The origins of intelligence.

Luc Steels
Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
E-mail: steels@arti.vub.ac.be
In: Proceedings of the Carlo Erba Foundation
Meeting on Artificial Life.
Fondazione Carlo Erba. Milano 1996.)

March 12, 1996

Abstract

Although substantial progress has been made on the question of the origin of life, less progress can be seen concerning the origins of intelligence. There is not even general agreement of what intelligence is. The paper proposes a definition of intelligence grounded in biology, which makes the question of the origins of intelligence seem more approachable. It then identifies two major transitions that must have been crucial in the development of intelligence: the origins of 'general purpose' neural networks and the origins of language. Some experimental work is reported that tries to recapitulate these major transitions using an artificial life perspective.

1 Introduction

Where does intelligence come from? How can we explain that in a physical world populated by living systems, the capacity that we call intelligence developed? Astonishingly enough, we have hardly any theory about this. Science has developed reasonable, although still debated, theories of the origin of the universe, such as the Big Bang theory. There are also theories of the origins of galaxies, of the earth and the moon, and of geological structures. There are theories of the origin of life, the diversity of species, and the origin of Man. So, why don't we have a theory of the origin of intelligence.

The reason is partly that for many people no such theory is needed. Mind is eternal, they say, belonging to a Platonic universe. Seeking an explanation for its origins is therefore absurd. Such a Platonic view is still common today with mathematicians like Penrose [15]. However it is not a scientific explanation. It is similar to the earlier view that the origin of the universe needs no explanation because it has always been there and will always be there, or that all the different species, including Man, were created in a few days by an omniscient being. If we want a scientific theory of the origins of intelligence, we must close the gap between the basic laws of physics and biology and theories of intelligence. Right now the gap is enormous and it can only be closed by working from both sides.

This paper raises a few issues and provides some directions and experimental approaches for addressing the question of the origins of intelligence. No clear definite answer can be given yet, although a way can be pointed out. The first section defines intelligence as a continuum with current biological views of living systems. It is only by having such a definition that we can hope to pinpoint precisely where intelligent systems outgrow living systems.

2 Defining Intelligence

Traditional definitions of intelligence involve a strong subjective component. For example, Turing has defined intelligence operationally by an experiment in which a human tries to identify whether he is interacting with a computer program or a real human being. If this distinction is not possible, the program is assumed to be intelligent. Newell [14] has defined a system to be intelligent if knowledge-level descriptions, beliefs, and intensions can be ascribed to it. Both definitions are not only subjective because they rely on human judgement but also ignore the embodied nature of human intelligence and the function of intelligence in survival.

This section proposes an alternative definition of intelligence which seeks to establish a continuum with life. It first identifies the class of evolving complex adaptive systems, then identifies progressively more complex instances from chemical systems to living systems, and then to intelligent systems.

2.1 Evolving Complex Adaptive Systems

Let us delineate a class of systems with four defining characteristics: self-maintenance, adaptivity, information preservation, and spontaneous increase of complexity. I propose to call such systems *evolving complex adaptive systems*. Living systems are an obvious subset but there are already autocatalytic chemical reactions with the same properties and intelligent or cultural systems could be seen as other examples.

- *Self-maintenance*: Self-maintenance means that the system is actively establishing itself. To avoid annihilation due to increased entropy, the system needs to constantly rebuild itself by drawing materials from the environment and establish a boundary between itself and the rest of the environment. Maturana and Varela have called this process autopoiesis [11]

- *Adaptivity*: The system is not only capable to maintain its own internal equilibrium for a constant environment, but also adapts when there are (small scale) changes to the environment in order to enhance its chances of further existence.

- *Information preservation*: The information defining the system is capable to be perserved so that the system does not depend on the continued existence of its components to survive. It is the *role* of the components that keeps the whole system together and if the various roles and their interrelations are preserved the whole system is preserved.

- *Spontaneous increase in complexity*: The most remarkable aspect is that the system is able to increase its own internal complexity. This could mean that there are increasingly more parts, more complex relations between parts, more complex behaviors of the parts, etc. Moreover often instances of the same system come together to form a larger whole that operates as a single system evolving complex adaptive system at a higher level.

