Intelligence with Representation

Luc Steels
Sony CSL Paris and
VUB AI-lab University of Brussels
Pleinlaan 2 1000 Brussels
steels@arti.vub.ac.be

Abstract

Behavior-based robotics has always been inspired by earlier cybernetics work such as that of Grey Walter. It emphasises that intelligence can be achieved without the kinds of representations common in symbolic AI systems. The paper argues that such representations might indeed not be needed for many aspects of sensori-motor intelligence but become a crucial issue when bootstrapping to higher levels of cognition. It proposes a scenario in the form of evolutionary language games by which embodied agents develop situated grounded representations adapted to their needs and the conventions emerging in the population.

Keywords: behavior-based robotics, representations, evolutionary language games

1 Introduction

In the nineteen sixties, Artificial Intelligence adopted a knowledge-based or symbolic approach to the synthesis of intelligence, by operationalising models from logic, generative linguistics, and cognitive psychology. This paradigm still dominates to a large extent most of A.I. research and practice [Russell and Norvig, 1995]. It has led to a stream of very interesting insights and powerful artificial systems, particularly in the domain of knowledge engineering and more recently 'semantic' web applications.

But in the nineteen nineties an alternative paradigm, known as the behavior-based approach emerged, trying to put the study of artificial intelligence on biological, instead of logical or psychological, grounds [Steels, 1994]. As a point of departure, this paradigm made contact again with the work of earlier cyberneticians, like Grey Walter. Many biological concepts, such as embodiment, adaptation, emergence, ecology, evolution, and self-organisation, proved to be highly relevant for autonomous robots operating in real-time in a highly dynamic environment [Pfeifer and Scheier, 2000]. However, pushing the behavior-based paradigm in the direction of higher cognition has been more difficult.

This paper first sketches our own initial experiments in behavior-based robotics and draws some of the major lessons learned with this approach. It then raises the question how the paradigm could touch on the issues for which the knowledge-based paradigm has been successful, particularly intelligent tasks that seem to imply necesssarily some form of conceptualisation and representation. For these tasks, the behavior-based approach has not yielded very convincing results so far. This raises two questions: (1) How can embodied cognitive agents construct representations and share them with others? In the spirit of the behavior-based approach, the representations cannot be imposed in a top-down manner by designers but need to evolve dynamically. (2) What is the role of communication and language in bootstrapping representational capacities? Perhaps language is the key motor that allows agents to develop and share conceptualisations of the world.

The paper then presents very briefly some recent experiments addressing these questions. They show how populations of embodied autonomous agents given the appropriate interactive behaviors can indeed construct shared external representations to communicate about their environment and thus bootstrap their cognitive capacities. Throughout the paper, I emphasise conceptual advances rather than technical detail, which can be found in the references.

2 The Behavior-based Approach to Robotics

Work in behavior-based approaches robotics usually shows three characteristics: There is first a basic control layer with embodied direct couplings between sensors and actuators. Then there is a second layer in which motivational variables are introduced to control when these behaviors should become active, thus leading to a hierarchical system. Finally there is a stage in which the collective dynamics of a group of robots becomes exploited.

2.1 Grey Walter Revisited

Around 1990, David McFarland, the Oxford ethologist, and myself started a project at the A.I. laboratory in Brussels (VUB) to reconstruct Grey Walter's Elmer and Elsie turtle experiment [Walter, 1950]. We built the robots with digital as opposed to analog technology and used Lego-technics with added sensors and actuators to have great flexibility in design and construction (see figure 2, 1), but the basic objective and also the general behavioral architecture of the robots was the same. The robots moved around randomly exploring their environment and used phototaxis to guide themselves towards sources of light. They had a battery and a way to monitor how much energy was left. Each robot had a charging rod on top. It could slide in a charging station detectable through visible light and thus recharge itself.

In the spirit of Grey Walter and the early cyberneticians, the robots had a repertoire of behaviors in the form of dynamical systems that established a direct coupling between sensing and acting. Here are some examples of these behaviors (see [Steels, 1994]):

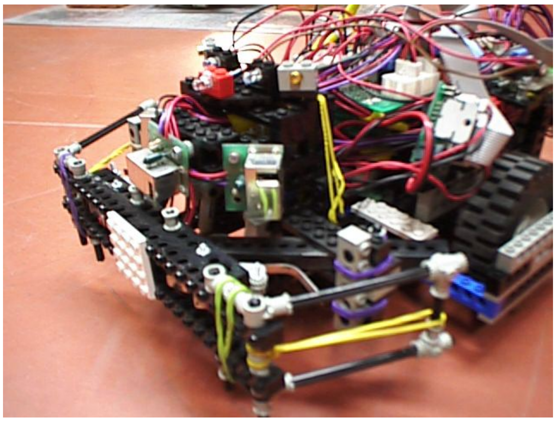

Figure 1: A close-up frontal view of a robot used in our reconstruction of the Grey Walter experiment. It shows the touch sensors mounted on the bumpers, and the infrared and visible light sensors. The robot was built out of Lego-technics had a custom-made digital 'brain brick' as computational engine and a charging rod on top to recharge itself.

