whence the detection of a procedural error (e.g. that electrical fields cannot be ignored) will issue in highly focused and detailed revisions to the AM which will in turn make equally sharpened empirical investigation of the whole situation possible. When learning success generates increased capacity for learning 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. Because SDAL results in synergistic skill acquisition, improvements in problem conception, and improved recognition of relevant information, there is reason to think it plays a key role in cognitive development. Indeed, this type of process is the essence of solving a divergent problem, where the correct problem definition, solution criteria and method are all progressively improved as the solution is arrived at. Solving such problems is characteristic of all basic life tasks.

6 Anticipative agency

This conception of a thoroughly anticipative, interactivist-constructivist (I-C) cognitive capacity extends to intentional agency generally, and we close this analysis with a brief introduction to developing an anticipative conception of agency.

Agents are entities which engage in normatively constrained, goal-directed, interaction with their environment. Intelligent agents have goals appropriate to their situation and interact with the environment in ways which adaptively achieve those goals. The core I-C conception of intentionality is focussed on the achievement of an adaptive interaction 'process flow'. Intentionality is measured by the capacity for fluid goal-directed management of interaction directed towards the achievement of system autonomy. It grades in strength in proportion to the *self*-directedness of the system's interaction capacity. Self-directed shaping of a complex interaction process flow will involve the full richness of the system's interaction window, which must be elaborated and integrated by the system during learning so as to bring about the simultaneous satisfaction of the system's many low and high-order norms (ONs and INs) that apply in the interaction circumstances. The result is a kind of 'maximum grip' that constitutes the achievement of skilled action (Merleau-Ponty 1962). In sum, the measure of intentional capacity is the extent to which the system can self-direct its interaction with the environment.

By contrast, the standard, information-processing conception of agency is dominated by a semantics and epistemics that is focused, respectively, on reference and truth. These notions focus attention on internal representational coherence and accuracy, and away from action, as the constitutive nature of agency. The essential nature of intentionality, e.g., becomes that of language-like reference (cf. Brentano 1960) rather than that of interactive effectiveness (cf. Merleau-Ponty). Referred to as *representationalism* or *cognitivism* in cognitive science, the assumption that the former is conceptually prior to the latter allows the theorist to abstract away from the details of dynamical interaction and treat intelligence and intentionality as functionally isolable capacities, respectively for problem solving and reference, from which adaptive interaction capacity follows trivially, or at least as a matter of mere engineering detail. Here we indicate how we would develop instead a thoroughly anticipative I-C conception of information and semantics.

The signal I to which a system is responding does not in itself have direct informational content for the system. Rather, the system exploits the signal as a source of order for organising its own processes. As such, the information value of I for the system is I's directive influence over **a**, the system's action in response. (Note equivalently that, while independently motivated, this conception of signal information is that which maximally resolves initial system uncertainty whether or not to do **a**, and so satisfies the Shannon/Weaver notion of information in signal transmission theory.) The major normative constraint on I's value for the system is the autonomy appropriateness of **a** in that circumstance. Then the meaning of I for the system is 'this is an **a**-appropriate action condition that will result in **a**'s closure conditions'.

Representational theories, by contrast, tie signal informational content to the situation originating the signal, which is (notoriously) ambiguous, and define success as accurate correspondence to the signal origin, which is (notoriously) not system-detectable. Such theories might be generically categorised as 'upstream, system-inaccessible' accounts of information content and contrasted with the 'downstream, system-accessible' account presented here. One of the key principles that an I-C approach is able to respect is the naturalist requirement that all constructed cognitive features be system constructible from system accessible interactions. While all natural systems have to make their own representations, meanings, etc., in all the standard theories this is impossible.[4]

Understanding the way information is exploited requires specifying the kind of interaction the signal has with the system's processes. Specifically, this leads to a distinction between *mere dynamical interaction, information utilisation* and *information processing*. A system S always interacts dynamically with an impinging signal I. If there are no significant process modulation effects then the interaction is merely dynamical. If S has a sufficiently differential dynamical response to I's features *and* uses this for downstream modulation of system processes, then S uses information in I and the constructed signal-information for

[4] Bickhard and Terveen 1995 provide a comprehensive critique of 'upstream, system-inaccessible' or encodingist theories of information and representation. Bickhard 1993 presents an alternative 'downstream, system-accessible' theory of representation compatible with the general position outlined here. Indeed, the position here owes much to these works.