We can identify different instantiations of this basic class of evolving complex adaptive systems, where each instantiation builds further upon the previous instantiations but adds more powerful machinery so that self-maintenance and adaptivity is more successful, information is better preserved and the growth of complexity becomes faster. Each time a major transition has been responsible for shifting to the next level of complexity, but the new level then 'slaves' the level below, or we can at least see a kind of co-evolution towards greater complexity of both. The major instantiations are (1) autocatalytic chemical reactions, (2) living systems, (3) intelligent systems, and (4) cultural systems. Conglomerations of these systems (groups of co-evolving reactions, species, colonies, societies) form in themselves evolving complex adaptive systems with their own dynamics.

1. Autocatalytic chemical reactions (uncoded life)

The various properties of evolving complex adaptive systems can already be seen in certain types of chemical reactions which are known as pre-life or uncoded life systems [8]:

- The reactions achieve self-maintenance by being autocatalytic. The substances to start the beginning of the reaction are regenerated, often after a long cycle and in larger quantities, so that the whole reaction chain can start again and proliferate. In some cases it is possible to show that boundaries form themselves [9].

- These reactions can be shown to be adaptive to changes in the environment. For example, the rate may slow down when temperature conditions change or when materials are less abundantly present. In some cases there are conditional pathways depending on the conditions in the environment.

- Autocatalytic reaction networks preserve information by making copies of themselves (with potential errors). Such copying has been synthesised in the laboratory.

- Autocatalytic reactions have recently been shown to be able to undergo evolution by natural selection, known in this case as molecular evolution. It is enough that there is a reaction that is autocatalytic and that variations occur in replication. When the environment (in this case the other chemicals present) provide selectionist pressures, then there is an evolution towards more complex molecules or reaction pathways that are capable to cope better with the selection pressures.

2. Living systems

Living systems clearly have all the properties of evolving complex adaptive systems. They most probably originated out of autocatalytic chemical reaction networks but achieve the characteristics of evolving complex adaptive systems differently:

- The simplest living systems (such as unicellular organisms) use metabolic pathways enclosed in cell membranes to maintain themselves while drawing materials from the environment. More complex living systems exhibit a much wider behavioral repertoire because groups of cells form organs with complex coordinated functions.

- Adaptivity is now not only achieved using chemical means but by changes in behavior, such as heavier breathing when oxygen content is lower or slower movement when it is very hot. Behavior is controlled using special-purpose neural networks.

- The most important innovation is however the preservation of information by coding the system in terms of genes. This requires the 'discovery' that proteins can function as interpreters of a code [5]. The code itself, in the form of DNA, is now copied as opposed to the whole organism. Additional proofreading while copying assures that much more complex information can be preserved, not only for creating the next generation of an individual but also for regenerating constantly parts of a single individual.

- The genetic mechanism provides also a much more powerful way to generate more complexity. The code is mutated or combined via crossover operations and then subjected to naturally occurring selection. A larger search space of possible life forms can thus be explored and it becomes easier to build further upon existing complex forms. Other ways are used to increase complexity as well, they include level formation and self-organisation. Based on these principles living systems have shown several transitions towards ever greater complexity. Recent overviews of the important transitions have been given by Maynard-Smith and Szathmary [12] and de Duve [3].

3. Intelligent systems

Intelligent systems can be defined as systems that have the same four properties (self-maintenance, adaptivity, information preservation, and increase in complexity) but use other means to achieve them. It is not yet completely obvious where the key lies, but two things are surely important:

- Neural networks, which initially were completely specific, have become general purpose structures which can store a large number of complex behavioral patterns, sustain processes for interpreting signals from the world and controlling at a fine grained level complex action patterns. Most importantly these networks and processes develop and adapt themselves continuously and very fast (compared to genetic evolution).

- At some point a symbolic capacity has developed: This is the ability to interpret the world in terms of concepts, to represent states of the world using these concepts, and to perform symbolic reasoning by manipulating these representations. This symbolic capacity also sustains symbolic learning.

These features result in superior capacity for all the four properties of evolving complex adaptive systems. Self-maintenance is enhanced by the ability to handle much more complex behavior, be responsive to much more environmental influences and control much more complex actuators (such as hands). Adaptivity is enhanced by the capacities of neural networks to acquire new knowledge and by symbolic learning. There is a vast increase in the amount of information that can be preserved compared with the genes. Finally there is a steady and fast build up of complexity, particularly during the developmental stages of the organism.

