Chapter 2
The Different Paradigms of Cognition

There are many positions on cognition, each taking a significantly different stance on the nature of cognition, what a cognitive system does, and how a cognitive system should be analyzed and synthesized. Among these, however, we can discern two broad classes (see Figure 2.1): the *cognitivist* approach based on symbolic information processing representational systems, and the *emergent systems* approach, embracing connectionist systems, dynamical systems, and enactive systems, all based to a lesser or greater extent on principles of self-organization [228, 39].

Cognitivist approaches correspond to the classical and still common view that 'cognition is a type of computation' defined on symbolic representations, and that cognitive systems 'instantiate such representations physically as cognitive codes and ... their behaviour is a causal consequence of operations carried out on these codes' [174]. Connectionist, dynamical, and enactive systems, grouped together under the general heading of emergent systems, argue against the information processing view, a view that sees cognition as 'symbolic, rational, encapsulated, structured, and algorithmic', and argue in favour of a position that treats cognition as emergent, self-organizing, and dynamical [221, 105].

As we will see, the difference between the cognitivist and emergent positions are deep and fundamental, and go far beyond a simple distinction based on symbol manipulation. Without wishing to preempt what is to follow, we can contrast the cognitivist and emergent paradigms on fourteen distinct characteristics:[1] computational operation, representational framework, semantic grounding, temporal constraints, inter-agent epistemology, embodiment, perception, action, anticipation, adaptation, motivation, autonomy, cognition, and philosophical foundation. Let us look briefly at each of these in turn (see Table 2.1 for a synopsis of the key issues).

Computational Operation

Cognitivist systems use rule-based manipulation of symbol tokens, typically but not necessarily in a sequential manner.

[1] These fourteen characteristics are based on the twelve proposed by [232] and augmented here by adding two more: the role of cognition and the underlying philosophy. The subsequent discussion is also an extended adaptation of the commentary in [232].

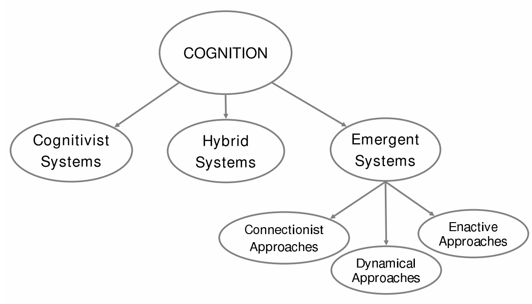

Fig. 2.1 The cognitivist, emergent, and hybrid paradigms of cognition.

Emergent systems exploit processes of self-organization, self-production, self-maintenance, and self-development, through the concurrent interaction of a network of distributed interacting components.

Representational Framework

Cognitivist systems use patterns of symbol tokens that refer to events in the external world. These are typically the descriptive[2] product of a human designer and are usually, but not necessarily, punctate rather than distributed.

Emergent systems representations are global system states encoded in the dynamic organization of the system's distributed network of components.

Semantic Grounding

Cognitivist systems symbolic representations are grounded through percept-symbol identification by either the designer or by learned association. These representations are accessible to direct human interpretation.

Emergent systems ground representations by autonomy-preserving anticipatory and adaptive skill construction. These representations only have meaning insofar as they contribute to the continued viability of the system and are inaccessible to direct human interpretation.

Temporal Constraints

Cognitivist systems are atemporal and are not necessarily entrained by the events in the external world.

Emergent systems are entrained and operate synchronously in real-time with events in its environment.

Inter-agent Epistemology

For cognitivist systems an absolute shared epistemology between agents is guaranteed by virtue of their positivist view of reality; that is, each agent is embedded in an environment, the structure and semantics of which are independent of the system's cognition.

The epistemology of emergent systems is the subjective agent-specific outcome

[2] Descriptive in the sense that the designer is a third-party observer of the relationship between a cognitive system and its environment so that the representational framework is how the designer sees the relationship.

of a history of shared consensual experiences among phylogentically-compatible agents.

Embodiment

Cognitivist systems do not need to be embodied, in principle, by virtue of their roots in functionalism (which holds that cognition is independent of the physical platform in which it is implemented [60]).

Emergent systems are intrinsically embodied and the physical instantiation plays a direct constitutive role in the cognitive process [229, 112, 64].

Perception

In cognitivist systems, perception provides an interface between the absolute external world and the symbolic representation of that world. The role of perception os to abstract faithful spatio-temporal representations of the external world from sensory data.

In emergent systems, perception is an agent-specific interpretation of perturbations of the system by the environment.

Action

In cognitivist systems, actions are causal consequences of symbolic processing of internal representations.

In emergent systems, actions are perturbations of the environment by the system, typically to maintain the viability of the system.

Anticipation

In cognitivist systems, anticipation typically takes the form of planning using some form of procedural or probabilistic reasoning with some *a priori* model.

Anticipation in the emergent paradigm requires the system to visit a number of states in its self-constructed perception-action state space without commiting to the associated actions.

Adaptation

For cognitivism, adaptation ususally implies the acquisition of new knowledge.

In emergent systems, adaptation entails a structural alteration or re-organization to effect a new set of dynamics. Adaptation can take the from of either learning or development.

Motivation

In cognitivist systems, motives provide the criteria which are used to select the goal to adopt and the associated actions.

In emergent systems, motives encapsulate the implicit value system that modulate the system dynamics of self-maintenance and self-development, impinging on perception (through attention), action (through action selection), and adaptation (through the mechanisms that govern change), such as enlarging the space of viable interaction.

Autonomy

Autonomy[3] The cognitivist paradigm does not necessarily entail autonomy. The

---

[3] There are many possible defi nitions of autonomy, ranging from the ability of a system to contribute to its own persistence [19] through to the self-maintaining organizational characteristic of living creatures —dissipative far-from equilibrium systems —that enables them to use their own

The Cognitivist Paradigm *vs.* the Emergent Paradigm		
Characteristic	**Cognitivist**	**Emergent**
Computational Operation	Syntactic manipulation of symbols	Concurrent self-organization of a network
Representational Framework	Patterns of symbol tokens	Global system states
Semantic Grounding	Percept-symbol association	Skill construction
Temporal Constraints	Atemporal	Synchronous real-time entrainment
Inter-agent epistemology	Agent-independent	Agent-dependent
Embodiment	No role implied: functionalist	Direct constitutive role: non-functionalist
Perception	Abstract symbolic representations	Perturbation by the environment
Action	Causal consequence of symbol manipulation	Perturbation by the system
Anticipation	Procedural or probabilistic reasoning	Traverse of perception-action state space
Adaptation	Learn new knowledge	Develop new dynamics
Motivation	Criteria for goal selection	Increase space of interaction
Autonomy	Not entailed	Cognition entails autonomy
Cognition	Rational goal-achievement	Self-maintenance adn self-development
Philosophical Foundation	Positivism	Phenomenology

Table 2.1 A comparison of cognitivist and emergent paradigms of cognition; refer to the text for a full explanation (adapted from [232] and extended).

emergent paradigm does since cognition is the process whereby an autonomous system becomes viable and effective.

Cognition

In the cognitivist paradigm, cognition is the rational process by which goals are achieved by reasoning with symbolic knowledge representations of the world in which the agent operates.

In the emergent paradigm, cognition is the dynamic process by which the system acts to maintain its identity and organizational coherence in the face of environmental perturbation. Cognition entails system development to improve its anticipatory capabilities and extend it space of autonomy-preserving actions.

Philosophical Foundations

The cognitivist paradigm is grounded in positivism [60].

The emergent paradigm is grounded in phenomenology [62, 231].

The sections that follow discuss the cognitivist and emergent paradigms, as well as hybrid approaches, and draw out each of these issues in more depth.

capacities to manage their interactions with the world, and with themselves, in order to remain viable [36].

