Intelligence - Dynamics and Representations

Luc Steels

Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
E-mail: steels@arti.vub.ac.be

Abstract. The paper explores a biologically inspired definition of intelligence. Intelligence is related to whether behavior of a system contributes to its self-maintenance. Behavior becomes more intelligent (or copes with more ecological pressures) when it is capable to create and use representations. The notion of representation should not be restricted to formal expressions with a truth-theoretic semantics. The dynamics at various levels of intelligent systems plays an essential role in forming representations. An example is given how behavioral diversity spontaneously emerges in a globally coupled network of agents.

Keywords: intelligence, self-organisation, representation, complex dynamical systems.

1 Introduction

Artificial intelligence research is concerned with an investigation into the phenomenon of intelligence using the methods of the artificial [24]. This means that systems are built which exhibit intelligent behavior and that this is seen as a way to progressively derive and test a theory of intelligence. After three decades of research, nobody denies that AI has resulted in many spinoffs for computer science, such as list processing, declarative programming, search algorithms, etc. Nobody denies that a whole range of programs have been written that exhibit features of (human) intelligence. For example, chess programs now compete at grandmaster level, expert systems have demonstrated human-level performance in difficult problems like scheduling, diagnosis, or design, natural language programs of high complexity have been built for parsing and producing natural language, and some machine learning programs have been capable to extract compact representations from examples. But substantial progress is still possible.

First of all, it seems that research efforts so far have not resulted in a coherent, widely accepted theory of intelligence. There is a body of engineering methods, techniques, and intuitive insights, which are usually taught using case studies [30]. This absence of theory is undoubtely due to the engineering bias of AI and the push from society to produce useful artefacts as opposed to theories. In addition, the bits of theory that do exist (for example Newell's cognitive architecture [19]) have no connection with theories of the physical and biological world, so that there is a wide gap between AI and other natural sciences.

Second, there is a continuing criticism that the achievements of AI systems rest completely on the intelligence of the AI programmers: They extract and formalise the knowledge from experts, they set the conceptual framework and determine the set of good and bad examples for machine learning algorithms, they synthetise the grammars going into the natural language systems. The problem of how structures for knowledge and behavior may develop is largely unresolved. This is particularly a bottleneck in the area of sensory-motor intelligence and common sense, where the task of analysis, formalisation, and explicit programming is so formidable that little success has come from using the classical AI approach.

Current research in intelligent autonomous agents promises to tackle these two gaps in a fundamental way: It is seeking a theory of intelligence compatible with the basic laws of physics and biology and a theory which explains how intelligence may come from non-intelligent, material processes. Obviously we are far from achieving these goals. Only the contours of the theory are visible and at the moment the artefacts that can be built are promising with respect to sensory-motor intelligence but still weak with respect to 'higher level' cognitive tasks. But a new methodological track has been opened in which solid work can proceed.

The main pillars of the new approach are as follows:

+ A biologically oriented definition of intelligence is the starting point of the investigation. Intelligence is defined with respect to the capability of an autonomous system to maintain itself. This gives an objective criterion, as opposed to a subjective criterion based on judgement of performance or the ascription of knowledge and reasoning. This definition is refined by considering the functionalities used to increase the chances of survival: representation, specialisation, cooperation, communication, reflection, etc.
+ A theory of intelligence must be compatible with the basic laws of physics and biology and it must be a universal theory, i.e. independent of a particular embodiment (wetware or silicon) or system level (brain component, individual agent, society). This universality can be achieved by using complex dynamical systems theory as a foundation. Intelligence then is seen as the result of a set of non-linear processes which exhibit properties also found in other physical systems. Phenomena like behavioral coherence, cooperation, or the emergence of diversity between agents can be explained using bifurcation theory, chaos, self-organisation, etc.

The goal of this paper is to discuss these two directions of research in more detail. First we focus on the biologically oriented definition of intelligence (section 2), further refining it with the notion of representation (section 3). Then we discuss in which way dynamical systems theory can act as the theoretical foundation of a theory of intelligence.

2 Defining intelligence

AI has wrestled since the beginning with the question of what intelligence is, which explains the controversies around the achievements of AI. Let us first look at some common definitions and then turn to a biologically oriented definition.

The first set of definitions is in terms of comparative performance with respect to human intelligence. The most famous instance of such a definition is the Turing test. Turing imagined interaction with either a human or an intelligent computer program through a terminal. When the program managed to trick the experimenter into believing that it was human often enough, it would qualify as artificial intelligence.

