Candidate No: 89879

CONCEPTS OF INTELLIGENCE

AMI Essay
EASY MSc
Department of Informatics
University of Sussex - Brighton - United Kingdom

1 ABSTRACT

Within this paper I aim to provide a brief overview of Knowledge-based systems and Behaviour-based systems, examining the adequacy of each as a model of intelligence. I also hope to highlight, clarify and tentatively extend some more recent work which is now beginning to narrow the gap between the two.

2 GOOD OLD FASHIONED KNOWLEDGE

In theory...

In 1936, Alan Turing proposed a thought experiment which would eventually propel the world into a brand new technological age. The Universal Turing Machine was an imaginary symbol manipulating device proven to be programmable to perform any conceivable computational function. For the first time, a single machine could be used to calculate anything that was logically calculable. This was the birth of the modern day computer and a founding moment for the field of Artificial Intelligence (AI).

A function is a rather simple mathematical concept. It receives a number of inputs and, using these, performs a sequential, mathematical process to produce an appropriate output. For example, the function of "circumference", takes one input (the radius of a circle), processes it (by multiplying it by two, and then multiplying it again by pi), and outputs the answer (the circumference).

Early practitioners of AI felt that higher functions could also be devised using the same logical mechanisms. For example, the function of "classify animal" would be given a suitable set of inputs ("presence of hair", "presence of backbone", "number of legs" etc.) and by following a properly formulated algorithm (e.g. if it has feathers, it is a bird) the function would output an appropriate "biological classification" for any given animal. Similarly, a function "diagnose", would conceivably take a suitable set of inputs (body temperature, presence of rash, nausea) and produce an appropriate diagnosis of a patient's illness [4]. Such "expert systems" were well studied in the 1970's and were the cutting edge of Artificial Intelligence research; a computer system, it was claimed, would someday take over the job of the local doctor, it would be more consistent and more reliable than a mere human, and it would possess the combined knowledge of all of the doctors in the world.

At the same time, similar approaches were taken in other branches of AI. In robotics a mobile computer would simply possess a function "move to doorway". This function would have access to a number of symbolic inputs (position in room, size and shape of room, position of obstacles), and would process this information to output the desired trajectory for successfully navigating towards the door. A higher function of "make a cup of tea" could simply be split into smaller functions of "move to kitchen", "get cup", "get tea bag", "boil water" etc. If a robot were built that could be asked to "go and make a cup of tea", surely it would possess a level of intelligence greater than that of a humble ant.

Early AI practitioners felt confident that given the right knowledge and the right logical program a computer would someday be created that would equal[1] humans in both intelligence and action.

Or so the story went. In practise however, things didn't work out so simply and the field, besieged from all sides, was doomed to fail.

… and in practice

First, philosophers questioned the feasibility of such models with relation the "harder problems" of intelligence. John Searle's infamous Chinese Room [1] questioned the capacity for a symbol manipulator to "understand" the symbols that it was manipulating. The deeper question of subjective experience still poses a major problem to the field, with most practitioners choosing to avoid the consciousness debate altogether; or else throwing it aside as a causally ineffective epiphenomenon.

In a similar attack against the functionalism of intelligence, mathematicians pointed to mathematical proofs such as Gödel's theorem. This theorem is apparent to a human observer but which could never be proved by any algorithm [2]. If a human can solve such a problem and yet a computer cannot, then a human cannot be a computer[2].

---

[1] Perhaps even surpass

[2] This is a general point still debated today; one counter being that "understanding" a problem is not the same as solving it.

From the standpoint of scientific plausibility[3], biologists were quick to scorn the lack of natural semblance within the models. A brain is a distributed array of millions of simple neurons, all firing in parallel; a computer, on the other hand, performs sequential processing within discrete functional units. As Francis Crick is claimed to have said, "*the brain isn't even a little bit like a computer*" [3].