2.1 Cognitivist Models

2.1.1 An Overview of Cognitivist Models

Cognitive science has its origins in cybernetics (1943-53) in the first efforts to formalize what had up to that point been metaphysical treatments of cognition [228]. The intention of the early cyberneticians was to create a science of mind, based on logic. Examples of progenitors include McCulloch and Pitts and their seminal paper ‘A logical calculus immanent in nervous activity’ [140]. This initial wave in the development of a science of cognition was followed in 1956 by the development of an approach referred to as *cognitivism*. Cognitivism holds that cognition involves computations defined over internal representations *qua* knowledge, in a process whereby information about the world is abstracted by perception, and represented using some appropriate symbolic data-structure, reasoned about, and then used to plan and act in the world. The approach has also been labelled by many as the *information processing* (or symbol manipulation) approach to cognition [129, 161, 84, 172, 107, 228, 221, 105]

Cognitivism has undoubtedly been the predominant approach to cognition to date and is still prevalent. The discipline of cognitive science is often identified with this particular approach [105, 60]: However, as we will see, it is by no means the only paradigm in cognitive science and there are indications that the discipline is migrating away from its stronger interpretations [39].

For cognitivist systems, cognition is representational in a strong and particular sense: it entails the manipulation of explicit symbolic representations of the state and behaviour of the external world to facilitate appropriate, adaptive, anticipatory, and effective interaction, and the storage of the knowledge gained from this experience to reason even more effectively in the future [94]. Perception is concerned with the abstraction of faithful spatio-temporal representations of the external world from sensory data. Reasoning itself is symbolic: a procedural process whereby explicit representations of an external world are manipulated to infer likely changes in the configuration of the world (and attendant perception of that altered configuration) arising from causal actions.

In most cognitivist approaches concerned with the creation of artificial cognitive systems, the symbolic representations (or representational frameworks, in the case of systems that are capable of learning) are the descriptive product of a human designer. This is significant because it means that they can be directly accessed and understood or interpreted by humans and that semantic knowledge can be embedded directly into and extracted directly from the system. However, it has been argued that this is also the key limiting factor of cognitivist vision systems: these programmer-dependent representations effectively bias the system (or ‘blind’ the system [243]) and constrain it to an idealized description that is dependent on and a consequence of the cognitive requirements of human activity. This approach works as long as the system doesn’t have to stray too far from the conditions under which these descriptions were formulated. The further one does stray, the larger the ‘semantic gap’

[208] between perception and possible interpretation, a gap that is normally plugged by the embedding of (even more) programmer knowledge or the enforcement of expectation-driven constraints [166] to render a system practicable in a given space of problems.

Cognitivism makes the positivist assumption that 'the world we perceive is isomorphic with our perceptions of it as a geometric environment' [203]. In cognitivism, the goal of cognition is to reason symbolically about these representations in order to effect the required adaptive, anticipatory, goal-directed, behaviour. Typically, this approach to cognition will deploy an arsenal of techniques including machine learning, probabilistic modelling, and other techniques in an attempt to deal with the inherently uncertain, time-varying, and incomplete nature of the sensory data that is being used to drive this representational framework. However, this doesn't alter the fact that the representational structure is still predicated on the descriptions of the designers. The significance of this will become apparent in later sections.

2.1.2 Cognitivism and Artificial Intelligence

Since cognitivism and artificial intelligence research have very strong links,[4] it is worth spending some time considering the relationship between cognitivist approaches and classical artificial intelligence, specifically the Newell's and Simon's 'Physical Symbol System' approach to artificial intelligence [161] which has been extraordinarily influential in shaping how we think about intelligence, natural as well as computational.

In Newell's and Simon's 1976 paper, two hypotheses are presented:

1. *The Physical Symbol System Hypothesis:* A physical symbol system has the necessary and sufficient means for general intelligent action.
2. *Heuristic Search Hypothesis*. The solutions to problems are represented as symbol structures. A physical-symbol system exercises its intelligence in problem-solving by search, that is, by generating and progressively modifying symbol structures until it produces a solution structure.

The first hypothesis implies that any system that exhibits general intelligence is a physical symbol system *and* any physical symbol system of sufficient size can be configured somehow ('organized further') to exhibit general intelligence.

The second hypothesis amounts to an assertion that symbol systems solve problems by heuristic search, *i.e.* 'successive generation of potential solution structures' in an effective and efficient manner. 'The task of intelligence, then, is is to avert the ever-present threat of the exponential explosion of search'.

---

[4] Some view AI as the direct descendent of cognitivism: "... the positivist and reductionist study of the mind gained an extraordinary popularity through a relatively recent doctrine called *Cognitivism*, a view that shaped the creation of a new field —*Cognitive Science* —and its most hard core offspring: Artificial Intelligence" (emphasis in the original). [60]

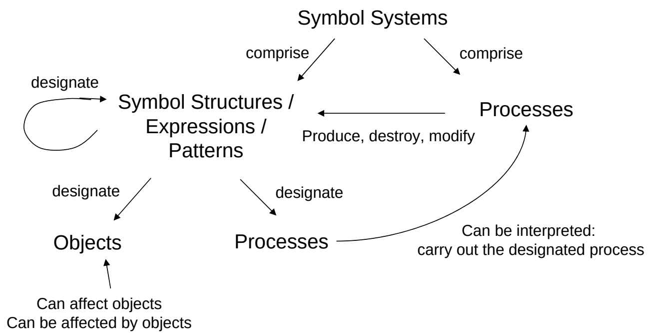

Fig. 2.2 The essence of a physical symbol system [161].

A physical symbol system is equivalent to an automatic formal system [84]. It is 'a machine that produces through time an evolving collection of symbol structures.' A symbol is a physical pattern that can occur as a component of another type of entity called an expression (or symbol structure): expressions/symbol structures are arrangements of symbols/tokens. As well as the symbol structures, the system also comprises processes that operate on expressions to produce other expressions: 'processes of creation, modification, reproduction, and destruction'. An expression can *designate* an object and thereby the system can either 'affect the object itself or behave in ways depending on the object', or, if the expression designates a process, then the system *interprets* the expression by carrying out the process (see Figure 2.2).

In the words of Newell and Simon,

> 'Symbol systems are collections of patterns and processes, the latter being capable of producing, destroying, and modifying the former. The most important properties of patterns is that they can designate objects, processes, or other patterns, and that when they designate processes, they can be interpreted. Interpretation means carrying out the designated process. The two most signifi cant classes of symbol systems with which we are acquainted are human beings and computers.'

What is important about this explanation of a symbol system is that it is more general than the usual portrayal of symbol-manipulation systems in which symbols designate only objects, in which case we have a system of processes that produces, destroys, and modifies symbols, and no more. Newell's and Simon's original view is more sophisticated. There are two recursive aspects to it: processes can produce processes, and patterns can designate patterns (which, of course, can be processes). These two recursive loops are closely linked. Not only can the system build ever more abstract representations and reason about those representation, but it can modify itself as a function both of its processing, *qua* current state/structure, and of its representations.

Symbol systems can be instantiated and the behaviour of these instantiated systems depend on the the details of the symbol system, its symbols, operations, and interpretations, and *not* on the particular form of the instantiation.

The *physical symbol system hypothesis* asserts that a physical symbol system has the necessary and sufficient means for general intelligence. From what we have just said about symbol systems, it follows that intelligent systems, either natural or artificial ones, are effectively equivalent because the instantiation is actually inconsequential, at least in principle.

To a very great extent, cognitivist systems are identically physical symbol systems.