If we consider more restricted versions of the Turing test, for example compare performance of chess programs with human performance, then an honest observer must by now agree that computer programs have reached levels of competence comparable to human intelligence. The problem is that it seems possible (given enough technological effort) to build highly complex programs which are indistinguishable in performance from human intelligence for a specific area, but these programs do not capture the evolution, nor the embedded (contextual) nature of intelligence. As a consequence 'intelligent' programs are often qualified as being no longer intelligent as soon as the person inspecting the program figures out how the problem has been solved. For example, chess programs carry out relatively deep searches in the search space and the impressive performance is therefore no longer thought to be due to intelligence. To find a firmer foundation it seems necessary to look for a definition of intelligence which is not related to subjective judgement.

The second set of definitions is in terms of knowledge and intensionality. For example, Newell has worked out the notion of a knowledge level description [19]. Such a description can be made of a system if its behavior is most coherently described in terms of the possession of knowledge and the application of this knowledge (principle of rationality). A system is defined to be intelligent if a knowledge-level description can be made of it and if it maximally uses the knowledge that it has in a given situation. It follows that artificial intelligence is (almost by definition) concerned with the extraction of knowledge and the formalisation and encoding in computer systems. This approach appears problematic from two points of view. First of all knowledge level descriptions can be made of many objects (such as thermostats) where the label 'intelligence' does not naturally apply. Second, the approach assumes a sharp discontinuum between intelligent and non-intelligent systems and hence does not help to explain how intelligence may have arisen in physical systems nor how knowledge and reasoning relates to neurophysiology.

There are still other definitions, which however are not used within AI itself. For example, several authors, most notably Roger Penrose, claim that intelligence is intimately tied up with consciousness and self-consciousness [21]. This in turn is defined as the capability to intuit mathematical truths or perform esthetic judgements. The topic of consciousness is so far not at the center of discussion in AI and no claims have ever been made that artificial intelligence systems exhibit consciousness (although see the discussion in [29]). Whether this means, as Penrose suggests, that consciousness falls outside the scope of artificial systems, is another matter. In any case it seems that the coupling of intelligence with consciousness unnecessarily restricts the scope of intelligent systems.

Let me now introduce an alternative to these definitions which finds its roots in biology (see for example [17], [6]). We start from the observation that intelligence is a property of living systems. Living systems are defined as systems which actively maintain themselves, using essentially two mechanisms:

+ They continuously replace their components and that way secure existence in the face of unreliable or short-lived components. The individual components of the system therefore do not matter, only the roles they play.
+ The system as a whole adapts/evolves to remain viable even if the environment changes, which is bound to happen.

The drive towards self-maintenance is found at many different levels:

The genetic level. This is the level which maintains the survivability of the species. Mechanisms of copying, mutation, and recombination together with selection pressures operating on the organisms carrying the genes, are the main mechanisms in which a coherent gene pool maintains itself and adapts itself to changing circumstances. At the moment we do not normally use such a level for artificial robotic agents, although we could say that the building plans, the design principles, and the initial structures of one type of agent when it starts its operation correspond to a kind of genetic level.

The structural level. This is the level of the components and processes making up the individual agents: cells, cell assemblies, organs, etc. Each of these components has its own defense mechanisms, renewal mechanisms, and adaptive processes. In the case of the brain, there are neurons, networks of neurons, neural assemblies, regions with particular functions, etc. What appropriate functional units are and how they coherently operate together is the main topic of interest in the study of intelligent autonomous agents. In artificial systems, they involve internal quantities, electronic and computational processes, behavior systems regulating relations between sensory states and actuator states, etc.

The individual level. This is the level of the individual agent which has to maintain itself by behaving appropriately in a given environment. In many biological systems (for example bacteria or ant colonies) individuals have no or little self-interest. But it is clear that the individual becomes gradually more important as evolution proceeded its path towards more complexity, and conflicts arise between genetic pressures, group pressures, and the tendency of the individual to maintain itself. Greater individuality seems to be linked tightly with the development of intelligence. In the case of artificial systems, the individual level corresponds to the level of the robotic agent as a whole which has to survive within its ecological niche.

The group level. This is the level where groups of individuals together form a coherent whole and maintain themselves as a group. This may include defense mechanisms, social differentiation according to the needs of the group, etc. In the case of artificial systems, the group level becomes relevant when there are groups of robotic agents which have to cooperate in order to survive within a particular ecosystem and accomplish tasks together.

Obviously there is a continuum for living systems in terms of the power they have to determine their own destiny, in other words the degrees of freedom or the available choices. A simple bacteria has very little control over its surroundings. It can at best move towards food sources and away from danger. For the rest it is at the mercy of environmental factors. Bacterial genes delineating the species are nevertheless still very successful mostly due to the rate of copying and the multitude and range of environments in which the individuals can survive. The more degrees of freedom living systems have, the more their chances of survival will increase but also the more structures and processes must be dedicated to making appropriate choices.