S is S's downstream response to I. Finally, if S transforms I (perhaps by modulating it using other signals) to extract the information it is using for directive purposes, i.e. if a directed signal transformation is part of the constructed signal-information, then S processes the information. Information processing is thus a kind of higher order directedness – the system has processes for modulation of information. This conception is wider than the classical notion of formal computational information processing currently dominating cognitive science and artificial intelligence while yet being specified in a systematic, principled manner. It is currently standard to define information processing systems within a language-like framework as systems possessing syntactically structured symbolic representations whose transformations are constrained to syntax-respecting formal computations. From an I-C perspective these are recent, sophisticated arrivals, capturing only a small specialised part of even intelligent functioning, though because of their familiarity to us most theorists easily assume them as fundamental.

In this conception semantics starts by being contextual, vague and implicit, the domain of systems responding to relatively undifferentiated classes of signals in relatively un-contextualised reflex-like ways. As systems become increasingly self-directed, e.g. as their capacity for emulation and their performance norms are elaborated and integrated, they increasingly differentiate signals and process them for information, and so their semantics becomes increasingly articulated, explicit and trans-contextual. Explicit reference, and referential semantics, emerges as a sophisticated construction in this scheme.[5]

Epistemics begins as undifferentiated from general reward, i.e. from usefulness (to the agent, for its autonomy), both in evolution and in infant development, and this is still the

---

[5] We do not speculate what *kind* of construction explicit reference might be, but one congenial possibility argued for by Bickhard and Ritchie (1983) is that representations of objects are constituted as certain kinds of invariances of indications of interaction potential. Thus there are many systematically interrelated interactions one can have with a book – one can walk around it, open it, turn it upside down, place a coffee cup on it, and read it. Each interaction opens up and closes off further interactive possibilities, such that the set of possible interactions forms a multiply interconnected web. On Bickhard's account the representation of the book itself as an object is a kind of emergent high-order invariance in the indications of interactive possibilities. Campbell and Bickhard (1986) also contains a discussion of the self as implicit in these higher order ('meta') coordinations of directed interactions, for implicit in these aspects of coordination are presuppositions about the nature and competencies of the directing system itself, without their necessarily being made explicit for the system. Though developed independently, our own account is sympathetic to this approach, especially its emphasis on the implicit roots of selfhood in interaction shaping.

ultimate bedrock test for our constructed semantics. But as systems become increasingly self-directed they differentiate actuality from possibility (especially as their capacity for AM emulation is elaborated, cf. Bickhard and Ritchie 1983, Grush 1997), and as their surrogate performance norms are elaborated they increasingly differentiate usefulness from truth (cf. the Piagetian model in Hooker 1995, chapters 5, 6).

The same characteristics apply to the development of selfhood. In coordinating its autonomy constraints with the environment I-C anticipative management generates an increasingly rich normative perspective as appropriate norms are elaborated and interrelated through learning to yield an increasingly integrated normative perspective from which anticipative action takes place. This is the 'self' proper of self-directedness. The system that anticipatively steers itself through its environment to satisfy its own integrated normative framework, learning to improve both its performance and its evaluation of that performance as it goes, that system displays a distinctively intentional selfhood (see Christensen and Hooker 1998, 1999b). 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. This stands in contrast to presenting them as the distinct capacities for reference and problem solving, as standard cognitivism does, and provides for a much richer articulation of their interrelations than cognitivism can naturally provide.