Later, Allen Newell [160] defines intelligence as the degree to which a system approximates [the ideal] of a knowledge-level system. A knowledge-level system is one which can bring to bear *all* its knowledge onto *every* problem it attempts to solve (or, equivalently, every goal it attempts to achieve). Perfect intelligence implies complete utilization of knowledge. It brings this knowledge to bear according to the *principle of maximum rationality* which was proposed by Newell in 1982 [159] as follows: 'If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action'. Anderson [2] later offered a different principle, the *principle of rationality*, sometimes referred to as rational analysis, stated as follows: 'the cognitive system optimizes the adaptation of the behaviour of the organism'. Note that Anderson's principle considers optimality to be necessary for rationality, something that Newell's principle doesn't.

The knowledge in such an artificial intelligence system, *i.e.* in a knowledge-level system, is represented by symbols. Symbols are abstract entities that may be instantiated and manipulated as 'tokens'. Newell characterizes a symbol system as follows [244]. It has:

- *Memory* to contain the symbolic information;
- *Symbols* to provide a pattern to match or index other symbolic information;
- *Operations* to manipulate symbols;
- *Interpretations* to allow symbols to specify operations;
- *Capacities* for *composability*, so that the operators may produce any symbol structure; for *interpretability*, so that the symbol structures are able to encode any meaningful arrangement of operations; and sufficient *memory* to facilitate both of the above.

Newell suggests a progression of four bands, from biological, to cognitive, to rational, to social. These bands are in fact the different levels of Newell's quintessential information processing cognitivist approach. According to Newell, each band is also characterized by a typical execution time:

- Biological: $10^{-4} - 10^{-2}$ seconds
- Cognitive: $10^{-1} - 10^{1}$ seconds
- Rational: $10^{2} - 10^{4}$ seconds
- Social: $10^{5} - 10^{7}$ seconds

Note that the restriction of biological processing to the millisecond timeframe in Newell's framework is somewhat inappropriate since it is clear that biological processes clearly extend over much greater timescales, as one can see from the brain's ability to recover from major damage and also its continuous development over the entire ontogenetic timeframe.

The biological band corresponds to the neurophysiological make-up of the system. Newell identifies three layers in this band: the organelle, the neuron, and the neural circuit. Connectionist systems are often focussed exclusively on this band.

In Newell's framework, the cognitive band corresponds to the symbol level and its physical instantiation as a concrete architecture. Newell identifies three layers:

1. deliberate acts such as reaching that take a very short amount of time, typically 100ms,
2. 'composed operations' such as shifting gear when driving that take on the order of a second,
3. actions that take up to ten seconds, such as steering a car through an entrance.

The rational band is concerned with actions that are typically characterized by tasks and typically require some reasoning. For example, the task navigating your way home. This is the knowledge level.

The social band extends activity to behaviours that occupy hours, days, or weeks, often involving interaction with other agents.

All knowledge is represented (symbolically) at the symbol level. All knowledge-level systems contain a symbol system. This is the strong interpretation of the physical symbol system hypothesis: not only is a physical symbol system sufficient for general intelligence, it is also necessary for intelligence.

2.1.3 Some Cognitivist Systems

Although we will survey cognitivist systems from an architectural point of view in Chapter 6, we mention here a sample of cognitivist systems to provide a preliminary impression of the approach.

The use of explict symbolic knowledge has been used in many cognitivist systems, *e.g.* a cognitive vision system [156] developed for the interpretation of video sequences of traffic behaviour and the generation of a natural language description of the observed environment. It proceeds from signal representations to symbolic representations through several layers of processing, ultimately representing vehicle behaviour with situation graph trees (SGT). Automatic interpretation of this representation of behaviour is effected by translating the SGT into a logic program (based on fuzzy metric temporal Horn logic). See also [9, 8, 10, 69, 68] for related work.

The cognitivist assumptions are also reflected well in the model-based approach described in [158, 152] which uses Description Logics, based on First Order Predicate Logic, to represent and reason about high-level concepts such as spatio-temporal object configurations and events.

Probabilistic frameworks have been proposed as an alternative (or sometimes an adjunct [158]) to these types of deterministic reasoning systems. For example, Buxton *et al.* describe a cognitive vision system for interpreting the activities of expert human operators. It exploits dynamic decision networks (DDN) — an extension of Bayesian belief networks to incorporate dynamic dependencies and utility theory [29] — for recognizing and reasoning about activities, and both time delay radial basis function networks (TDRBFN) and hidden markov models (HMM) for recognition of gestures. Although this system does incorporate learning to create the gesture models, the overall symbolic reasoning process, albeit a probabilistic one, still requires the system designer to identify the contextual constraints and their causal dependencies (for the present at least: on-going research is directed at automatically learning the task-based context dependent control strategies) [189, 30, 31].[5] Recent progress in autonomously constructing and using symbolic models of behaviour from sensory input using inductive logic programming is reported in [41].

The dependence of cognitivist approaches on designer-oriented world-representations is also well exemplified by knowledge-based systems such as those based on ontologies. For example, Maillot *et al.* [128] describe a framework for an ontology-based cognitive vision system which focusses on mapping between domain knowledge and image processing knowledge using a visual concept ontology incorporating spatio-temporal, textural, and colour concepts.

Another architecture for a cognitive vision system is described in [34], This system comprises a sub-symbolic level, exploiting a viewer-centred $2\frac{1}{2}$D representation based on sensory data, an intermediate pre-linguistic conceptual level based on object-centred 3D superquadric representations, and a linguistic level which uses a symbolic knowledge base. An attentional process links the conceptual and linguistic level.

An adaptable system architecture for observation and interpretation of human activity that dynamically configures its processing to deal with the context in which it is operating is decribed in [44] while a cognitive vision system for autonomous control of cars is described in [49].

Town and Sinclair present a cognitive framework that combines low-level processing (motion estimation, edge tracking, region classification, face detection, shape models, perceputal grouping operators) with high-level processing using a language-based ontology and adaptive Bayesian networks. The system is self-referential in the sense that it maintains an internal representation of its goals and current hypotheses. Visual inference can then be performed by processing sentence structures in this ontological language. It adopts a quintessentially cognitivist symbolic representationalist approach, albeit that it uses probabilistic models, since it requires that a designer identify the "right structural assumptions" and prior probability distributions.

[5] See [29] for a survey of probabilistic generative models for learning and understanding activities in dynamic scenes.

2.2 Emergent Approaches

Emergent approaches take a very different view of cognition. Here, cognition is the process whereby an autonomous system becomes viable and effective in its environment. It does so through a process of self-organization through which the system is continually re-constituting itself in real-time to maintain its operational identity through moderation of mutual system-environment interaction and co-determination [133]. Co-determination implies that the cognitive agent is specified by its environment and at the same time that the cognitive process determines what is real or meaningful for the agent. In a sense, co-determination means that the agent constructs its reality (its world) as a result of its operation in that world. In this context, cognitive behaviour is sometimes defined as the automatic induction of an ontology: such an ontology will be inherently specific to the embodiment and dependent on the system's history of interactions, *i.e.*, its experiences. Thus, for emergent approaches, perception is concerned with the acquisition of sensory data in order to enable effective action [133] and is dependent on the richness of the action interface [78]. It is not a process whereby the structure of an absolute external environment is abstracted and represented in a more or less isomorphic manner.

Sandini *et al.* have argued that cognition is also the complement of perception [191]. Perception deals with the immediate and cognition deals with longer timeframes. Thus cognition reflects the mechanism by which an agent compensates for the immediate nature of perception and can therefore adapt to and anticipate environmental action that occurs over much longer timescales. That is, cognition is intrinsically linked with the ability of an agent to act prospectively: to operate in the future and deal with what might be, not just what is.

In contrast to the cognitivist approach, many emergent approaches assert that the primary model for cognitive learning is anticipative skill construction rather than knowledge acquisition and that processes which both guide action and improve the capacity to guide action while doing so are taken to be the root capacity for all intelligent systems [36]. While cognitivism entails a self-contained abstract model that is disembodied in principle, the physical instantiation of the systems plays no part in the model of cognition [229, 230]. In contrast, emergent approaches are intrinsically embodied and the physical instantiation plays a pivotal role in cognition. They are neither functionalist nor positivist.