4. Cultural systems

It is useful to define yet another type of instance of the general class of evolving complex adaptive systems, namely cultural systems of which language is one of the main examples. Other examples are religious systems and social systems. These cultural systems appear to have their own internal dynamics which cause them to maintain themselves, adapt, preserve information and become more complex. Thus languages originate, develop, and sometimes die, like organisms. They are formed by the joined distributed action of millions of individuals that speak a language. Languages constantly adapt to changes in the meanings that users want to express and have become more complex to cope with the pressure of reducing cognitive overload for speaker and hearer as well as pressure to say more in a shorter time frame.

Interrelationships

There are complex interrelations between these different types of systems because one system is built on top of another one: living systems embody a multitude of autocatalytic reaction networks, intelligent systems have grown out of living systems, and cultural systems have evolved through intelligent systems. Often a 'slaving' relation can be observed between the different layers. For example, once living systems come into existence, they enslave autocatalytic reaction networks to become metabolic pathways under the control of the genes. Similarly, the vocal apparatus necessary for speech is an adaptation from the earlier vocal apparatus, observed in chimpanzees, which could not make so many distinctive sounds. The development of the vocal apparatus in Homo Sapiens results however in a high risk of choking, which was non-existent before, and a dislocation of certain teeth (wisdom teeth) which often have to be removed surgically. These two features are disadvantageous from a purely biological survival point of view but developed nevertheless to support the complexification of language.

The complex self-enforcing relations between levels is probably crucial for understanding intelligence. On the one hand, intelligent systems have grown out of living systems. On the other hand, intelligent systems are the substrate on which cultural systems have evolved. Language is a cultural system that clearly seems to have played a primordial role in pushing intelligent systems towards the extreme plasticity that is known today. Cultural systems have their own dynamics which sometimes (as in the case of wars due to religious or nationalistic tendencies) enslaves the individuals to act against their own self-interest or the interest of others.

3 Steps towards intelligent systems

The problem of the origins of intelligence can now be posed with much greater precision. First of all we need to understand how neural networks, which initially were completely specific, could have become general purpose dynamical systems. Second we need to understand how an independent symbolic level could have emerged. For none of these questions is there a plausible answer today. This section provides some more discussion and then sketches possible technical and experimental approaches.

3.1 The plasticity of the neural substrate

We need to find the major transitions through which neural networks, which were initially special-purpose and hence the subject of genetic evolution by natural selection, have become general-purpose and moldable by developmental and learning processes. There is so far no theory to explain this, partly because there is not yet an adequate theory that explains the plasticity of neural networks as such.

Two approaches have domininated research on plasticity so far. The first approach focuses on mechanisms that perform induction based on a large number of example behaviors, either supervised or unsupervised. Various neural network techniques have been proposed and applied with varying degrees of success [23]. This approach relies however on the prior availability of enough examples and assumes enough time to perform the inductive process. These conditions are seldom satisfied for agents having to stay viable in an unknown environment. The second approach performs a kind of genetic evolution. The behavioral networks are subjected to random variation by mutation and crossover and consequent selection [7]. This approach cannot work on a single agent that has to stay viable as it acquires new behavior, because it relies on the exploration of a population of agents of which most members will fail to survive and genetic evolution is in general too slow to be responsive to changes to the environment fast enough. When studying intelligent systems we seek precisely mechanisms which are not genetic.

Our own approach is selectionist in the sense that behavioral networks are generated independently of the environment in which they have to operate and then subjected to selectionist pressures. But the selectionist process takes place during the lifetime of the individual. Different variations are tested after each other. The exploration strategy must be such that the agent remains viable. The proposed mechanism has some strong relations between the 'neural darwinism' hypothesis of Edelman [4] which focuses however on the acquisition of categorisation competence as opposed to behavioral regulation for remaining viable in a challenging ecosystem.