The qualification 'intelligent' must be seen against this general background. A biologically inspired definition of intelligence focuses on the interaction between a living system (at whatever level) and the environment, which includes other living systems. Such an interaction is typically called a behavior. The main requirement for a behavior to be intelligent is that it contributes to the continued survival of the system, directly or indirectly. Thus when an animal which is starving from hunger does not go to a food source but performs some other action such as fight another animal, then we would say that this behavior is non-intelligent. Obviously intelligent behavior depends strongly on the environment and a particular system will only behave intelligently within a certain environmental niche to which it is adapted. Because environments tend to evolve, a necessary aspect of intelligence is that it is adaptive.

The advantage of this definition is that it can be precisely quantified. It is possible to identify the characteristic pressures in an ecosystem (for example the availability of resources, the presence of dangers, etc.) and to measure behavior and its impact on the viability [16]. Well worked out theories already exist for example at the population dynamics level or at the genetic level. At this point all this is still a somewhat academic exercise for robotic agents because there are no autonomous agents yet in the world that are viable for a sufficiently long time in real world environments. But once the technology matures, it will be common place to take this perspective.

3 Representations.

Many researchers would find a definition of intelligence in terms of survivability not strong enough. They would argue, rightfully, that the appropriate metabolism, a powerful immune system, etc., are also critical to the survival of organisms (in the case of artificial systems the equivalent is the life time of the batteries, the reliability of microprocessors, the physical robustness of the body). They would also argue that many biological systems (like fungi) would then be more intelligent than humans because they manage to survive for much longer periods of time. So we need to sharpen the definition of intelligence by considering what kind of functionalities intelligent systems use to achieve viability.

Here we quickly arrive at the notion of representation. The term representation is used in its broadest possible sense here. Representations are physical structures (for example electro-chemical states) which have correlations with aspects of the environment and thus have a predictive power for the system. These correlations are maintained by processes which are themselves quite complex and indirect, for example sensors or actuators which act as transducers of energy of one form into energy of another form. Representations support processes that in turn influence behavior. What makes representations unique is that processes operating over representations can have their own dynamics independently of the dynamics of the world that they represent.

Although it seems obvious that the ability to handle representations is the most distinguishing characteristic of intelligent systems, this has lately become a controversial point. Autonomous agents researchers have been arguing 'against representations'. For example, Brooks [3] has claimed that intelligence can be realised without representations. Researchers in situated cognition [4], [22] and in 'constructivist' cognitive science [14] have argued that representations do not play the important role that is traditionally assigned to them. Researchers in neural networks in general reject 'symbolic representations' in favor of subsymbolic or non-symbolic processing [26]. All this is resulting in a strong debate of representationalists vs. non-representationalists [8]. Let me attempt to clarify the issues.

In classical AI, physical structures acting as representations are usually called symbols and the processes operating over them are called symbol processing operations. In addition the symbol processing is subjected to strong constraints: Symbols need to be defined using a formal system and symbolic expressions need to have a strict correspondence to the objects they represent in the sense of Tarskian truth-theoretic semantics. The operations that can be performed to obtain predictive power must be truth-preserving.

These restrictions on representations are obviously too narrow. States in dynamical systems [11] may also behave as representations. Representations should not be restricted to those amenable to formal semantics nor should processing be restricted to logically justified inferences. The relation between representations and reality can and usually is very undisciplined, partly due to the problem of maintaining strict correspondence between the environment and the representation. For example, it is known that the signals received by sonar sensors are only for 20 percent effectively due to reflection from objects. Sonar sensors therefore do not function directly as object detectors and they do not produce a 'clean representation' of whether there is an object or not in the environment. Rather they establish a (weak) correlation between external states (the presence of obstacles in the environment) and internal states (hypothesised positions of obstacles in an analogical map) which may be usefully exploited by the behavioral models.

Second, classical AI restricts itself mostly to *explicit representations*. A representation in general is a structure which has an influence on behavior. Explicit representations enact this influence by categorising concepts of the reality concerned and by deriving descriptions of future states of reality. An implicit (or emergent) representation occurs when an agent has a particular behavior which is appropriate with respect to the motivations and action patterns of other agents and the environment but there is no model. The appropriate behavior is for example due to an historical evolution which has selected for the behavior. The implicit representation is still grounded in explicit representations but they are at a different level.