2.2.1 Connectionist Systems

Connectionist systems rely on parallel processing of non-symbolic distributed activation patterns using statistical properties, rather than logical rules, to process information and achieve effective behaviour [141]. In this sense, the neural network instatiations of the connectionist model 'are dynamical systems which compute functions that best capture the statistical regularities in training data' [210].

A comprehensive review of connectionism is beyond the scope of this course. For an overview of the foundation of the field and a selection of seminal papers on connectionism, see Anderson's and Rosenfeld's *Neurocomputing: Foundations of Research* [3] and *Neurocomputing 2: Directions of Research* [4]. Medler provides a succinct survey of the development of connectionism in [141], while Smolensky reviews the field from a mathematical perspective, addressing computational, dynamical, and statistical issues [210, 211, 212, 213]. Arbib's *Handbook of Brain Theory and Neural Networks* provides very accessible summaries of much of the relevant literature [7].

The roots of connectionism reach back well before the computational era. Although Feldman and Ballard [54] are normally credited with the introduction of the term 'connectionist models' in 1982, the term connectionism has been used as early as 1932 in psychology by Thorndike [222, 223] to signal an expanded form of associationism based, for example, on the connectionist principles clearly evident in William James' model of associative memory,[6] but also anticipating such mechanisms as Hebbian learning. In fact, the introduction to Hebb's book *The Organization of Behaviour* [87], in which he presents an unsupervised neural training algorithm whereby the synaptic strength is increased if both the source and target neurons are active at the same time, contains one of the first usages of the term connectionism [3], p. 43.

We have already noted that cognitivism has some of its roots in earlier work in cognitive science and in McCulloch and Pitts seminal work in particular [140]. McCulloch and Pitts showed that any statement within propositional logic could be represented by a network of simple processing units and, furthermore, that such nets have, in principle, the computational power of a Universal Turing Machine. Depending on how you read this equivalence, McCulloch and Pitts contributed to the foundation of both cognitivism and connectionism.

The connectionist approach was advanced significantly in the late 1950s with the introduction of Rosenblatt's *perceptron* [183] and Selfridge's *Pandemonium* model of learning [195]. Rosenblatt showed that any pattern classification problem expressed in binary notation can be solved by a perceptron network. Although network learning advanced in 1960 with the introduction of the Widrow-Hoff rule, or delta rule, for supervised training in the *Adeline* neural model [241], the problem with perceptron networks was that no learning algorithm existed to allow the adjustment of the weights of the connections between input units and hidden associative units. Consequently, perceptron networks were effectively single-layer networks since learning algorithms could only adjust the connection strength between the hidden units and the output units, the weights governing the connection strength between input and hidden units being fixed by design.

In 1969, Minsky and Papert [151] showed that these perceptrons can only be trained to solve linearly separable problems and couldn't be trained to solve more general problems. As a result, research on neural networks and connectionist models suffered.

---

[6] Anderson's and Rosenfeld's collection of seminal papers on neurocomputing [3] opens with Chapter XVI 'Association' from William James' 1890 Psychology, Briefer Course [99].

With the apparent limitations of perceptions clouding work on network learning, research focussed more on memory and information retrieval and, in particular, on parallel models of associative memory (*e.g.* see [88]). Landmark contributions in this period include McClelland's Interactive Activation and Competition (IAC) model [136] which introduced the idea of competitive pools of mutually-inhibitory neurons and demonstrated the ability of connectionist systems to retrieve specific and general information from stored knowledge about specific instances.

During this period too alternative connectionist models were being put forward in, for example, Grossberg's Adaptive Resonance Theory (ART) [82] and Kohonen's self-organizing maps (SOM) [109], ofter referred to simply as Kohonen networks. ART, introduced in 1976, has evolved and expanded considerably in the past 30 years to address real-time supervised and unsupervised category learning, pattern classification, and prediction (see [33] for a summary). Kohonen networks produce topological maps in which proximate points in the input space are mapped by an unsupervised self-organizing learning process to an internal network state which preserves this topology: that is, input points (points in pattern space) which are close together are represented in the mapping by points (in weight space) which are close together. Once the unsupervised self-organization is complete, the Kohonen network can be used as either an auto-associative memory or a pattern classifier.

Perceptron-like neural networks underwent a resurgence in the mid 1980s with the development of the parallel distributed processing (PDP) architecture [188] in general and with the introduction by Rumelhart, Hinton, and Williams of the back-propagation algorithm [186, 187]. The back-propagation learning algorithm, also known as the generalized delta rule or GDR as it is an generalization of the Widrow-Hoff delta rule for training Adaline units, overcame the limitation cited by Minsky and Papert by allowing the connections weights between the input units and the hidden units be modified, thereby enabling multi-layer perceptrons to *learn* solutions to problems that are not linearly separable. Although the back-propagation learning rule made its great impact through the work of Rumelhart *et al.*, it had previously been derived independently by Werbos [240], among others [141].

In cognitive science, PDP made a significant contribution to the move away from the sequential view of computational models of mind, towards a view of concurrently-operating networks of mutually-cooperating and competing units, and also in raising an awareness of the importance of the structure of the computing system on the computation.

The standard PDP model represents a static mapping between the input vectors as a consequence of the feed-forward configuration. On the other hand, recurrent networks which have connections that loop back to form circuits, *i.e.* networks in which either the output or the hidden units' activations signals are fed back to the network as inputs, exhibit dynamic behaviour.[7] Perhaps the best known type of recurrent network is the Hopfield net [95]. Hopfield nets are fully recurrent networks

[7] This recurrent feed-back has nothing to do with the feed-back of error signals by, for example, back-propagation to effect weight adjustment during learning

that act as auto-associative memory[8] or content-addressable memory that can effect pattern completion. Other recurrent networks include Elman nets [53] (with recurrent connections from the hidden to the input units) and Jordan nets [102] (with recurrent connections from the output to the input units). Boltzman machines [89] are variants of Hopfield nets that use stochastic rather than deterministic weight update procedures to avoid problems with the network becoming trapped in local minima during learning.

Multi-layer perceptrons and other PDP connectionist networks typically use monotonic functions, such as hard-limiting threshold functions or sigmoid functions, to activate neurons. The use of non-monotonic activation functions, such as the Gaussian function, can offer computational advantages, *e.g.* faster and more reliable convergence on problems that are not linearly separable.

Radial basis function (RBF) networks [153] also use Gaussian functions but differ from multi-layer perceptrons in that the Gaussian function is used only for the hidden layer, with the input and output layers using linear activation functions.

Connectionist systems continue to have a strong influence on cognitive science, either in a strictly PDP sense such as McClelland's and Rogers' PDP approach to semantic cognition [138]) or in the guise of hybrid systems such as Smolensky's and Legendre's connectionist/symbolic computational architecture for cognition [214, 209].

One of the original motivations for work on emergent systems was disaffection with the sequential, atemporal, and localized character of symbol-manipulation based cognitivism [228]. Emergent systems, on the other hand, depend on parallel, real-time, and distributed architectures. Of itself, however, this shift in emphasis isn't sufficient to constitute a new paradigm and, as we have seen, there are several other pivotal characteristics of emergent systems. Indeed, Freeman and Núñez have argued that more recent systems — what they term neo-cognitivist systems — exploit parallel and distributed computing in the form of artificial neural networks and associative memories but, nonetheless, still adhere to the original cognitivist assumptions [60]. A similar point was made by Van Gelder and Port [67]. We discuss these hybrid systems in Section 2.3.