- Forward Movement: This was realised by a feedback control system which maintained a certain speed (measured by a wheel counter) by increasing or decreasing the speed on the left or right mounted motors.

- Touch-based Obstacle Avoidance: This was implemented by pulling down the motor speed from the left or right motors depending on touch sensing using contact switches mounted on the left, front, or right side of the robot. Forward movement and touch-based obstacle avoidance operating in parallel enabled a robot to wander around randomly in the arena. (see figure 3)

- Infra-red based Obstacle Avoidance: This was implemented by modulating the motor speed, depending on how much infra-red, which was actively projected by the sensor at a certain frequency, was received back. The active infrared sensors were mounted on the front, the left and the right side of the robot.

- Visible light orientation: This was implemented by modulating the left and right motor speed in such a way that the left and right photosensor received equal amounts of visible light. When there was more light on the left, the right motor was increased and the left motor speed decreased. When there was more on the right, the left motor was increased and the right motor speed decreased.

- Charging behavior: This meant sitting still, i.e. suppressing movement, while

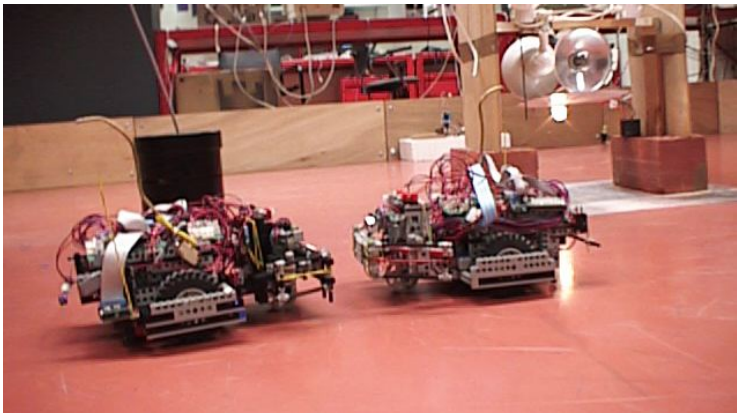

Figure 2: A view of the ecosystem with two of the robots. It contains a charging station with a light mounted on top (shown on the right in the picture) and competitors in the form of black boxes with a light mounted in them that took away energy flowing into the system (one is shown behind the robot on the left).

recharging batteries. This behavior system was activated when the contact switch mounted on the charging rod closed. Inflow of energy into the battery was monitored and the behavior made inactive when the battery was full or when there was no more energy left inside the charging station's battery.

There was no centralised coordination of these behaviors, nor any kind of internal central world model. All behaviors were active in parallel and each interacted with a dynamically changing environment in real time. The coordination took place implicitly, through the environment and through the cumulative impact each behavior exercised on the actuators. For example, after the touch-based obstacle avoidance behavior system had pulled down the speed, the forward movement behavior system would automatically bring it back up to the default speed. Phototaxis towards the charging station required two behavior systems: forward movement and visual light orientation. When operated in parallel, these behaviors bring the robot near the charging station. The robot then usually hits the sides of the charging station, because the light is too diffuse to dock with visible light alone. But obstacle avoidance then pushes the robot away from the charging station again and phototaxis pulls it back towards the station, so that eventually (usually after one or two trials) the robot indeed ends up in the charging station.

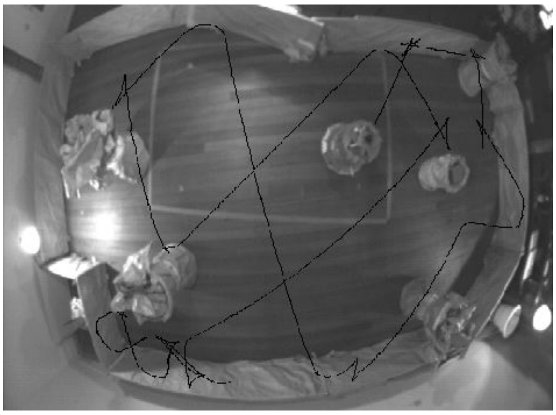

Figure 3: A typical example of wandering behavior in the arena resulting from a parallel combination of forward movement and obstacle avoidance. Behavior of robots could be monitored by an overhead camera and tracking programs.

2.2 Motivated Behaviors

Once Grey Walter's designs were operational, we added an important additional ingredient, directly inspired by McFarland's insights from ethological research in animal behavior [McFarland and Bosser, 1994]. The robots' behaviors needed to be motivated, which meant that we needed to conceive of an ecology in which these behaviors were necessary and hence meaningful. Why indeed would a robot get out of the charging station if staying guaranteed eternal life? So we introduced competitors for the energy in the charging station in the form of boxes in which a light was mounted (figure 2). The light takes away some of the energy that is flowing into the total system and therefore if a robot remains inside the charging station there is less and less energy available for it. However, if the robot pushes against a box, its light dims and thus more energy flows into the charging station so that it becomes available for the next recharge. The robot has a drive to survive and its behaviors have become meaningful with a particular ecosystem.