In our laboratory we have set up a robotic ecosystem to experimentally investigate this selectionist development process. The ecosystem includes a

Figure 1: Robotic ecosystem as physically realised at the VUB AI laboratory. There are a number of robots which can recharge themselves in a charging station. There is competition for the energy flowing in the charging station in the form of lamps which can be put out by the robots by pushing against the boxes in which they are housed.

set of small robots which have about 20 sensors, a series of actuators including two motors driving left and right wheels, and their own computational capacity and batteries. The ecosystem is generic for situations where a developing agent is confronted with a growing population p of competitors for its energy resources. There is a steady inflow of these resources in the ecosystem to reach a level g. The developing robot has an internal energy level e which decreases due to normal bodily operation and active behavior. Internal energy can be replenished by recharging at a certain location. The resource availability at this location b is replenished from the globally available resources. The competitors which take the form of lamps, grow by consuming as well from the globally available resources. So the more competitors there are, the less resources will be available for the robot. But the robot can combat the competitors and thus maintain an adequate supply of resources for itself. The situation is moreover such that one robot cannot survive on its own. It needs to cooperate with other robots so that they take turns to work and recharge.

Given this ecosystem, a cyclic series of activities is observed in which the robot seeks the charging station where resources are available, consumes resources to replenish its battery, moves out of the area to seek out the

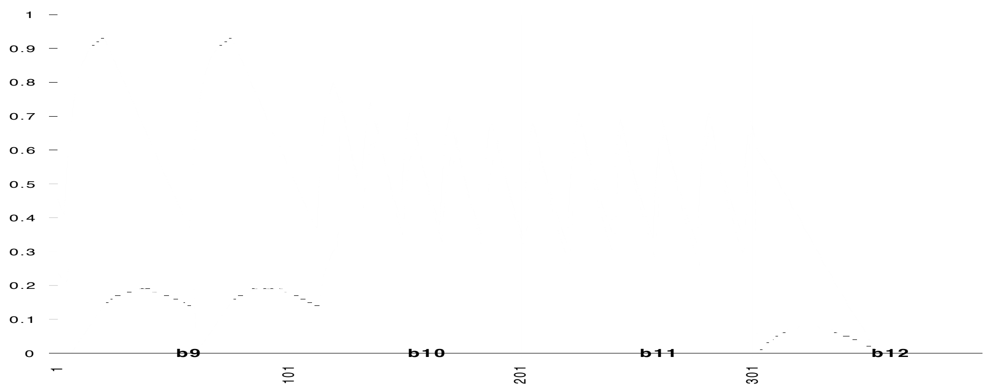

Figure 2: The different combinations of the forward movement motivation and the contact sensor are tried for 100 time steps each. Actual energy level and forward movement are shown. Only b9 (the first one) produces a significant improvement in performance.

competitors, and looks for the charging station again when its energy level is getting low. A developing robot should be able to learn these behaviors while interacting with the environment. At the start of the experiment, the robot performs a default random walk behavior which, due to the benign initial conditions, nevertheless results in viable behavior. The robot is assumed to have a variety of basic behavior systems in the spirit of the behavior-oriented approach [17]. Each behavior system is a goal-seeking feedback control system modulated by a motivational quantity. For example, for alignment to visible light, the goal is to have a zero difference between the left and right photosensors. The motivation is linked to energy deficit.

Figure 2. shows a snapshot of the learning process which takes the form of the successive exploration of different alternative behaviors by instantiating certain couplings between sensor or motivational quantities and motor quantities. In this example, the robot is trying a connection between a contact sensor which detects that the robot is in the charging station and forward movement. The robot 'discovers' that it should stop in the charging station. More details of these experiments can be found in [21], [22].

Although we have managed to get some build up of behavioral complexity, it is still too early to say that this kind of selectionist process gives the desired flexibility and plasticity that we see in intelligent behavior and then there still remains the problem how the dynamics implied by this scheme could have evolved out of special-purpose dynamical networks.

3.2 Language as the key of a symbolic layer

Classical symbolic AI systems exhibit great complexity for particular functionalities such as expert problem solving, chess playing, etc. But this functionality is typically completely programmed by hand based on an analysis of human competence. These systems are therefore frozen instances of intelligence as opposed to evolving adaptive intelligent systems. Moreover they mostly do not have any direct relationship to reality but need a human to interpret reality and supply the symbolic descriptions that are needed. Unfortunately not much progress can be seen yet on how the gap between subsymbolic and symbolic capacities should be bridged, nor on the question of the origin of the symbolic layer.