One of the key features of emergent systems, in general, and connectionism, in particular, is that 'the system's connectivity becomes inseparable *from its history of transformations*, and related to the kind of task defined for the system' [228]. Furthermore, symbols play no role.[9] Whereas in the cognitivist approach the symbols are distinct from what they stand for, in the connectionist approach, "meaning relates to the global state of the system" [228]. Indeed, meaning is something attributed by an external third-party observer to the correspondence of a system state with that of the world in which the emergent system is embedded. Meaning is a description attributed by an outside agent: it is not something that is intrinsic to the cognitive

[8] Hetero-associative memory —or simply associative memory —produces an output vector that is different from the input vector

[9] It would be more accurate to say that symbols should play no role since it has been noted that connectionist systems often fall back in the cognitivist paradigm by treating neural weights as a distributed symbolic representation [67].

system except in the sense that the dynamics of the system reflect the effectiveness of its ability to interact with the world.

Examples of the application of associative learning systems in robotics can be found in [101, 142] where hand-eye coordination is learned by a Kohonen neural network from the association of proprioceptive and exteroceptive stimuli. As well as attempting to model cognitive behaviour, connectionist systems can self-organize to produce feature-analyzing capabilities similar to those of the first few processing stages of the mammalian visual system (*e.g.* centre-surround cells and orientation-selective cells) [127]. An example of a connectionist system which exploits the co-dependency of perception and action in a developmental setting can be found in [146]. This is a biologically-motivated system that learns goal-directed reaching using colour-segmented images derived from a retina-like log-polar sensor camera. The system adopts a developmental approach: beginning with innate inbuilt primitive reflexes, it learns sensorimotor coordination. Radial basis function networks have also been used in cognitive vision systems, for example, to accomplish face detection [30].

2.2.2 Dynamical Systems

Dynamical systems theory is very general and can be deployed to model many different types of systems in such diverse areas as biology, astronomy, ecology, economonics, physics, and many more. It has been used to complement classical approaches in artificial intelligence [177] and it has also been deployed to model natural and artificial cognitive systems [105, 221, 67]. Advocates of the dynamical systems approach to cognition argue that motoric and perceptual systems are both dynamical systems, each of which self-organizes into meta-stable patterns of behaviour.

A dynamical system defines a particular pattern of behaviour. The system is characterized by a state vector $\mathbf{q}$ and its time derivative $\dot{\mathbf{q}}$ is a function of the state vector, control parameters $\mathbf{p}$ and noise n. It is a self-organizing system because the system dynamics are defined by, and only by, the system state $\dot{\mathbf{q}} = \mathbf{N}(\mathbf{q}, \mathbf{p}, n)$.

In general, a dynamical system is an open dissipative non-linear non-equilibrium system: a system in the sense of a large number of interacting components with large number of degrees of freedom, dissipative in the sense that it diffuses energy (its phase space decreases in volume with time implying preferential sub-spaces), non-equilibrium in the sense that it is unable to maintain structure or function without external sources of energy, material, information (and, hence, open). The non-linearity is crucial: as well as providing for complex behaviour, it means that the dissipation is not uniform and that only a small number of the system's degrees of freedom contribute to its behaviour. These are termed *order parameters* (or *collective variables*). Each order parameter defines the evolution of the system, leading to meta-stable states in a multi-stable state space (or phase space). It is this ability to characterize the behaviour of a high-dimensional system with a low-dimensional model that is one of the features that distinguishes dynamical systems from connectionist systems [105].

Certain conditions must prevail before a system qualifies as a cognitive dynamical system. The components of the system must be related and interact with one another: any change in one component or aspect of the system must be dependent on and only on the states of the other components: 'they must be interactive and self contained' [67]. As we will see shortly, this is very reminiscent of the requirement for operational closure in enactive systems, the topic of the next section.

Proponents of dynamical systems point to the fact that they provide one directly with many of the characteristics inherent in natural cognitive systems such as multistability, adaptability, pattern formation and recognition, intentionality, and learning. These are achieved purely as a function of dynamical laws and consequent self-organization. They require no recourse to symbolic representations, especially those that are the result of human design.

However, Clark [39] has pointed out that the antipathy which proponents of dynamical systems approaches display toward cognitivist approaches rests on rather weak ground insofar as the scenarios they use to support their own case are not ones that require higher level reasoning: they are not 'representation hungry' and, therefore, are not well suited to be used in a general anti-representationalist (or anti-cognitivist) argument. At the same time, Clark also notes that this antipathy is actually less focussed on representations *per se* (dynamical systems readily admit internal states that can be construed as representations) but more on objectivist representations which form an isomorphic symbolic surrogate of an absolute external reality.

It has been argued that dynamical systems allow for the development of higher order cognitive functions, such as intentionality and learning, in a straight-forward manner, at least in principle. For example, intentionality — purposive or goal-directed behaviour — is achieved by the superposition of an intentional potential function on the intrinsic potential function [105]. Similarly, learning is viewed as the modification of already-existing behavioural patterns that take place in a historical context whereby the entire attractor layout (the phase-space configuration) of the dynamical system is modified. Thus, learning changes the whole system as a new attractor is developed.

Although dynamical models can account for several non-trivial behaviours that require the integration of visual stimuli and motoric control, including the perception of affordances, perception of time to contact, and figure-ground bi-stability [71, 73, 105, 108, 233], the principled feasibility of higher-order cognitive faculties has yet to be validated.

The implications of dynamical models are many: as noted in [221], 'cognition is non-symbolic, nonrepresentational and all mental activity is emergent, situated, historical, and embodied'. It is also socially constructed, meaning that certain levels of cognition emerge from the dynamical interaction between cognitive agents. Furthermore, dynamical cognitive systems are, of necessity, embodied. This requirement arises directly from the fact that the dynamics depend on self-organizing processes whereby the system differentiates itself as a distinct entity through it dynamical configuration and its interactive exploration of the environment.

With emergent systems in general, and dynamical systems in particular, one of the key issues is that cognitive processes are temporal processes that ‘unfold’ in real-time and synchronously with events in their environment. This strong requirement for synchronous development in the context of its environment again echoes the enactive systems approach set out in the next section. It is significant for two reasons. First, it places a strong limitation on the rate at which the ontogenetic[10] learning of the cognitive system can proceed: it is constrained by the speed of coupling (*i.e.* the interaction) and not by the speed at which internal changes can occur [243]. Natural cognitive systems have a learning cycle measured in weeks, months, and years and, while it might be possible to collapse it into minutes and hours for an artificial system because of increases in the rate of internal adaptation and change, it cannot be reduced below the time-scale of the interaction (or structural coupling; see next section). If the system has to develop a cognitive ability that, *e.g.*, allows it to anticipate or predict action and events that occur over an extended time-scale (*e.g.* hours), it will take at least that length of time to learn. Second, taken together with the requirement for embodiment, we see that the consequent historical and situated nature of the systems means that one cannot short-circuit the ontogenetic development. Specifically, you can’t bootstrap an emergent dynamical system into an advanced state of learned behaviour. With that said, recall from the Introduction that an important characteristic of cognitive systems is their anticipatory capability: their ability to break free of the present. There appears to be a contradiction here. On the one hand, we are saying that emergent cognitive systems are entrained by events in the environment and that their development must proceed in real-time synchronously with the environment, but at the same time that they can break free from this entrainment. In fact, as we will see in Chapter 6, there isn’t a contradiction. The synchronous entrainment is associated with the system’s interaction with the environment, but the anticipatory capability arises from the internal dynamics of the cognitive system: its capacity for self-organization and self-development involving processes for mirroring and simulating events based on prior experience (brought about historically by the synchronous interaction) but operating internally by self-perturbation and free from the synchronous environmental perturbations of perception and action.