Eliminating the competitors required a single new behavior: infrared phototaxis (with infrared at another frequency than that emitted by the active infrared sensors used for obstacle avoidance). Figure 4 shows an example of the internal states of sensors and actuators as the robot is moving towards a box and then starts pushing against it. The thick graphs are tracing the left and right motor speeds. The robot picks up speed until the default speed is reached. It moves forward into a straight line until the infrared sensors start to sense infrared. Then a zig-zag movement starts as the robot attempts to keep itself oriented towards the source of the infrared light. If the robot is near the box,

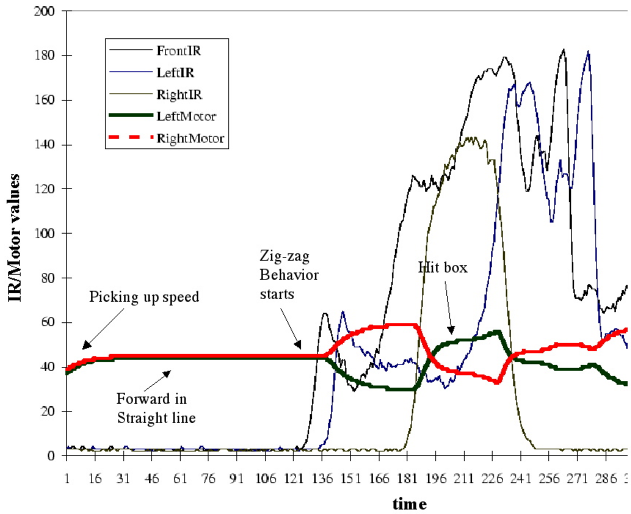

Figure 4: Snapshot of the infrared sensors and the motor speeds as the robot is approaching a box and starts pushing it. Thick graphs plot the left and right motor speeds. The other three graphs plot the infrared sensors.

it will hit it, at which moment the obstacle avoidance behaviors become active and the robot retracts. But the taxis behavior system is still active and pulls the robot once more against the box. This keeps going on until the infrared light of the box diminishes, at which time the attraction diminishes and the robot moves on in a random direction. So the box pushing behavior is emergent. It is not explicitly present as a separate subsystem in the behavioral repertoire of the robot. It is a side effect of a number of simpler behaviors operating together in this particular environment [Steels, 1991].

Handling the competitors also introduces the need for internal motivational states that regulate whether the behaviors connected with looking for the charging station should become active or whether the robot should seek out the competitors and dim its lights. Motivational states are internal dynamical variables. For example, the 'charging-motivation' variable is a function of the energy left in the battery. Motivational states modulate the response intensity of behavior systems. Thus when the charging-motivation becomes higher (as the energy in the battery gets lower), behaviors relevant to phototaxis towards visible light become more responsive. Using dynamical behavioral control as opposed to discrete action selection after logic style planning enable the robot to remain opportunitistic. For example, if the robot is already attracted to the charging station but passes a box on the way, it may still push against it.

The dynamical motivational architecture that has been pioneered in this project goes beyond Grey Walter's experiments. It has become a standard feature of more sophisticated behavior-based robots. For example, in KISMET [Breazeal, 1998], which is an animated head that attempts to maintain a turn-taking interaction with a human, there is a layer of motivational variables in function of the ongoing interaction with a human. Similarly Sony's AIBO [Fujita and Kitano, 1998], which is a dog-like pet robot featuring about a thousand behaviors each very tightly coupled to the world, features a large number of motivational variables reflecting the physical state of the robot and the need to track the emotional states of people interacting with it.

2.3 Collective Dynamics

We used the ecosystem as a platform for learning, both the learning of basic behaviors (using selectionist [Steels, 1950] or re-enforcement techniques [Birk, 1998]) and the learning of correct parameters for each of the behaviors. I will briefly report on one such learning experiment, because of its surprising results and also because it introduces an important additional ingredient with respect to Grey Walter's original designs, namely collective dynamics and self-organisation.

In the experiment, the robots go through cycles, doing work (i.e. pushing against the boxes) and recharging themselves. But the parameterisation of these cycles has to be learned. Although it is possible to develop mathematical models of optimal behavior based on the notion of cost and utility [McFarland and Bosser, 1994], it is impossible in practice to calculate the parameters in advance because the recharging behavior of the batteries (on the robot and in the charging station) as well as the exact behavior of a robot (how long it would take for example to reach the charging station from a particular position) are all unknown in advance and unpredictable.

Learning has been implemented through simple parameter adaptation: When the robot's battery is full and there is still energy left in the charging station, the 'work' parameter, which influences how eagerly a robot engages in taxis towards the boxes, is diminished slightly so that less work is done in the next cycle. When not enough energy is left in the batteries while recharging, this is a sign that more work needed to be done and that the work parameter therefore has to be increased. Based on this simple adaptive strategy, a robot gradually optimises its cyclic behavior in tune with the environment.