7 Diagnosing Genuine Anticipation: Reflections on Rosen

The autonomous systems account of anticipatory systems thus provided offers an illuminating perspective from which to assess other accounts, in particular here the important treatment of Rosen 1985. Intuitively, Rosen considers a system S_2, the intended anticipatory system, that contains a model M of another system S_1, the intended anticipated system, and such that M's state brings about changes in some other non-M variables X of S_2, on the basis of which S_2 interacts with S_1. Rosen then sets down five necessary (and presumably jointly sufficient) conditions for S_2 to be genuinely anticipatory; simplified, these are: (1) M exists, (2) X exists, non-null, and X does not determine M (they are 'unlinked'), (3) M modulates X, (4) S_2, through X, interacts with S_1, (5) M is a predictive model of S_1 (modelled as M's state sequence being a time-compressed homomorphism of S_1's state sequence). There is in addition a general rationale offered for anticipation: (6) to render invariant some set of properties of S_1. As a first simple example of this schema, Rosen considers a general substrate-enzyme chain model of a biosynthetic pathway where a sequence of substrate transformations are each catalysed by their distinctive enzymes. In Rosen's example all reaction rates are constant except the n'th transform, from substrate A_{n-1} to substrate A_n, catalysed by enzyme E_n, where the first substrate A_0 also regulates E_n in such a way that A_{n-1} is held constant. Taking $S_1 = A_{n-1}$, $S_2 = A_0 \otimes E_n$, $M = A_0$, $X = E_n$, this model satisfies Rosen's five conditions and rationale (6).

It will be immediately clear that our general autonomous system model of an anticipatory system provided here also satisfies Rosen's five conditions. A cheetah (S_2) which is a successful hunter certainly is equipped with a predictive model (M) of its prey (S_1) and M keys its sensory-motor subsystems (X) so as to interact with the prey, namely to stalk, chase and kill it. (Normally an autonomous system will act on its environment, but note that it may also act on parts of itself, as when a wolf licks a wound to heal it or chews off its foot to escape a trap.) Yet, as we shall now show, ours is also a distinctively richer conception of anticipation, and in ways that are essential, in our view. That is, Rosen's five conditions are not jointly sufficient for genuine anticipation. Conversely, that extra content illuminates what we see as telling inadequacies in Rosen's treatment, including in his proffered rationale.

The essence of the differences between us is that we confine the notion of anticipation to autonomous systems. We do so because we believe that only these define a principled sense of *normative system-referenced* anticipation, i.e. where it is the *system itself* that is to be anticipatory, rather than anticipation deriving from some larger, suppressed context (see further below). Only autonomous systems have an explicit normative system perspective against which process is evaluated. Being genuinely anticipative, we contend, requires this normative perspective, or selfhood, against which the success or otherwise of anticipative states is evaluated. It is this context that provides a rationale for the system to form expectancies of future outcomes of actions, and the basis for their evaluation. To see the force of these considerations we return to re-consider Rosen's account.

The need for a system self-perspective clearly motivates Rosen's treatment. The rationale for requiring that X be non-null (condition 2), e.g., is that otherwise S_2 could not physically act upon S_1 based on its anticipations and, since M would be all of S_2, S_1 would effectively determine S_2; the nett effect being to reduce S_2 to a mere 'time shadower' of S_1. And the effect of the rationale (condition 6) is to provide a goal for S_2 to pursue in intervening anticipatively in S_1. All this suggests that Rosen intends to capture the basic conditions for a robust sense of agency in S_2's acting anticipatively, even if these conditions must also grade off to ones a simple system like a biosynthetic pathway can satisfy. But Rosen's formalism provides no means to introduce goals, or a system self-perspective, no matter how rudimentary. Why should S_2 care about S_1, or want to stabilise some property of S_1? S_2 is not equipped to have any perspective of its own. The substrate A_0, e.g., can neither evaluate its actions on either A_{n-1} or E_n, nor learn from them. In contrast, cheetahs do have a normative perspective of their own, and they don't intrinsically care about predicting their prey, they have to learn to anticipate prey behaviour only because of the valuable consequences (for themselves) of doing so. Without that self-perspective, S_2 is merely a brute 'predictive intervenor' in S_1, just as much without rationale as it would be if it were merely a brute time shadower.