Although dynamical systems theory approaches often differ from connectionist systems on several fronts [105, 221, 67], it is better perhaps to consider them complementary ways of describing cognitive systems, dynamical systems addressing macroscopic behaviour at an emergent level and connectionist systems addressing microscopic behaviour at a mechanistic level [139]. Connectionist systems themselves are, after all, dynamical systems with temporal properties and structures such as attractors, instabilities, and transitions [242]. Typically, however, connectionist systems describe the dynamics in a very high dimensional space of activation potentials and connection strengths whereas dynamical systems theory models describe

[10] Ontogeny is concerned with the development of the system over its lifetime.

the dynamics in a low dimensional space where a small number of state variables capture the behaviour of the system as a whole. Schöner argues that this is possible because the macroscopic states of high-dimensional dynamics and their long-term evolution are captured by the dynamics in that part of the space where instabilities occur: the low-dimensional Center-Manifold [193]. Much of the power of dynamical perspectives comes from this higher-level abstraction of the dynamics [212]. The complementary nature of dynamical systems and connectionist descriptions is emphasized by Schöner and by Kelso [105, 194] who argue that non-linear dynamical systems should be modelled simultaneously at three distinct levels: a boundary constraint level that determines the task or goals (initial conditions, non-specific conditions), a collective variables level which characterize coordinated states, and a component level which forms the realized system (*e.g.* nonlinearly coupled oscillators or neural networks). This is significant because it contrasts strongly with the cognitivist approach, best epitomized by David Marr's advocacy of a three-level hierarchy of abstraction (computational theory, representations and algorithms, and hardware implementation), with modelling at the computational theory level being effected without strong reference to the lower and less abstract levels [130]. This complementary perspective of dynamical systems theory and connectionism enables the investigation of the emergent dynamical properties of connectionist systems in terms of attractors, meta-stability, and state transition, all of which arise from the underlying mechanistic dynamics, and, *vice versa*, it offers the possibility of implementing dynamical systems theory models with connectionist architectures.

As already noted, the benefit of the dynamical systems formulation of system dynamics is that it collapses a very high dimensionality system defined by the complete set of system variables onto a low dimensional space defined by the collective variables. The collective variables are the subset of the system variables that govern the system behaviour.

This formulation of dynamical systems theory effectively describes an emergent spatio-temporal pattern generator that can be perturbed by environmental conditions (the control variables). It is significant that these variables do not define the dynamics: they just act as a perturbing influence, knocking the system ideally from one meta-stable basin of attraction to another. Therefore, as presented, this is essentially a reactive system (*i.e.* a level one autopoietic system; see next section). Two points are significant about such a system:

1. There is no latitude for development;
2. There is no indication of how the required configuration is achieved in the first place and, therefore, the phylogenetic configuration must be identified *a priori*. This problem is tied closely to the decomposition issue mentioned above.

If we wish to have a co-development and co-determination (and therfore a capacity to learn) so that the system can develop a new state space attractors over time, we require a new set of dynamics. This implies that the system must be self-modifying. This is equivalent to the introduction of a nervous system in an autopoietic system, which forms the core of enactive systems models, the topic of the next section.

2.2.3 Enactive Systems

Enactive systems take the emergent paradigm even further. In contradistinction to cognitivism, which involves a view of cognition that requires the representation of a given objective pre-determined world [67, 228], enaction [131, 132, 134, 133, 227, 228, 243] asserts that cognition is a process whereby the issues that are important for the continued existence of a cognitive entity brought out or enacted: co-determined by the entity as it interacts with the environment in which it is embedded. Thus, nothing is ‘pre-given’, and hence there is no need for symbolic representations. Instead there is an enactive interpretation: a real-time context-based choosing of relevance.

In this sense, the philosophical ground of enaction is Husserlian phenomenology, in contradistinction to the objectivist realism of the cognitivist approach. Whilst this might sound out of place here, and indeed irrelevant for those interested in engineering cognitive systems, it has very practical implications. It comes down to a simple choice of axioms upon which to build a cognitive system.

For cognitivism, the role of cognition is to abstract objective structure and meaning through perception and reasoning. For enactive systems, the purpose of cognition is to uncover unspecified regularity and order that can then be construed as meaningful because they facilitate the continuing operation and development of the cognitive system. In adopting this stance, the enactive position challenges the conventional assumption that the world *as the system experiences it* is independent of the cognitive system (‘the knower’). Instead, knower and known ‘stand in relation to each other as mutual specification: they arise together’ [228].

This type of statement is normally anathema to scientists as it seems to be positing a position of extreme subjectivism, the very antithesis of modern science. However, this is not what is intended at all. On the contrary, the enactive approach is an attempt to avoid the problems of both the realist (representationalist) and the solipsist (ungrounded subjectivism) positions.

The only condition that is required of an enactive system is *effective action*: that it permit the continued integrity of the system involved. It is essentially a very neutral position, assuming only that there is the basis of order in the environment in which the cognitive system is embedded. From this point of view, cognition is exactly the process by which that order or some aspect of it is uncovered (or constructed) by the system. This immediately allows that there are different forms of reality (or relevance) that are dependent directly on the nature of the dynamics making up the cognitive system. This is not a solipsist position of ungrounded subjectivism, but neither is it the commonly-held position of unique — representable — realism. It is fundamentally a phenomenological position.

The enactive systems research agenda stretches back to the early 1970s in the work of computational biologists Maturana and Varela and has been taken up by others, including some in the main-stream of classical AI [131, 132, 134, 227, 228, 243, 133].

The goal of enactive systems research is the complete treatment of the nature and emergence of autonomous, cognitive, social systems. It is founded on the concept of autopoiesis – literally *self-production* – whereby a system emerges as a coherent systemic entity, distinct from its environment, as a consequence of processes of self-organization. However, enaction involves different degrees of autopoeisis and three orders of system can be distinguished.

First-order autopoietic systems correspond to cellular entities that achieve a physical identity through structural coupling with their environment. As the system couples with its environment, it interacts with it in the sense that the environmental perturbations trigger structural changes 'that permit it to continue operating'.

Second-order systems are meta-cellular systems that engage in structural coupling with their environment, this time through a nervous system that enables the association of many internal states with the different interactions in which the organism is involved. In addition to processes of self-production, these systems also have processes of self-development. Maturana and Varela use the term operational closure for second-order systems instead of autopoiesis to reflect this increased level of flexibility [133].

Third-order systems exhibit coupling between second-order (*i.e.* cognitive) systems, *i.e.* between distinct cognitive agents. It is significant that second- and third-order systems possess the ability to perturb their own organizational processes and attendant structures. Third-order couplings allow a recurrent (common) ontogenetic drift in which the systems are reciprocally-coupled. The resultant structural adaptation – mutually shared by the coupled systems – gives rise to new phenomonological domains: language and a shared epistemology that reflects (but not abstracts) the common medium in which they are coupled. Such systems are capable of three types of behaviour: (i) the instinctive behaviours that derive from the organizational principles that define it as an autopoietic system (and that emerge from the phylogenetic evolution of the system), (ii) ontogenetic behaviours that derive from the development of the system over its lifetime, and (iii) communicative behaviours that are a result of the third-order structural coupling between members of the society of entities.

Linguistic behaviours are the intersection of ontogenetic and communication behaviours and they facilitate the creation of a common understanding of the shared world that is the environment of the coupled systems. That is, language is the emergent consequence of the third-order structural coupling of a socially-cohesive group of cognitive entities.

Consciousness and the concept of mind emerges when language and linguistic behaviour are applied recursively to the entity engaged in that linguistic behaviour. Issues of consciousness, thought, and mind, then, belong to the domain of social coupling. They don't have any independent existence and as such can't be modelled but are emergent properties of self-referential linguistic behaviours between structurally-coupled and socially-coupled cogitive entities having a shared mutually-constructed emergent epistemology.