When we let two robots roam in the same environment with this kind of adaptive behavior, some unexpected rather dramatic emergent behaviors could be observed. One might expect that the robots evolve towards a balanced workload, in other words that they share equally the amount of work needed to have adequate energy for survival. Indeed this is a possible attractor of this dynamical system and the learning strategy finds this global optimum. But sometimes we also saw a situation emerge in which one robot would perform double as much work as another one. Apparently this was also an attractor of the same dynamical system. How could such a situation ever arise?

Suppose we have two robots A and B. A, due to inevitable stochastic factors, works slightly less, which causes less energy to be available for B. B increases its work parameter and therefore works slightly more on his next cycle. But this provides slightly more energy for A which hence decides to work even less on his next cycle. So there is a positive feedback loop: The less one robot works, the more the other one is forced to work. This progressively brings the robots towards a new stable situation but one in which robot A performs much less work than robot B! A social inequality spontaneously emerges between the robots. There is obviously an upper limit because otherwise B would no longer survive and A would die as well.

This experiment dramatically illustrates an insight which is very common in research in complex dynamical systems, namely that a set of local behaviors may give rise to certain forms of organisation when they are applied by many elements in a shared environment [Langton, 1995]. This principle of self-organisation has yielded a rich harvest of explanations in many areas of biology, including morphogenesis and collective animal behavior [Camazine et al., 2001] and has since been explored with great sophistication by several researchers interested in collective robotics, see e.g. [Melhuish, 2001].

2.4 Insights and Limitations

What are the major insights coming out of this research as well as the limitations? The self-sufficiency experiment sketched above is an example of the 'behavior-based' approach to AI introduced in 1989 [Steels, 1989]. The approach has since become in widespread use [Arkin, 1998]. It sets out an 'artificial life' route towards artificial intelligence [Steels and Brooks, 1994] and in many aspects revives some of the key intuitions of early cybernetics [Braitenberg, 1984]. What are the key 'dogmas' of this behavior-based approach?

1. *Embodiment and dynamics:* The behavior-based approach starts from the physical body operating in a concrete environment. Sensing and actuating is not simply a matter of input and output for a program running without time or space constraints. Instead the primary objective is smooth behavior in real-time and within the resources imposed by the robot. Programs are no longer discrete decision makers but dynamical systems that establish a more or less direct coupling between continuous sensing and continuous actuating.

2. *Situatedness and ecology:* It is necessary to set up the environment in such a way that behaviors become contextualised and are motivated, so that they can be learned or chosen in view of actual needs, like individual or group survival. This emphasis on the concrete and situated nature of behavior and knowledge is shared by researchers in situated cognition [Clancey, 1997].

3. *Emergent behavior:* Putting together simple behaviors in interaction with a specific environment may give rise to unexpected complexity. Such a strategy is different from a hierarchical top-down design normally practised in engineering where components reflect a functional decomposition. Emergent behavior tends

to be more robust and adapted to the environment and brings us a step closer to explain how intelligence may evolve.

4. *Adaptation and evolution:* Organisms (and autonomous robots) must continuously adapt themselves to their environment, because this environment is open-ended and likely to change, sometimes in dramatic ways. Rather than assuming a kind of mature steady state that can be implemented by hand or obtained by an optimisation process, the behavior-based approach emphasises never ending adaptation. Every component of the system must be designed so that it can adapt itself at any moment. The optimal behavior is always transient. Short-term adaptation leads to long-term evolution.

5. *Collective dynamics and self-organisation:* In a multi-agent system, agents form a coupled dynamical system in which individuals influence the whole and in turn get influenced by it. Self-organisation is a typical example of this. It involves a positive feedback loop which re-enforces spontaneous fluctuations in agent behavior so that global coherence arises [Prigogine and Stengers, 1984].

The behavior-based approach has paid off in the sense that many physical robots have now been built based on these principles. They operate in real time in the real world with hardware and software of relatively low complexity [Arkin, 1998]. The robots KISMET [Breazeal, 1998] and AIBO [Fujita and Kitano, 1998] mentioned earlier are particularly nice recent examples. The astonishing humanoid robots which are currently becoming operational, such as the Sony SDR humanoids [Kuroki et al., 2001], are similarly rooted in the behavior-based approach.

3 Representations

The question to be raised next is: How can we go further? Behavior-based robots exhibit smooth integration with dynamic environments but they fall far short of human intelligence in a very important respect, namely the use of representations. Several behavior-based roboticists (most notably Rodney Brooks [Brooks, 1991]) have argued explicitly against representations and it is therefore no surprise that no real progress has been made on this front. De-emphasising representations has worked well for embodied situated behavior, but fails to address issues related to higher level cognition, such as communication in natural language or episodic memory.