This makes a crucial difference to the formulation of anticipation. A cheetah is born with its basic operational hunting norm, satiation (a surrogate for adequate nutrition), and its correlative basic goal, eating, and the result of its learning to hunt is that it develops new intermediate goals, such as stalking near enough before chasing. All these goals are there only because they serve to satisfy the basic satiation norm *of the cheetah*. This is the valuable consequence that is the reference for the cheetah's actions. Put in Rosen's terminology, the rationale for S_2's interaction with S_1 must lie in the consequences *for S_2*, not S_1, as Rosen suggests in his rationale, 6. Any consequences for S_1 will be derivative on these. Further, to do this requires that S_1 act upon S_2, as well as S_2 acting upon S_1, and in a way that informs both the anticipative modulation (now dodging left, catch achieved, eating begun) and the evaluative matrix (satiation increasing). This must of course have always been a condition for anticipation, otherwise the means to satisfy Rosen's condition 1 would remain a mystery; but, tellingly, Rosen's account ignores it. Now we see its importance. It is essentially the success and failure of anticipations (an action of S_1 on S_2) that drives learning. Without this feedback effect learning is impossible and the formation of anticipation mysterious. Anticipative action, and especially learning, can only be properly understood in terms of the modulation of the full interactive cycle, not just Rosen's initial action half.

Nor will it be the case that invariance of some low level property, even of S_2, must always be the outcome, let alone invariance of a property of S_1. As it happens one consequence of cheetah hunting is to maintain various cheetah properties, like blood sugar concentrations, relevantly invariant (more precisely between some tolerance levels) - but even here the variation over time of blood sugar is itself an important factor in the anticipative process. Ultimately it is only the very high level organisational property of autonomy itself that must remain invariant; a system might undergo all manner of changes as it develops, e.g. as it learns, and it can do so successfully so long only as autonomy is preserved (cf. our earlier critique of autopoiesis). And this may or may not call for any equivalent outcome for S_1. Cheetahs cannot anticipatively regulate prey populations, and insofar as they happen to do so (principally regulating genetic quality by culling the old, defective and sick) they precisely act as brute intervenors, in exactly the manner of Rosen's initial substrate, and do not act genuinely anticipatively.

It is not accidental, we suggest, though no less damaging in consequence, that it is these important features which are typically suppressed in the problem formulations of both anticipative control engineering and biological functional analysis. In the case of control engineering they are hidden in the human agency that poses the original engineering design problem. Perhaps a pulp and paper process needs to be continually optimised (an S_2 goal, though scarcely one requiring any simple S_1 invariances) and this requires predictive control. But this S_2 goal derives from a more basic one, the profitability to its owners of an optimised process, the valuable S_2 consequence really sought. This latter, however, is suppressed in the technical design problem the design engineer receives. Nonetheless, in utilising the predictive control design it is the *owners* who act anticipatively, not the operating control system, which is merely a predictive part of the anticipative machinery and *by itself* no more than a brute intervenor. A predictive, even a predictive regulating, element should not be mistaken for anticipation proper. Even to give the predictive controller the derivative title of anticipator is misleading because it is possible to single out such parts at best only in suitably modular systems. While simple engineering devices may show sufficient modularity, this is typically not the case in nature - nor, increasingly, in sophisticated engineering - where process multiplexing and multibonding characterise system organisation.

A similar obfuscation arises in functional biology when the systems context of a function is, often again for good practical reasons, taken for granted and suppressed. Rosen's initial substrate regulates enzymes, in whatever way it does, not because it is a genuinely anticipatory system but because it's doing so forms an essential part of a larger autonomous system. Here further confusion awaits; while this function was perhaps selected for, that does not make it even derivatively anticipatory for evolution, since evolution is not, cannot be, an anticipatory process (because its interactive and temporal structure precludes autonomy and hence coherent S_1 to S_2 feedback). The only place a genuine normative perspective occurs is for autonomous systems (where asynchronous processes do permit coherent S_1 to S_2 feedback to occur). But those autonomous systems which have Rosen's biosynthetic pathway as a cellular component likely no more than very indirectly evaluate its functioning (perhaps against general 'feeling well' norms) and only in company with many other contributing processes, so it is at best only weakly even derivatively anticipatory for them. Rather, it is more accurately specified as simply a predictive regulatory component among many components of their overall anticipatory organisation.

In sum, on our account Rosen's formulation does indeed capture some of the essence of genuine anticipation, and we have learned much here (as elsewhere) from him, but we consider that he moves to formalisation too fast and uncritically, missing the essential rationale of anticipation and some of its essential form.