The core of the enactive approach is that cognition is a process whereby a system identifies regularities as a consequence of co-determination of the cognitive activities themselves, such that the integrity of the system is preserved. In this approach, the nervous system (and a cognitive agent) does not abstract or 'pick up information' from the environment and therefore the metaphor of calling the brain an information processing device is 'not only ambiguous but patently wrong' [133]. On the contrary, 'all knowing is doing as sensory-effector correlations in the realm of structural coupling in which the nervous system exists'. Knowledge is particular to the system's history of interaction. If that knowledge is shared among a society of cognitive agents, it is not because of any intrinsic abstract universality, but because of the consensual history of experiences shared between cognitive agents with similar phylogeny and compatible ontogeny.

A key postulate of enactive systems is that reasoning, as we commonly conceive it, is the consequence of reflexive[11] use of the linguistic descriptive abilities to the cognitive agent itself [133]. Linguistic capability is in turn developed as a consequence of the consensual co-development of an epistemology in a society of phylogenetically-identical cognitive agents. This is significant: reasoning in this sense is a descriptive phenomenon and is quite distinct from the self-organizing mechanism (*i.e.* structural coupling and operational closure [133]) by which the system/agent develops its cognitive and linguistic behaviours. Since language (and all inter-agent communication) is a manifestation of high-order cognition, specifically co-determination of consensual understanding amongst phylogenetically-identical and ontogenetically-compatible agents, symbolic or linguistic reasoning is actually the product of higher-order social cognitive systems rather than a generative process of the cognition of an individual agent.

As with dynamical systems, enactive systems operate in synchronous real-time: cognitive processes must proceed synchronously with events in the systems environment as a direct consequence of the structural coupling and co-determination between system and environment. However, exactly the same point we made about the complementary process of anticipation in dynamical systems applies equally here. And, again, enactive systems are necessarily embodied systems. This is a direct consequence of the requirement of structural coupling of enactive systems. There is no semantic gap in emergent systems (connectionist, dynamical, or enactive): the system builds its own understanding as it develops and cognitive understanding emerges by co-determined exploratory learning. Overall, enactive systems offer a framework by which successively-richer orders of cognitive capability can be achieved, from autonomy of a system throught to the emergence of linguistic and communicative behaviours in societies of cognitive agents.

While the enactive systems agenda is very compelling, and is frequently referred to by researchers in, for example, developmental psychology, it hasn't achieved great acceptance in main-stream computatational cognitive science and artificial intellegence. The main reason for this is that it is more a meta-theory than a theory *per se*: it is a philosophy of science but it hasn't offered any formal models by which cognitive systems can be either analysed or synthesized. However, it does have a great deal in common with the research agenda in dynamical systems which *is* a scientific theory but is perhaps lacking it ability to prescribe how higher-order cognitive functions can be realized. It has been noted that dynamical systems theory so

[11] Reflexive in the sense of self-referential, not in the sense of a reflex action.

far has been employed more as an analysis tool and less as a tool for the design and synthesis of cognitive systems [148, 37]. The coalescence of the tenets of enactive systems into dynamical systems approaches may well provide the way forward for both communities, and for emergent approaches in general.

The emergent position in general and the enactive position in particular are supported by recent results which have shown that a biological organism's perception of its body and the dimensionality and geometry of the space in which it is embedded can be deduced (learned or discovered) by the organism from an analysis of the dependencies between motoric commands and consequent sensory data, without any knowledge or reference to an external model of the world or the physical structure of the organism [169, 170]. Thus, the perceived structure of reality could therefore be a consequence of an effort on the part of brains to account for the dependency between their inputs and their outputs in terms of a small number of parameters. Thus, there is in fact no need to rely on the classical idea of an *a priori* model of the external world that is mapped by the sensory apparatus to 'some kind of objective archetype'. The conceptions of space, geometry, and the world that the body distinguishes itself from arises from the sensorimotor interaction of the system, exactly the position advocated in developmental psychology [221]. Furthermore, it is the analysis of the sensory consequences of motor commands that gives rise to these concepts. Significantly, the motor commands are *not* derived as a function of the sensory data. The primary issue is that sensory and motor information are treated simultaneously, and not from either a stimulus perspective or a motor control point of view. As we will see in Section 2.3 and 4.4, this perception-action co-dependency forms the basis of many artificial cognitive systems.

The enactive approach is mirrored in the work of others. For example, Bickhard [19] introduces the ideas of self-maintenant system and recursive self-maintenant systems. He asserts that

> 'The grounds of cognition are adaptive far-from-equilibrium autonomy —recursively self-maintenant autonomy —not symbol processing nor connectionist input processing. The foundations of cognition are not akin to the computer foundations of program execution, nor to passive connectionist activation vectors.'

Bickhard defines autonomy as the property of a system to contribute to its own persistence. Since there are different grades of contribution, there are therefore different levels of autonomy.

Bickhard introduces a distinction between two types of self-organizing autonomous system:

1. *Self-Maintenant Systems* that make active contributions to their own persistence but do now contribute to the maintenance of the conditions for persistence. Bickhard uses a lighted candle as an example. The flame vapourizes the wax which in turn combusts to form the flame.
2. *Recursive Self-Maintenant Systems* that do contribute actively to the conditions for persistence. These systems can deploy different processes of self-maintenance depending on environmental conditions: "they shift their self-maintenant processes so as to maintain self-maintenance as the environment shifts".

He also distinguishes between two types of stability: (a) *energy well stability* which is equivalent to the stability of systems in thermodynamic equilibrium — no interaction with its environment is required to maintain this equilibrium — and (b) *far from equilibrium stability* which is equivalent to non-thermodynamic equilibrium. Persistence of this state of equilibrium requires that the process or system does not go to thermodynamic equilibrium. These systems are completely dependent for their continued existence on continued contributions of external factors: they do required environmental interaction and are necessarily open processes (which nonetheless exhibit closed self-organization).

Self-maintenant and recursive self-maintenant systems are both example of far-from-equilibrium stability systems.

On the issue of representations in emergent systems, he notes that recursive self-maintenant systems do in fact yield the emergence of representation. Function emerges in self-maintenant systems and representation emerges as a particular type of function ('indications of potential interactions') in recursively self-maintenant systems.

2.3 Hybrid Approaches

Considerable effort has also gone into developing approaches which combine aspects of the emergent systems and cognitivist systems [78, 79, 80]. Typically, hybrid systems exploit symbolic knowledge to represent the agent's world and logical rule-based systems to reason about this knowledge in order to achieve goals and select actions while at the same time using emergent models of perception and action to explore the world and build these representations. These hybrid approaches have their roots in arguments against the use of explicit programmer-based knowledge in the creation of artificially-intelligent systems [52] and in the development of active 'animate' perceptual systems [15] in which perception-action behaviours become the focus, rather than the perceptual abstraction of representations. Such systems still use representations and representational invariances but it has been argued that these representations should only be constructed by the system itself as it interacts with and explores the world rather than through *a priori* specification or programming so that objects should be represented as 'invariant combinations of percepts and responses where the invariances (which are not restricted to geometric properties) need to be learned through interaction rather than specified or programmed *a priori*' [78]. Thus, a system's ability to interpret objects and the external world is dependent on its ability to flexibly interact with it and interaction is an organizing mechanism that drives a coherence of association between perception and action. There are two important consequences of this approach of action-dependent perception. First, one cannot have any meaningful direct access to the internal semantic representations, and second cognitive systems must be embodied (at least during the learning phase) [79]. According to Granlund, for instance, action precedes perception and 'cognitive systems need to acquire information about the external world through learning or association' … 'Ultimately, a key issue is to achieve behavioural plasticity, *i.e.*, the ability of an embodied system to learn to do a task it was not explicitly designed for.' Thus, hybrid systems are in many ways consistent with emergent systems while still exploiting programmer-centred representational frameworks (for example, see [163]).

Recent results in building a cognitive vision system on these principles can be found in [75, 76, 77]. This system architecture combines a neural-network based perception-action component (in which percepts are mediated by actions through exploratory learning) and a symbolic component (based on concepts — invariant descriptions stripped of unnecessary spatial context — can be used in more prospective processing such as planning or communication).