Indeed, representations can be postulated at all levels of intelligent systems:

- *Genetic level*: The DNA molecules are the physical structures which act as explicit representations at the genetic level. More specifically, they represent directly the presence of particular structures in the organism using a code based on the position of the individual molecules. Of course DNA does not 'represent' in a Tarskian-style truth-theoretic sense. The relation between the DNA molecules and the resulting structure is complex and what structure is found in the organism is to a large extent determined by the environment as well. At the same time, the genes represent implicitly a set of environments in which the resulting organism can survive.
- *Structural level*: The neural structures causally responsible for behavior have their own explicit representations although there is no consensus on what they are: levels of activation, patterns of spikes, mass behavior

of neurons. Also in artificial systems explicit internal representations in the form of electro-magnetic states are common and computer technology has made it possible to create and manipulate millions of representational states in very short time periods.
- *Individual level*: It seems appropriate to make a distinction between representations at the structural level and at the level of the individual. These are the representations that we are most familiar with. They take the form of drawings, models, conventions, internal languages, etc. The individual's representations are implemented by the representations and processes at the structural level.
- *Group level*: Groups of agents may use representations at the group level to maintain coherence and increase their chances of survival, from the pheromone trails deposited by ant societies to languages and literature used in human societies.

For a long time, science has made progress by reducing the complexity at one level by looking at the underlying components. Behavior at a particular level is explained by clarifying the behavior of the components at the next level down. For example, properties of chemical reactions are explained (and thus predicted) by the properties of the molecules engaged in the reactions, the properties of the molecules are explained in terms of atoms, the properties of atoms in terms of elementary particles, etc. Also in the case of intelligence, we see that many researchers hope that an understanding of intelligence will come from understanding the behavior of the underlying components. For example, most neurophysiologists believe that a theory of intelligence will result from understanding the behavior of neural networks in the brain. Some physicists go even so far as to claim that only a reduction of the biochemical structures and processes in the brain to the quantum level will provide an explanation of intelligence ([21]).

At the moment there is however a strong opposing tendency to take a wholistic point of view, also in the basic sciences [5]. This means that it is now understood that there are properties at each level which cannot be reduced to the level below, but follow from the dynamics at that level, and from interactions (resonances) between the dynamics of the different levels ([20]). In the case of intelligence, this means that it will not be possible to understand intelligence by only focusing on the structures and processes causally determining observable behavior. Part of the explanation of intelligence will come from the dynamics in interaction with the structures and processes in the environment, and the coupling between the different levels.

This viewpoint is adopted in the strongest possible sense in our own work, and is common in current research on intelligent autonomous agents. One implication is for example, that it makes less sense from a methodological point of view to build a particular robot which executes a particular task. In our laboratory, we have created a complete robotic ecosystem (figure 3) which involves an environment with different pressures for the robots (e.g.

the need to collect energy and ensure that it is available), different robotic agents which have to cooperate but are also in competition with each other, and a growing repertoire of adaptive structural components (called behavior systems) which are causally responsible for behavior. (see [27], [17]).

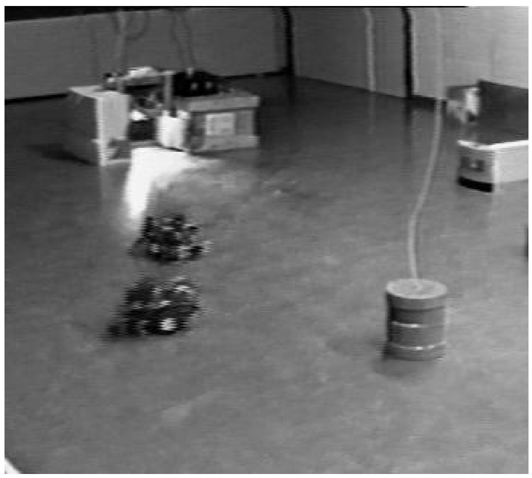

Fig. 1. Robotic ecosystem constructed at the VUB AI laboratory. There is a charging station which robots can use to recharge their batteries. There are also 'parasites' in the form of lamps which take energy away from the charging station. Robots temporarily kill off parasites by pushing against the boxes.

Such an integrated experimental environment ensures that all the different levels (genetic, structural, individual, group) are present at the same time, each with strong interactions to the environment. This way a wholistic approach to the study of intelligence is possible.

A wholistic position must also be adopted for intelligent behavior and for representations. When a particular behavior is observed, it is not at all clear at which level the representation causally co-responsible for it, should be located. Classical AI is too much focused on representations at the individual level and assumes that there is a direct correspondence between the individual and the structural level. But many representations of the world could be implicit, i.e. they are assumed by particular behavior patters due to a historical selection process.