Why has there been this aversion to representations? The first point of critique (similar to that raised by philosophers like Searle) is that many of the internal representations assumed in A.I. systems are not grounded in reality, i.e. no process is proposed that relates these representations to the world through a sensori-motor apparatus. Thus the processing remains in the symbolic domain and humans must supply and interpret the initial representations. Of course, it is in principle possible to ground symbolic representations, as one of the first AI robots Shakey [Nilson, 1984] has demonstrated, but further experiments with machine vision and real world robot control have shown that this is very difficult to do it in a reliable way. Many proposed internal representations turn out to require (human-level) intelligence or take so much time that they become non-computable in practice. The second critique is that a centralised executive operating over a global symbolic world model is too slow and brittle to allow the flexible, fast, reactive control needed for physically embodied agents. So it has been proposed that control emerges from the dynamic interaction of independent behavioral components whose activation is modulated by events in the world and by internal motivational states. This strategy makes the centralised rich internal representations assumed by symbolic A.I. superfluous [Steels and Brooks, 1994].

But does all this mean that the notion of representation has to be thrown out with the proverbial bathwater? I don't think so. The first thing to note is that originally the term representation did not have the same connotations as today. Representation meant *external* representation: physical objects, like images, sculptures, pretend play, language sentences, etc. Representations act as a stand-in for objects in the world [Gombrich, 1963]. They imply a conceptualisation of the world, which is expressed using the properties of the medium and the set of emerging conventions in the group. It is only more recently under the influence of logic and early A.I. work that representations have become viewed as something purely internal.

Research by Piaget and others has shown that different forms of representation all arise roughly around the same time and if one form of representation-making is impaired others are as well [Piaget, 1970]. Moreover the ability to construct and interpret (external) representations has been shown to be crucial in the development of the child. If it does not happen it is an early indicator of mental retardation. These findings suggest that representation-making might be a crucial bootstrapping device for higer mental function. At some point external representation-making becomes internalised to form the basis of thinking, in the sense of inner dialogs or mental imaging [Steels, 2003]. So internal representations do not come before external representations but follow or co-evolve with them.

The remainder of this paper reports on experiments focusing on the first step. We try to find out whether behavior-based principles can be applied to the emergence, adaptation, evolution and self-organisation of external representations.

3.1 Emergent Representations

How can we explain how new representations may emerge in a group of agents? Let me explore an analogy with architectural structures (see also [Keller, 1994]). Consider a grass lawn in the form of a square between two buildings on a university campus. The buildings are on diagonally opposite sides. There is a path around the square but people who need to go from one building to another, naturally take the shortest path, which cuts right through the lawn. Even though the gardener has planted a nice smooth grass lawn (and perhaps put up a little sign saying 'Don't step on the grass'), a natural path arises sooner or later as the grass starts to fade away in the places where people step on it. The gardener can try to fight this, but is probably better off creating a real path by clearly marking the naturally emerging path with some sort of material structure and by using gravel on the path so that the grass will not grow. Now everybody, even someone who has never been on campus, will recognise instantly that this is the logical path to take.

The first characteristic of the path, which it shares with representations in general, is that it is a material structure. (External) representations are never purely abstract mathematical entities. Second, the primary purpose of the path is to organise human activity concerned with the solution of a particular problem. In this case, people need to go from one building to the other and the path helps them to do so. So the path regulates behavior to achieve a particular goal. That is what makes it meaningful. The same is true for representations. *Representations are primarily viewed here as organisers of activity rather than abstract models of some aspect of reality.*

How could such a path emerge? Initially we can imagine a phase in which a set of naturally occurring processes (people crossing the lawn) generates as a side effect a physical structure in the environment (an area where there is less grass). Once this has started, a positive feedback loop sets in: The more the physical structure organises the activity, the less grass remains on the path and the better it therefore plays its role as organiser of this activity. Soon everybody crosses the lawn using this path. The intervention of the gardener further enforces this role.

The path has a collective dimension. Many people are involved, both in establishing the path and in using it afterwards. So the path has become a vehicle of communication. Its physical appearance signals that this is the way to cross the lawn. People using the path and thus establishing or re-establishing it indirectly tell other people that this is the normal way to cross the lawn. If the gardener delineates the path, he communicates even more explicitly that it is OK to cross the lawn using the path.

Besides organising activity and enabling communication, the path illustrates another important role of representations. It acts as a record of the activities of people and thus as a memory for how people tend to cross the lawn. The path itself is not the memory, it is the categorisation and interpretation of the path that makes it function as memory. I do not regard memory as a kind of storehouse in which objects are put for future retrieval, but as an active process, categorising and re-categorising structures or events as being meaningful with respect to certain goals [Rosenfield, 1988], [Clancey, 1997].

Our urban environments have plenty of structures like the path and architects and urbanists are very much aware of this. They analyse the desired or naturally occuring social and physical activities of inhabitants and try to invent structures that support this activity [Alexander et al., 1977]. This is not always done correctly, unfortunately. We have all been found struggling with doorknobs which had some esthetic abstract quality but were cognitively impenetrable. We all know buildings or urban structures that have been re-appropriated by inhabitants to satisfy quite different purposes from those intended by the designers.