A biologically-motivated system, modelled on brain function and cortical pathways and exploiting optical flow as its primary visual stimulus, has demonstrated the development of object segmentation, recognition, and localization capabilities without any prior knowledge of visual appearance though exploratory reaching and simple manipulation [145]. This hybrid extension of the connectionist system [146] also exhibits the ability to learn a simple object affordance and use it to mimic the actions of another (human) agent.

An alternative hybrid approach, based on subspace learning, is used in [100] to build an embodied robotic system that can achieve appearance-based self-localization using a catadioptric panoramic camera and an incrementally-constructed robust eigenspace model of the environment.

2.4 Who is Right?

To summarize, Table 2.2 contrasts the four approaches (the cognitivist and the three emergent approaches) under three broad questions: *What is cognition? How does it work?* and *What does a good cognitive system do?* You will see that the perspectives of each approach on these questions are quite different.

The foregoing paradigms have their own strengths and weaknesses, their proponents and critics, and they stand at different stages of scientific maturity. The arguments in favour of connectionist, dynamical, and enactive systems are compelling but the current capabilities of cognitivist systems are actually more advanced.

Several authors have provided detailed critiques of the various approaches. These include, for example, Clark [39], Christensen and Hooker [37], and Crutchfield [45].[12]

Christiansen and Hooker argued [37] that cognitivist systems suffer from three problems: the symbol grounding problem, the frame problem (the need to differentiate the significant in a very large data-set and then generalize to accommodate new

[12] The following is abstracted from [232].

Approaches to Cognition	
What is cognition?	
Cognitivist:	Symbolic computation: rule-based manipulation of symbols
Connectionist:	The emergence of global states in a network of simple components
Dynamical:	A history of activity that brings forth change and activity
Enactive:	Effective action: history of structural coupling that enacts (brings forth) a world
How does it work?	
Cognitivist:	Through any device that can manipulate symbols
Connectionist:	Through local rules and changes in the connectivity of the elements
Dynamical:	Through the self-organizing processes of interconnected sensorimotor subnetworks
Enactive:	Through a network of interconnected elements capable of structural changes
What does a good cognitive system do?	
Cognitivist:	Represents the stable truths of the real world
Connectionist:	Develops emergent properties that yield stable solutions to tasks
Dynamical:	Becomes an active and adaptive part of an ongoing and continually changing world
Enactive:	Becomes a viable part of an existing world of meaning (ontogeny) or shapes a new one (phylogeny)

Table 2.2 Attributes of different approaches to cognition (from [229] and adapted from [221] and [228]).

data),[13] and the combinatorial problem. These problems are one of the reasons why cognitivist models have difficulties in creating systems that exhibit robust sensorimotor interactions in complex, noisy, dynamic environments. They also have difficulties modelling the higher-order cognitive abilities such as generalization, creativity, and learning [37]. According to the Christensen and Hooker, and as we have remarked on several occasions, cognitivist systems are poor at functioning effectively outside narrow, well-defined problem domains.

Enactive and dynamical systems should in theory be much less brittle because they emerge — and develop — through mutual specification and co-determination with the environment, but our ability to build artificial cognitive systems based on these principles is actually very limited at present. To date, dynamical systems theory has provided more of a general modelling framework rather than a model of cognition [37] and has so far been employed more as an analysis tool than as a tool for the design and synthesis of cognitive systems [148, 37]. The extent to which this will change, and the speed with which it will do so, is uncertain. Hybrid approaches appear, to some at least, to offer the best of both worlds: the adaptability of emergent systems (because they populate their representational frameworks through learning and experience) but the advanced starting point of cognitivist systems (because the representational invariances and representational frameworks don't have to be learned but are designed in). However, it is unclear how well one can combine what are ultimately highly antagonistic underlying philosophies. Opinion is divided, with arguments both for (*e.g.* [39, 45, 76]) and against (*e.g.* [37]).

[13] In the cognitivist paradigm, the frame problem has been expressed in slightly different but essentially equivalent terms: how can one build a program capable of inferring the effects of an action without reasoning explicitly about all its perhaps very many non-effects? [201]

Clark suggests that one way forward is the development of a form of 'dynamic computationalism' in which dynamical elements form part of an information-processing system [39]. This idea is echoed by Crutchfield [45] who, whilst agreeing that dynamics are certainly involved in cognition, argues that dynamics *per se* are "not a substitute for information processing and computation in cognitive processes" but neither are the two approaches incompatible. He holds that a synthesis of the two can be developed to provide an approach that does allow dynamical state space structures to support computation. He proposes 'computational mechanics' as the way to tackle this synthesis of dynamics and computation. However, this development requires that dynamics itself needs to be extended significantly from one which is deterministic, low-dimensional, and time asymptotic, to one which is stochastic, distributed and high dimensional, and reacts over transient rather than asymptotic time scales. In addition, the identification of computation with digital or discrete computation has to be relaxed to allow for other interpretations of what it is to compute.

It might be opportune to remark at this point on the dichotomy between cognitivist and emergent systems. As we have seen, there are some fundamental differences these two general paradigms — the principled disembodiment of physical symbol systems *vs.* the mandatory embodiment of emergent developmental systems [230], and the manner in which cognitivist systems often preempt development by embedding externally-derived domain knowledge and processing structures, for example — but the gap between the two shows some signs of narrowing. This is mainly due (i) to a fairly recent movement on the part of proponents of the cognitivist paradigm to assert the fundamentally important role played by action and perception in the realization of a cognitive system; (ii) to the move away from the view that internal symbolic representations are the only valid form of respresentation [39]; and (iii) to the weakening of the dependence on embedded *a priori* knowledge and the attendant increased reliance on machine learning and statistical frameworks both for tuning system parameters and the acquisition of new knowledge both for the representation of objects and the formation of new representations. However, cognitivist systems still have some way to go to address the issue of true ontogenetic development with all that it entails for autonomy, embodiment, architecture plasticity, and system-centred construction of knowledge mediated by exploratory and social motivations and innate value systems.

2.5 *Caveat*

Cognition can no longer be equated with symbolic reasoning and it is now viewed by most people — even those drawn from very disparate viewpoints — as being intrinsically related to the issues of perception and action. For example, consider Anderson *et al.* [6]: 'There is reason to suppose that the nature of cognition is strongly determined by the perceptual-motor systems, as the proponents of embodied and situated cognition have argued' and Langley [120] 'mental states are always grounded in real or imagined physical states, and problem-space operators always expand to primitive skills with executable actions'. Anderson and Langley are prominent exponents of the cognitivist approach, with a deep background in traditional AI. Obviously, AI has shifted its position since the days when the physical symbol systems hypothesis held sway, even if there is much more to be done.

Let us close this chapter with a reminder that, to date, no one has actually designed a complete cognitive system. Futhermore, notwithstanding the attractiveness of some paradigms *vis-à-vis* others, there is still considerable disagreement about the right approach to take. The following advice on how to engage (properly) in a scientific controversy [38] is very relevant.

- Beware an either-or mentality
- Try both narrow and broad interpretations of terms
- Given a dichotomy, ask what both assume
- Beware imposing spatial metaphors
- Beware locating relations
- Try viewing "independent" levels as co-determined
- Don't equate a descriptive model with the causal process being described
- Recognize that first approximations are often overstatements
- Beware that words can sometimes mean their opposites
- Enduring dilemmas are possibly important clues
- Periodically revisit what you have chosen to ignore
- Beware of building your theory into the data
- Locate your work within historical debates and trends
- "It's not new" does not refute a hypothesis
- Beware errors of logical typing
- Recognize conceptual barriers to change
- To understand an incomprehensible position, start with what the person is against
- Recognize that the "born again" mentality conceives sharp contrasts
- Recognize how different disciplines study and use as tools different aspects of intelligence
- Recognize the different mental styles of your colleagues