So 'intelligence' is no longer viewed here as restricted to a unique capability that only (conscious) humans have. Intelligence occurs at all levels, although one can say that the most developed forms are at the level of humans and their cultures; most complex in the sense that at all levels we find the most complex representations, the most versatile behavior creation, and strong and elaborate forms of cooperation and communication.

Nor is intelligence viewed here as an all-or-none phenomenon. In each case, it is possible to trace an evolutionary path (with many co-evolutionary couplings re-enforcing the build up of complexity) in which the first signs of a particular functionality (e.g. communication) are becoming apparent and the functionality then gradually develops under the ecological pressures into the complex forms that we can observe in humans. Understanding such evolutionary paths is one of the main challenges of autonomous agents research.

Science proceeds by formulating abstract mathematical theories with which it is possible to describe a wide range of natural phenomena. The mathematical deduction or calculation based on the theory can then be mapped onto predictions of reality in order to check whether the theoretical description is valid. At the moment the most worked out formal theory used in theories of intelligence is based on logic (e.g. [9]). But the search for a theory of intelligence which is compatible with physics and biology and which sees intelligence as a universal phenomenon present at many different levels of biological systems, pushes us into another direction. Most theories of complex natural phenomena are phrased in terms of the recently developed theory of complex dynamical systems, which includes theories of chaos and self-organisation. It is therefore no surprise that several researchers in the field of intelligent autonomous agents have been seeking a foundation in the same direction ([7], [25], [27], [13]).

4 Emergent diversification in agent behavior

To illustrate many points of the paper, I will now develop a concrete example in the context of the robotic experiments mentioned earlier. We will be interested in the question how in a group of agents, individual differences between agents may arise.

Three aspects need to be described: (i) the contextual setting of the experiment, (ii) the experimental results, and (iii) the underlying theory.

(i) Contextual setting.

The experiment involves a group (at least 2) of robotic agents. Each agent has 30 sensors (for infrared, touch, visible light, battery level, forward movement, sound) mounted on a body which also houses batteries and motors connected to a left and right wheel (figure 2).

The robots have a limited behavioral repertoire described in more detail in [28]: It includes forward movement, touch-based obstacle avoidance, infrared-based (smooth) obstacle avoidance, phototaxis towards a light mounted on a charging station, phototaxis towards lamps mounted in boxes, etc. Each behavior is a dynamical process implemented in a dedicated programming language PDL implemented on top of C.

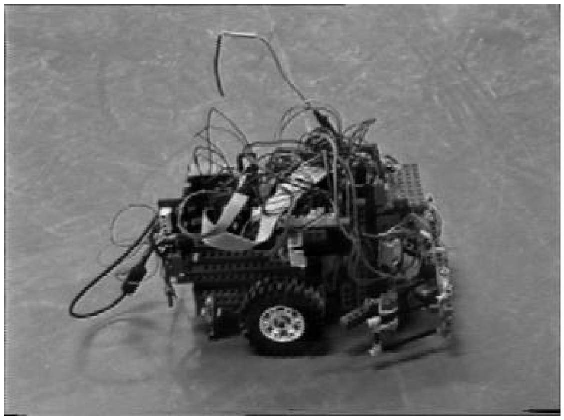

Fig. 2. Robotic agents used in the experiments. Each robot is autonomous from the viewpoint of energy (thanks to batteries) and processing (thanks to onboard dedicated electronics, microcontrollers located on a sensory-motor unit, and the main processing unit).

The processes directly relate sensory states (which are continuously varying) to actuator parameters. They typically take the form of feedback control processes with three components:

1. A cue c, for example the amount of infrared light perceived. In the case of infra-red based obstacle avoidance, this infrared light needs to be minimised.

2. A rate r, which determines the slope with which the desired state is approached.

3. An intensity i, which is determined by motivational processes.

These components exert an influence on an actuator parameter a based on the formula:

$a \leftarrow i\ r\ c$

The PDL system freezes at time step t all quantities, executes all processes, sums the influences and enacts them on the quantities. It then sends the final values to actuators, reads new sensory values and proceeds to the next time step (t+1). Such a cycle takes place 100 times per second (for a repertoire of up to 500 processes).

Motivational processes monitor dimensions related to viability of the agent (for example presence of enough energy in the batteries). A motivational quantity (such as EnergyNeed) increases as its monitored quantity (Energy in battery) decreases. This quantity is then coupled to the behavioral intensity of those behaviors that may contribute to decreasing the motivational quantity. A motivational process has also a cue (for example the difference between the maximum battery level and the current battery level) and a rate, which is the rate at which the motivational quantity should increase w.r.t. the cue.