But to what extent are other external representations, like language utterances or graphical signs, representations? Suppose there is a road on which you can turn left or right. The traffic police want you to go to the left only. One way they could achieve this is to introduce an obstacle so that you cannot go right. This obstacle acts as a representation similar to the path on the lawn. Another way however is to hang a notice above the street which says something like ‘left turn only’. The phrase ‘left turn only’ has the same powers as the road block. There could also be traffic signs, one showing a blue left arrow and another one a red right arrow crossed out. Again they have the same effect as blocking the road. There is of course an important difference. The words ’left turn only’ or the signs with arrows organise activity because their meaning is conventionally known and accepted. This makes it much easier to establish such representations and easier to change their meaning.

Note that the cognitive behavior of human beings has itself a causal impact on whether the organisational role of the path is achieved or not. Gradually the path is categorised as the best way to cross the lawn and this enforces the path. We can look at rivers, how they form, why some of them meander, etc. But this cognitive activity has no impact whatsoever on the river itself. The causal force cutting out the river bed is purely physical. This makes the river no longer a representation. The intervention of a cognitive agent is essential for a material structure to become a representation. The structure must trigger categorisation and then action selection which depends on the outcome of this categorisation. That is why the path on the grass lawn is similar to words in a language. Its meaning is cognitively established. Of course there is a natural relation between the meaning and the material structure of the path. That is why this path is a ‘natural representation’. In the case of language the meaning is established by convention, although it is not necessarily entirely arbitrary. Analogy may play an important role in expressing new ideas with existing words. Drawings and three dimensional models are somewhere in between. They visually reflect the material form of what is represented but do not have the same causal force. Analogy is now the main way in which the meaning of the representation becomes clear to an observer.

3.2 New Experiments

The ideas sketched in the previous subsection have informed a series of experiments [Steels, 2001] which all have the same structure:

1. We set up a population of agents by introducing several robots in the same environment and by ‘teleporting’ software states into these robot bodies so that we can have very large (sometimes thousands of) agents. This way we can experiment with populations (which is absolutely necessary to explore collective dynamics) without giving up the situatedness and embodiment. We have used many different types of robot platforms, including the Lego robots used in the self-sufficiency experiment discussed earlier [Steels and Vogt, 1997], movable cameras in the ‘Talking Heads’ experiment [Steels et al., 2002] 5, and the Sony AIBO robot [Steels and Kaplan, 2002].

2. The agents engage in interactions with each other thus establishing a collective dynamics. For example, one agent draws the attention to an object in the environment. This can be achieved without verbal communication, for example by

Figure 5: The Talking Heads experiment was a large-scale experiment in which a population with thousands of agents used movable cameras to capture images of the environment and play language games about them

gesturing, or with a representation, namely by exchanging a sound or a structured sequence of sounds. When the interaction involves verbal elements we call it a language game. Agents take turns being speaker and hearer.

3. In the beginning of the experiment, the agents do not have any shared external representation system (i.e. a lexicon) nor ways of categorising and conceptualising reality as semantic basis for constructing external representations. If an interaction fails, they expand their categorial repertoire, develop new conceptualisations, or invent a new word or grammatical construction. The 'speaker' is the one that chooses what to say and may invent new representational conventions while doing so. He will however try to maximise communicative success by keeping a score of the success of each construction in the past and choosing constructions with the highest success. The 'hearer' either already shares the same convention and the interaction succeeds, or the hearer attempts to detect the conventions and conceptualisations used by the speaker. So, as part of each interaction the agents adapt their internal structures to become better communicators in the future.

The following phenomena could be observed in these experiments (see references for technical details).

1. The population of agents have been shown to establish a communication system based on the exchange of symbolic representations (words or syntactically structured sequences of words). Communicative success can be measured and

typically rises rapidly (depending on the size of the population and the number of meanings to be expressed) [Steels et al., 2002].

2. The main mechanism by which conventions become shared is self-organisation, applied in a way that is similar to the flocking and path formation shown on autonomous robots [Mataric, 2003]. There are random fluctuations in the sense that different agents construct and choose ways of expressing something partly based on random choices or choices influenced by their individual developmental history. But there is a positive feedback loop between use and success. Conventions which are used more have more success and will therefore be chosen more. Consequently a winner-take all effect is seen where one convention progressively dominates (see figure 6).

3. Agents feature mechanisms (such as radial basis function networks or the random expansion of discrimination trees) which expand or adapt their categorisations and conceptualisations. These mechanisms operate on an individual basis and there is no guarantee that agents who do not interact with each other arrive at the same repertoire of categories. Typically they do not. What we have seen however is that these repertoires nevertheless can become shared through communication, namely due to a structural coupling between conceptalisation and verbal expression.