Most of the present work assumes that there are abstraction facilities in neural networks or a new higher level dynamics that may emerge. In our own work, we take a quite different approach. We assume that language has played a key role in the formation of a symbolic layer in human intelligence and therefore focus on experiments in which the origin of language could take place. We are exploring two hypotheses:

[1] Language is an autonomous adaptive system which forms itself in a self-organizing process. Language is therefore similar to other self-organizing phenomena observed in biosystems, such as paths in an ant society, clouds of birds, etc. A language is viewed as an adaptive system in the sense that it has to allow its users to express an open-ended, ever growing or changing set of meanings with an open-ended but finite set of building blocks and combinations of building blocks. The speakers and hearers are distributed agents that through their localised linguistic behavior (namely the carrying out of conversations) shape and reshape the language. No agent has a complete view of the language and no agent can control the linguistic behavior of the whole group. Moreover no separate mechanism for language acquisition is necessary because the mechanisms that explain the origin of language also explain how it is acquired by new agents entering the community.

[2] Language spontaneously becomes more complex based on the same mechanisms that give rise to complexity in biosystems in general. The development and evolution of language is primarily driven by the need to optimize

communicative success and handle the very strong constraints which hold for open-ended real world languages, namely limited time to communicate, limited time to process the utterance, weak and error-prone acoustic transmission, limited feedback about success, constraints of the vocal apparatus, etc.

Self-organization is a common phenomenon in certain evolving complex adaptive systems. To support self-organization a system must exhibit a series of spontaneous fluctuations and a feedback process that enforces a particular fluctuation so that it eventually forms a (dissipative) structure [16]. The feedback process is related to a particular condition in the environment, for example an influx of materials that keeps the system in a non-equilibrium state. As long as the condition is present, the dissipative structure will be maintained. Some standard applications of self-organization can be seen in morphogenetic processes, or the formation of a path in an ant society or a termite nest [2].

A language can be viewed as a dissipative structure similar to a path in an ant society. Each agent is assumed to create and continuously change his own language in a random fashion, resulting in a fluctuating linguistic community. Language must be shared in order to obtain the benefit of cooperating through communication. Hence the changes are coupled to communicative success: the higher the success the less probable a change. This results in a feedback process. When more agents use the same word for the same meaning, communicative success increases and therefore the word-meaning association becomes more stable. It has been shown that coherence indeed emerges [18].

In one experiment (reported in [19]), agents developed spontaneously and autonomously a vocabulary to talk about themselves and identify spatial relations among themselves. Here is an example dialog where the object is introduced by a-25 using a spatial description (straight in front of me) expressed as 'b u v a j a' and confirmed by a-23 using another spatial description (behind me to the left) expressed as 'b a t u l o'.

```
Dialog 1142 with a-25 a-23
 => a-25: (a-25)
   a-25: a-25  -> (B U)  <- a-23: a-25
   a-25: FRONT  -> (V A)  <- a-23: FRONT
   a-25: STRAIGHT  -> (J A)  <- a-23: STRAIGHT
```

Figure 3: The results of a typical experiment with 10 agents, 5 possible words, and 1 meaning. It plots the communicative success of each word (y-axis) over time (x-axis). We see a search period in which different words compete until one gains complete dominance.

```
=> a-23: (a-25)
  a-23: a-23  -> (B A)  <- a-25: a-23
  a-23: BEHIND  -> (T U)  <- a-25: BEHIND
  a-23: LEFT  -> (L O)  <- a-25: LEFT
=> a-25: (a-25)
  a-25: confirm -> 'yes' <- confirm
```

Additional experiments are currently being performed to explore issues like ambiguity and semantic indeterminacy [20], the formation of morphological and syntactic structures, the indirect mapping of meanings to words (where one word may capture many different meanings), the emergence and handling of ambiguity, the grounding of language in robotic agents, the creation of new meaning, etc. Through such experiments we expect to understand better how complex symbolic representations form themselves. Similar processes must be going on internally in the brain to form and shape the mental language in which knowledge is expressed, although we have not yet carried our experiments in this area.