$m \leftarrow r\ c$

A further important characteristic is that all processes (both behavioral and motivational) are adaptive in the sense that their rate r changes depending on the resulting behavior. For example, if the impact of the behavioral processes monitoring infrared does not cause the robot to turn away from obstacles and collisions occur, then the rate with which there is an influence on the actuators must be increased. These adaptations are carried out by a third set of processes (the adaptation processes) which monitor the performance of behaviors and couple it to the increase or decrease of the rate parameters.

Although the basic control mechanisms are linear, there are two sources of non-linearity in the overall system. First of all, each behavioral process is adaptive. As a consequence its rate may increase or decrease, thus making the process non-linear. Second, the cue strength typically increases non-linearly as a particular behavior is enacted. For example, the amount of light received during phototaxis will increase non-linearly as the robot approaches the light.

It is impossible within the limited space available here to explain in full detail how such a collection of distributed adaptive processes may give rise to coherent behavior, particularly because there is no central action selection mechanism that decides what the most appropriate behavior is at any point in time. Instead, all behavior systems are active at all times and they cooperate to give the globally coherent behavior. But the most important point for the further discussion is that the interactions between the different behaviors and the environment causes an activity cycle to emerge which has a characteristic cycle time. The activities in the cycle are:

- seek and destroy competitors
- seek charging station
- recharge

McFarland [17] has worked out the optimality conditions for such a cycle and also the possible evolutions towards more optimality.

Each agent individually goes through these characteristic activity cycles spending varying amounts of time depending on the parameter settings. For example, if the motivation to recharge overtakes the motivation to do more work because the battery is getting low, there is a switch from the 'seek parasites' activity to the 'seek charging station' activity, which means that the various behaviors involved with these activities have a stronger tendency to occur. But the individual cycles are coupled: to the environment, because the ability to recharge will depend on the availability of energy in the charging station, and to the other agents, because if one agent is in the charging station, another agent cannot have access.

The question now is whether the different agents could smoothly cooperate in exploiting the available resources. This means that the right setting must be found of the various parameters in the different motivational processes. From an analytic point of view this problem is extremely difficult (see [18]). And the question is whether the robots would be able to discover themselves the right parameter settings using adaptation processes. The answer turns out to be yes, and there are some surprising side effects to be explained in the next paragraphs.

We focus on one parameter w, which could be called 'the amount of work that a robot believes it has to do before going to the charging station'. This parameter has an impact on the process that relates the units U of work already done to the motivation M to do more work. One unit of work is equal to one push against the box that houses a lamp, i.e. a parasite. This process is defined as follows:

$M \leftarrow w(Max - U)$

Max is equal to the maximum amount of work.

Fig. 3. Function relating the amount of work done with the motivation to do more work. When the parameter w is high (e.g. w = 0.5, left figure) the robot 'believes' that it has to do more work than when the parameter is low (e.g. w = 0.2, right figure)

The adaptation process is as follows:

1. When the robot arrives in the charging station and is forced to leave the charging station before its battery is full due to the parasites which take away too much energy, then the parameter w is increased. The rationale behind this is that not enough work (i.e. killing of parasites) was done, so next time more needs to be done.

2. When the robot has been able to recharge itself fully and there is still energy left in the charging station, then the parameter w is decreased. The rationale here is that too much work was done.

(ii) Experimental results

Simulations in the context of the complete ecosystem for a period of 4 hours with 2 robots show that after a while three different types of situations occur:

1. There is a situation, illustrated in fig. 4, in which the robots are both viable and oscillate around the same nearly optimal values of the parameters.

2. There is a situation in which one of the robots is no longer viable. This is partly due to the fact that in this experiment no communication exists between the robots. One robot locks the other one out of going into the charging station.

3. There is a third situation, which is the most interesting and also the most common one, in which two types of robots emerge (fig. 5). The first robot has a high rate for w which means that it will do a lot more work than the second robot. Effectively there has been a diversification between a 'hard working' and a 'less working' behavior. Because the second robot has more free time, it will have an opportunity to develop other behaviors.

Fig. 4. Situation in which both robots oscillate around the same nearly optimal parameter value for w.

It is interesting to raise in this context the question what kind of representations are used. Clearly each invididual agent does not have an explicit model of the energy left in the charging station, what other agents do, how much energy it will need to find back the charging station, etc. Consequently the agents can not do any planning (in the traditional sense). An agent only has explicit representations of the rates relating the amount of work done with motivation to work and the amount of energy left in the battery with motivation to go recharging. The agent has implicit representations of the aspects of the environment relevant for their decision making. Nevertheless the agents behave optimally with respect to an exploitation of the energy available in the overall environment.