4. Finally we observe evolution, not at a genetic but at a cultural level. For example, a word may become associated with a new meaning, a word that had multiple meanings may become disentangled so as to have one meaning, a meaning may become expressed with another word, etc. All this is due to the inevitable stochasticity of situated embodied language games and the influx of new members in the population [Steels and Kaplan, 1998].

These experiments show how autonomous agents can construct their own external representations (particularly in the medium of sound) and how these representations may become conventionalised and thus shared in a group, without prior design nor any global control. On the other hand, this is only the first part of the story. What we still need to show is how these external representations may led to the significant bootstrapping effect that we see in human development, where representations (drawings, language, pretend play) are a primary motor of cognitive development. This can only be achieved by using the representations emerging from communication for other tasks, including ‘knowledge-based’ tasks, such as planning, thinking, or memory. These are tasks which have so far remained firmly in the province of symbolic A.I. but the path sketched here suggests a new route to approach them.

4 Conclusions

The behavior-based approach to A.I., which has its direct roots in a revival and reconstruction of the early cybernetics work so well illustrated by Grey Walter, has lead to a

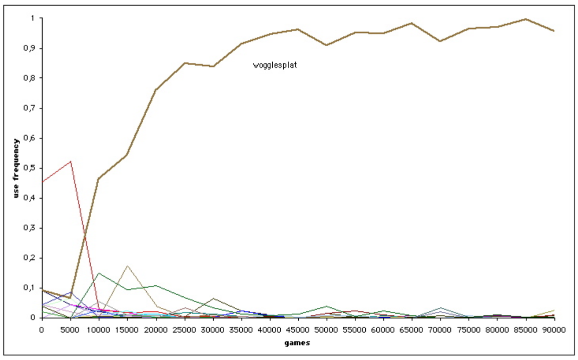

Figure 6: A meaning-form diagram which graphs for a specific meaning all the possible forms and their frequency of use. A winner-take-all situation is clearly observed. The x-axis shows language games and y-axis the frequency of particular forms.

new practice for building autonomous robots. This new practice emphasises embodiment and a tight coupling between the behavior of the robot and the environment. The behavior is realised by a network of dynamical behavior systems. It emphasises adaptation, evolution, and self-organisation as ways of generating complex behavior from simple components.

This paper reported on one of the early experiments in behavior-based A.I. and then sketched a line of research which attempts to go beyond physical behavior towards cognition via an exploration of adaptive (external) representations. It turns out that the same biologically-inspired mechanisms which have informed behavior-based robotics are relevant for evolving adaptive representations. Much remains to be done of course and there are many open questions remaining, but results already achieved show a new fruitful path towards embodied artificial intelligence.

Acknowledgement

Many people have been involved in the projects briefly reported here. David McFarland has had a tremendous influence during his many years spent as a visiting scientist at the VUB AI-lab. Another visiting scientist, Tim Smithers brought to the lab the early lego-vehicle experiments from Edinburgh. In the late nineteen eighties, Rodney Brooks and Maja Mataric transferred some early behavior-based technologies to our lab, during their own extended visits. Rolf Pfeifer, who was also a visiting scientist, has made many important contributions for the conceptual foundations of the behavior-based paradigm. Finally, Kerstin Dautenhahn infused the experiments with her characteristic creativity and biological grounding.

The chief developers of the robots and the experimental setup of the first ecosystem experiments were Danny Vereertbrugghen and Peter Stuer. The second wave of experimentation was primarily done by Andreas Birk, Holger Kenn, and Tony Belpaeme, who built new versions of the robot's brain brick, a vision based object detection system, and an environment for monitoring the course of an experiment. Paul Vogt's Ph.D. thesis was a major milestone for communication experiments on the legorobot platform. Much of all this work was financed by a IUAP and GOA grant from the Belgian government. The Talking Heads experiment was another large collective effort, mainly started and set up at the Sony Computer Science Laboratory in Paris in collaboration with Frederic Kaplan and Angus McIntyre, and at the VUB AI Lab with Joris Van Looveren and Tony Belpaeme. Frederic Kaplan has been the main force in the more recent experiments in language games with the AIBO. Marleen Wynants' comments have improved the present paper.

References

[Alexander et al., 1977] Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., and Angel, S. (1977). *A Pattern Language.* Oxford University Press, New York.

[Arkin, 1998] Arkin, R. (1998). *Behavior-Based Robotics*. MIT Press, Cambridge, MA.

[Birk, 1998] Birk, A. (1998). Robot learning and self-sufficiency: What the energy-level can tell us about a robot's performance. In Birk, A. and Demiris, K., editors, *Learning Robots.*, pages 451–461. Springer Verlag, Berlin.

[Braitenberg, 1984] Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.

[Breazeal, 1998] Breazeal, C. (1998). A motivational system for regulating human-robot interaction. In *Proceedings of AAAI 98*, pages 436–477. AAAI Press, Menlo Park Ca.

[Brooks, 1991] Brooks, R. (1991). New approaches to robotics. *Artificial Intelligence*, 47:139–159.