4 Conclusions

Much remains to be discovered before a theory of the origins of intelligence can begin to take shape. We need to understand how the brain is capable to exhibit its remarkable behavioral plasticity and how a symbolic layer might have emerged. Experiments using software agents and robotic agents appear to be a very difficult but at the same time rewarding way to pose questions, imagine possible answers, and experimentally test them.

5 Acknowledgement

The research and writing of this paper has been financed by the Belgian Federal government FKFO project on emergent functionality (FKFO contract nr. G.0014.95) and the IUAP 'Construct' Project (nr. 20) of the Belgian government, with additional support from the external researcher program of the Sony Computer Science Laboratory in Tokyo. The thoughts expressed in this paper have benefited from a number of lively discussions at the Cortona Woche organised by Prof. Luisi from the ETH, Zuerich in Cortona Italy, september 1995.

References

[1] ASHBY, R.: Design for a brain. Chapman and Hall., London. 1952.

[2] DENEUBOURG, J-L.: Application de l'ordre par fluctuations a la description de certaines etapes de la construction du nid chez les termites. Insectes Sociaux, Tome 24, 2, p. 117-130. 1977.

[3] DE DUVE, C. Vital Dust: Life as a Cosmic Imperative. Basic Books, New York. 1995.

[4] EDELMAN, G.: Neural Darwinism: The Theory of Neuronal Group Selection. Basic Books, New York. 1987.

[5] EIGEN, M. ET.AL. The origin of genetic information. Scientific American, 244, 88-118. 1981.

[6] HOLLAND, J.H. Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, Michigan. 1975

[7] CLIFFS, D, P. HARVEY AND P. HUSBANDS Using genetic evolution for autonomous agents. In: Journal of Robotics and Autonomous Systems, 1995, 1-2. North-Holland Pub., Amsterdam. 1995.

[8] KAUFFMAN, S.A. The origins of order: self organization and selection in evolution. Oxford University Press, Oxford. 1993.

[9] FLEISCHACKER, G., S. COLONNA, L. LUISI: Self-production of supramolecular structures. From synthetic structures to models of minimal living systems. Kluwer Academic Publishers, Dordrecht. 1994.

[10] MARGULIS, AND T. TOFFOLI: Cellular Automata Machines. MIT Press, Cambridge Ma. 1987.

[11] MATURANA, H.R. AND F.J. VARELA The Tree of Knowledge: The Biological roots of Human Understanding. Shamhala Press, Boston. 1987.

[12] MAYNARD-SMITH, J. AND E. SZATHMARY: The major transitions in evolution. Oxford University Press, Oxford. 1994.

[13] MCCLELLAND, J.L. AND D.E. RUMELHART (EDS.) Explorations in Parallel Distributed Processing. The MIT Press/Bradford Books, Cambridge Ma. 1986.

[14] NEWELL, A. The Knowledge Level. Artificial Intelligence Journal. Vol 18. p. 87-127. 1982.

[15] PENROSE, R. The Emperor's New Mind. Oxford University Press, Oxford. 1990.

[16] PRIGOGINE, I. AND I. STENGERS: Order Out of Chaos. Bantam Books, New York. 1984.

[17] STEELS, L. AND R. BROOKS, EDS. Building Situated Embodied Agents. The Alife route to AI. Lawrence Erlbaum Assoc., New Haven. 1995.

[18] STEELS, L.: Self-organizing vocabularies. Fifth Conference on Artificial Life. Nara. Japan. The MIT Press, Cambridge Ma. 1996.

[19] STEELS, L.: A self-organizing spatial vocabulary. Artificial Life Journal. Vol 3. 1996b.

[20] STEELS, L.: Emergent adaptive lexicons. Proceedings of the Simulation of Adaptive Behavior Conf. The MIT Press, Cambridge Ma. 1996b.

[21] STEELS, L.: A selectionist mechanism for autonomous behavior acquisition. Journal of Robotics and Autonomous Systems. Vol. 16. North-Holland Pub., Amsterdam. 1996d.

[22] STEELS, L.: Discovering the Competitors. Journal of Adaptive Systems. 4(2) 1996. The MIT Press, Cambridge Ma. 1996e.

[23] TORRAS, C.: An overview of neural network learning. In: Steels, L., ed.: The biology and technology of intelligent autonomous agents. Springer-Verlag, Berlin. 1995