8 Conclusion

This paper has presented a account of anticipation in autonomous systems as a central theme of an interactivist-constructivist paradigm for modelling adaptive intelligence. The theory of autonomy provides an account of norms in the form of the closure conditions required for successful self-generation, and an account of anticipation as an aspect of the relations between system actions and these closure conditions. This serves as the grounding point for modelling intelligence as the ability of the system to adaptively direct its interaction processes in complex variable conditions. More specifically, intelligence is understood as emerging through increasing self-directedness. Self-directed systems anticipate and evaluate the interaction flow, directively modifying the interaction process so as to achieve goals that regenerate or improve the system's autonomous closure conditions. Learning arises out of the drive to improve anticipation, that starts by being contextual, vague, and implicit, and becomes increasingly articulated and explicit as the system constructs anticipative models and goals for interaction. Cognitive development occurs through *self-directed anticipative learning* (SDAL), in which a pushme-pullyou effect is generated as increasingly rich anticipation increases the directedness of learning by improving error localisation, context recognition and the construction of improved anticipation.

References.

Bickhard, M.H.: 1992, 'Scaffolding and self-scaffolding: central aspects of development', in Winegar, I.T. and J. Valsiner (eds.): 1992, *Children's development within social contexts: Research and methodology.* (Hillsdale, NJ.: Lawrence Erlbaum.)

Bickhard, M.H.: 1993, 'Representational Content in Humans and Machines', *Journal of Experimental and Theoretical Artificial Intelligence* **5**: 285-333.

Bickhard, M. H., D. M. Ritchie: 1983, *On the nature of representation: A case study of James J. Gibson's theory of perception*. New York: Praeger.

Bickhard, M. H., L. Terveen.: 1995, *Foundational Issues in Artificial Intelligence and Cognitive Science — Impasse and Solution*, Amsterdam: Elsevier Scientific.

Brentano, F.: 1960, 'The Distinction Between Mental and Physical Phenomena', in R.M. Chisholm (ed.) *Realism and the Background of Phenomenology*, Glencoe, Il.: Free Press.

Campbell, R.L., M.H. Bickhard: 1986, *Knowing Levels and Developmental Stages*, Basel: Karger.

Christensen, W.D., C.A. Hooker: 1998, 'From Cell to Scientist: Toward an organisational theory of life and mind', in J. Bigelow (ed.): 1998, *Our Cultural Heritage*, Australian Academy of Humanities, University House, Canberra. Pp.275-326.

Christensen, W.D., C.A. Hooker: 1999a, 'The organisation of knowing: Autonomy, evolution, and evolutionary epistemology'. To appear in C. Heyes and D. Hull (eds.), *Festschrift for Donald Campbell*.

Christensen, W.D., C.A. Hooker: 1999b, 'An interactivist-constructivist approach to intelligence: self-directed anticipative learning'. To appear in *Philosophical Psychology*.

Collier, J.D., C.A. Hooker: 1999, 'Complexly Organised Dynamical Systems'. *Open Systems and Information Dynamics*, **36**, 1-62.

Clark, A.: 1997, *Being there: Putting brain, body and world together again*. Cambridge, MA: The MIT Press.

Grush, R.: 1997, 'The architecture of representation', *Philosophical Psychology,* **10** (1): 5-23.

Hooker, C.A.: 1995, *Reason, Regulation and Realism: Toward a Naturalistic, Regulatory Systems Theory of Reason*. Albany, N.Y.: State University of New York Press.

Maturana, H.R.: 1981, 'Autopoiesis', in M. Zeleny (ed.) *Autopoiesis: A theory of the living organization*. Boulder: Westview press.

Merleau-Ponty, M.: 1962, *Phenomenology of perception.* (transl. C. Smith.) London: Routledge & Kegan-Paul.

Mingers, J.: 1995, *Self-producing Systems: Implications and Applications of Autopoiesis.* London: Plenum.

Rosen, R.: 1985, *Anticipatory Systems: philosophical, mathematical, and methodological foundations*, New York: Pergamon.

Smithers, T.: 1995, 'Are Autonomous Agents Information Processing Systems?', L. Steels, R.A. Brooks (eds.): 1995, *The Artificial Life Route to 'Artificial Intelligence': Building Situated Embodied Agents*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Varela, F., H.R. Maturana, R.B. Uribe: 1974, 'Autopoiesis: The organization of living systems, its characterization and a model', *Biosystems* **5**: 187-196.