Fig. 5. Situation in which diversification between different behaviors emerges. There is now a hard working robot with high value of w and a less working robot with low value of w. An equilibrium situation has however developed making both robots viable.

(iii) Underlying theory

These experimental results are fascinating but do not yet make up a scientific theory. For that we need to go one step further and find abstract mathematical objects which exhibit the same properties. Then we know that the behavior is due to the dynamics itself (and not to some artefact of the robotic ecosystem or other intervening factors). We will also have a predictive theory and possibly learn about other possible situations which we have not been able to investigate yet experimentally. If the particular dynamics also explains other natural phenomena, particularly in biology, then this is further evidence that aspects of intelligence can be explained using concepts from the natural sciences.

It turns out that there is a theory which is appropriate here. This is the theory of coupled map lattices, developed by Kaneko [12] and various co-workers. Kaneko has used this theory for quite different biological phenomena, such as cell differentiation or the maintenance of diversity in population dynamics.

Globally coupled maps consist of a network of elements which each have a certain (possibly chaotic) dynamics. An example is the following:

$$x_{n+1}(i) = (1-\epsilon) f(x_n(i)) + \frac{\epsilon}{N} \sum_{j=1}^{N} f(x_n(j))$$

where n is a discrete time step, i is the index of an element and $f(x) = 1 - ax^2$. Synchronous oscillation occurs due to the interaction between the different elements, whereas the chaotic instability inherent in f introduces the potential for destruction of the coherence. A cluster is defined as a set of elements for which x(i) is the same. k is the number of clusters in the total network.

Kaneko [12] has shown that depending on the degree of non-linearity (captured by the parameter a) four phases occur:

1. There is a coherent phase in which all elements oscillate in harmony (k = 1).
2. There is ordered phase in which few clusters are observed.
3. There is partially ordered phase with a coexistence of attractors with many and few clusters.
4. All attractors have N clusters, where N is the number of elements in the network.

Kaneko has also observed phenomena of intermittency, where the self-organisation towards coherent structure is seen in cascade with the occurrence of high-dimensional disordered motion.

The situation being investigated maps onto this model as follows. The individual cycles of the agents map onto the function f which is present in each of the elements. The elements are globally coupled through the charging station and the parasites. The degree of non-linearity is related to the rate r with which the rate w is adapted, because adaptation causes the linear functions to become non-linear. In the experiments, an adaptation rate r was apparently chosen so that the second regime was observed (an ordered phase with few clusters). The theory predicts that under other parameter selections high-dimensional disordered motion may occur or the other two situations. These predictions have not yet been tested experimentally.

The example illustrates the following points:

- It shows how behavioral diversity (and thus diversity between different agents) can emerge spontaneously even though every agent starts with the same initial structure, and even though every agent has the same self-interest to survive in optimal circumstances.
- It shows that the diversity is not due to structure inside each agent, but through the coupling between the different agents through the environment (and particularly the global constraints imposed by the charging station and the parasites). This illustrates that behavior (at one level) cannot be understood in isolation but that a wholistic point of view must be taken.
- It shows that the diversity is due to the dynamics. In other words, no new principle or explanation is required because universal properties of abstract dynamical systems (in this case globally coupled maps) already exhibit these phenomena. More generally, the clustering phenomena shown here are a special case of the theory in which order arises through fluctuations causing bifurcation and thus evolution in systems ([20]).
- The agents have no explicit models of the strategies of other agents, of the time it takes to recharge, the average time to find the charging station, etc. Their behavior could nevertheless be called intelligent from the viewpoint of optimal use of resources within the given ecological constraints. This is an example where implicit representations dominate.

5 Conclusions

The paper discussed approaches towards a theory of intelligence which are grounded in biological theory and the theory of complex dynamical systems. The approach starts from the idea that intelligence centers around the ability of a system to maintain itself through the creation and use of representations. Intelligence is seen at many different levels and is partly due to the coupling between the different levels. Representations are not necessarily explicit but may be implicit in distributed behavior. This has been illustrated with a concrete example of how behavioral diversity may originate from adaptive processes.

6 Acknowledgement

The viewpoints discussed in this paper have been shaped and greatly enhanced by discussions with many people, including discussions at the Trento NATO ASI. Thanks are due in particular to Thomas Christaller, David McFarland, Rolf Pfeifer, Tim Smithers, and Walter Van de Velde. Danny Vereertbrugghe, Peter Stuer, and Filip Vertommen have constructed the physical ecosystem and the robots. Johan Myny has contributed in the construction and investigation of the robotic ecosystem. This research was partially sponsored by the Esprit basic research project SUBSYM and the DPWB concerted action (IUAP) CONSTRUCT of the Belgian government.