[Camazine et al., 2001] Camazine, S., Deneubourg, J.-L., Franks, N., Sneyd, J., Theraulaz, G., and Bonabeau, E. (2001). *Self-Organization in Biological Systems.* Princeton University, Princeton, NJ.

[Clancey, 1997] Clancey, W. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge University Press, Cambridge, UK.

[Fujita and Kitano, 1998] Fujita, M. and Kitano, H. (1998). Development of an autonomous quadruped robot for robot entertainment. *Autonomous Robotics*, 5:1–14.

[Gombrich, 1963] Gombrich, E. (1963). *Meditations on a Hobby Horse and other essays on the theory of art.* Phaidon, London.

[Keller, 1994] Keller, R. (1994). *On Language Change. The Invisible hand in language.* Routledge, London.

[Kuroki et al., 2001] Kuroki, Y., Ishida, T., Yamaguchi, J., Fujita, M., and Doi, T. (2001). A small biped entertainment robot. In Hashimoto, S., editor, *Proceedings of the IEEE-RAS International Conference on Humanoid Robots.*, pages 181–186. Waseda University, Tokyo.

[Langton, 1995] Langton, C. (1995). *Artifical Life: an overview*. The MIT Press, Cambridge, MA.

[Mataric, 2003] Mataric, M. (2003). Designing and understanding adaptive group behavior. *Adaptive Behavior*, 4 (1):51–80.

[McFarland and Bosser, 1994] McFarland, D. and Bosser, T. (1994). *Intelligent Behavior in Animals and Robots.* MIT Press, Cambridge Ma.

[Melhuish, 2001] Melhuish, C. (2001). *Strategies for Collective Minimalist Mobile Robots.* Professional Engineering Publishing Ltd., London.

[Nilson, 1984] Nilson, N. (1984). Shakey the robot.

[Pfeifer and Scheier, 2000] Pfeifer, R. and Scheier, C. (2000). *Understanding intelligence*. MIT Press, Cambridge Ma., Boston, MA, USA.

[Piaget, 1970] Piaget, J. (1970). *La Formation du Symbole Chez L'Enfant*. Delachaut et Nestlie.

[Prigogine and Stengers, 1984] Prigogine, I. and Stengers, I. (1984). *Order Out of Chaos*. Bantam Books, New York.

[Rosenfield, 1988] Rosenfield, I. (1988). *The Invention of Memory: A New View of the Brain.* Basic Books, New York.

[Russell and Norvig, 1995] Russell, S. and Norvig, P. (1995). *Artificial Intelligence. A Modern Approach.* Prentice-Hall, Englewood Cliffs.

[Steels, 1950] Steels, L. (1950). Discovering the competitors. *Journal of Adaptive Systems*, 2(4):173–199.

[Steels, 1989] Steels, L. (1989). Connectionist problem solving: an ai perspective. In Pfeifer, R., Schreter, Z., Fogelman-Souli, F., and Steels, L., editors, *Connectionism in Perspective*, pages 215–228. Elsevier Science Publishers (North-Holland), Amsterdam.

[Steels, 1991] Steels, L. (1991). Towards a theory of emergent functionality. In Meyer, J. and (eds.), S. W., editors, *From Animals to Animats. Proceedings of the 1st SAB Conference*, pages 451–461. MIT Press, Cambridge Ma.

[Steels, 1994] Steels, L. (1994). A case-study in behavior-oriented design. In Edmonds, B. and Dautenhan, K., editors, *Proceedings of the 3d SAB Conference*, pages 4–16. MIT Press, Cambridge Ma.

[Steels, 2001] Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent systems*, pages 17–22.

[Steels, 2003] Steels, L. (2003). Language re-entrance and the 'inner voice'. *Journal of Consciousness Studies*, 10 (4).

[Steels and Brooks, 1994] Steels, L. and Brooks, R. (1994). *The 'Artificial life' Route to 'Artificial Intelligence'. Building Situated Embodied Agents.* Lawrence Erlbaum Ass, New Haven.

[Steels and Kaplan, 1998] Steels, L. and Kaplan, F. (1998). Stochasticity as a source of innovation in language games. In Adami, C., Belew, R., Kitano, H., and Taylor, C., editors, *Proceedings of Artificial Life VI*, pages 368–376, Cambridge, MA. The MIT Press.

[Steels and Kaplan, 2002] Steels, L. and Kaplan, F. (2002). Aibo s first words. *Evolution of Communication*, 4(1).

[Steels et al., 2002] Steels, L., Kaplan, F., McIntyre, A., and Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In Wray, A., editor, *The Transition to Language*. Oxford University Press, Oxford, UK.

[Steels and Vogt, 1997] Steels, L. and Vogt, P. (1997). Grounding adaptive language games in robotic agents. In Husbands, C. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life.*, pages 451–461. MIT Press, Cambridge Ma.

[Walter, 1950] Walter, W. (1950). An imitation of life. *Scientific American*, 2(185):42–45.