References

1. Arkin, R. (1989) Motor Schema based mobile robot navigation. Int. Journal of Robotics Research. Vol 8, 4 p. 92-112.
2. Brooks, R. 1991. Intelligence without reason, IJCAI-91, Sydney, Australia, pp 569–595.
3. Brooks, R. 1991. Intelligence without representation, AI Journal.
4. Clancey, W.J. (1993) Situated Action: A Neuropsychological Interpretation. Cognitive Science 17(1), 87-116.
5. Cohen, J. and I. Stewart (1994) The Collapse of Chaos.
6. Dawkins, R. (1976) The Selfish Gene. Oxford University Press. Oxford.
7. Engels, C. and G. Schoener (1994) Dynamic fields endow behavior-based robots with representations. Robotics and Autonomous Systems. 1994.
8. Ford, K. and P. Hayes (1994) The baby and the bathwater. AI Journal.
9. Genesereth, M. and N. Nilsson (1987) Logical Foundations of Artificial Intelligence. Morgan Kaufmann Pub. Los Altos.
10. Haken, H. (1983). *Advanced synergetics: instability hierarchies of self-organisating systems and devices*, Springer, Berlin.
11. Jaeger, H. (1994) *Dynamic Symbol Systems* Ph.D. thesis. Faculty of Technology. Bielefeld.

12. Kaneko, K. (1994) Relevance of dynamic clustering to biological networks. Physica D 75, 55-73.
13. Kiss, G. (1993) Autonomous Agents, AI and Chaos Theory. In: Meyer, J.A., et.al. (eds.) *From Animals to Animats 2* Proceedings of the Second Int. Conference on Simulation of Adaptive Behavior. MIT Press, Cambridge. pp. 518-524.
14. Maturana, H.R. and F.J. Varela (1987) The Tree of Knowledge: The Biological roots of Human Understanding. Shamhala Press, Boston.
15. McFarland, D. (1990) Animal behaviour. Oxford University Press, Oxford.
16. McFarland, D. and T. Boesser (1994) Intelligent Behavior in Animals and Robots. MIT Press/Bradford Books, Cambridge Ma.
17. McFarland, D. (1994) Towards Robot Cooperation. Proceedings of the Simulation of Adaptive Behavior Conference. Brighton. MIT Press.
18. McFarland, D., E. Spier, and P. Stuer (1994)
19. Newell,A. (1981). The knowledge level, *Journal of Artificial Intelligence*, vol 18, no 1, pp 87–127.
20. Nicolis, G. and I. Prigogine (1985) Exploring Complexity. Piper, Munchen.
21. Penrose, R. (1990) The Emperor's New Mind. Oxford University Press. Oxford.
22. Pfeifer, R. and P. Verschure (1992) Distributed Adaptive Control: A Paradigm for Designing Autonomous Agents. In: Varela, F.J. and P. Bourgine (eds.) (1992) Toward a Practice of Autonomous Systems. Proceedings of the First European Conference on Artificial Life. MIT Press/Bradford Books, Cambridge Ma. p. 21-30.
23. Schöner, G. and M. Dose (1993) A dynamical systems approach to task-level system integration used to plan and control autonomous vehicle motion, *Journal of Robotics and Autonomous Systems*, vol 10, pp 253–267.
24. Simon, H. (1969) The Sciences of the Artificial. MIT Press, Cambridge Ma.
25. Smithers, T. (1994) Are autonomous agents information processing systems? In: Steels, L. and R. Brooks, (Eds.), *The 'artificial life' route to 'artifical intelligence': building situated embodied agents*, Lawrence Erlbaum Associates, New Haven.
26. Smolensky, P. (1986) Information Processing in Dynamical Systems. Foundations of Harmony Theory. In: Rumelhart, D.E., J.L. McClelland, (eds.) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition.* Vol 1. MIT Press, Cambridge Ma. pp. 194-281.
27. Steels, L. (1994a) The artificial life roots of artificial intelligence. Artificial Life Journal, Vol 1,1. MIT Press, Cambridge.
28. Steels, L. (1994b) A case study in the behavior-oriented design of autonomous agents. Proceedings of the Simulation of Adaptive Behavior Conference. Brighton. MIT Press. Cambridge.
29. Trautteur, G. (ed.) (1994) Approaches to consciousness. Kluwer Academic Publishing. Amsterdam.
30. Winston, P. (1992) Artificial Intelligence. Addison-Wesley Pub. Cy. Reading Ma.

This article was processed using the LaTeX macro package with LLNCS style