However, the most damaging blows came from within the field itself. What seemed like a good idea in theory, turned out to be impossible to implement in practice. Expert medical diagnoses, it turns out, requires a lot more than just the following of a simple set of rules about what diseases have what symptoms [4]. It requires a "human" insight, the ability to judge a patient's behaviour, their mannerisms, and their past history. A lot of diagnoses is based not of rules, but on hunches and on years of experienced "gut instincts". It is not a passive job of rule following, it is a complex pursuit of creative problem solving within a dynamic and innovative field full of change. The collation and logical programming of such knowledge remains to this day an impossible task of Herculean proportions.

But building the "thinking processes" was only a half of it. It was soon discovered that obtaining the information that you wanted to think about was doubly hard. This became most striking within the field of Computer Vision; where building internal "representations" of even the most simple block worlds was found to be nearly impossible. Things that were simply taken for granted suddenly became the most unwieldy of engineering problems. Shadows; partial obscurance (objects that are partly hidden by other objects); invariance (the ability to recognise an object as the same object from a different perspective); suddenly there were a plethora of problems that simply couldn't be solved.

The final, and possibly most damaging, problem was "time". How can a system which needs SO MUCH computational power to process just a single static visual image ever cope within a fast, dynamic, real-time environment. More importantly, how can a "thinker" ever "choose" the appropriate action to take in such a dynamic world. By the time he has made a decision, the world model will have changed and thus have an effect the legitimacy of his decision; for clarity, imagine a robot trying to step onto a swinging bridge. The problem wasn't a lack of computer power, it was a lack of feasibility.

And so, these early Knowledge Based approaches (now somewhat unaffectionately dubbed Good Old Fashioned AI - GOFAI) went out of vogue; but its importance should not necessarily be cast aside with complete disregard. By bringing together mathematicians, biologists, philosophers and computer scientists a new kind of science was awakening; and the field of Artificial Life was born.

---

[3] Science is the study of reality; anything else just isn't science.

3 COMPLEX ADAPTIVE BEHAVIOUR

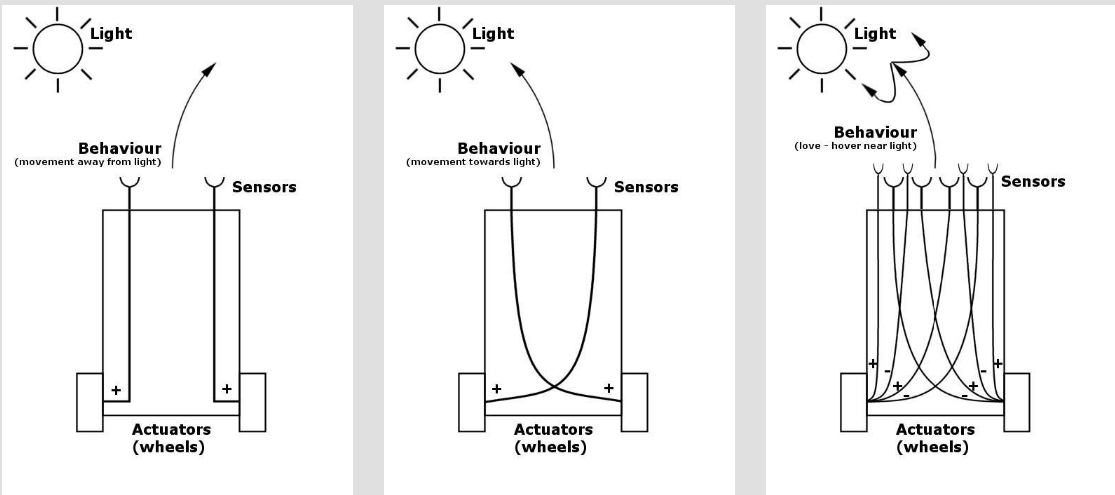
In theory…

Each of our senses (and we are now thought to have at least nine [5]) works in a similar way. The environment stimulates a collection of receptors which encode[4] the "information" into electrical impulses and transmit it to the central nervous system (CNS).

Once a signal enters the CNS it joins an already busy network of electrical impulses running around a huge maze of over 100 billion connected neurons. Within this maze, there are a number of efferent neurons which feed these impulses to motor neurons and, in turn, to muscle fibres. A coordinated innervation of muscle fibres causes arms to lift, hands to grab and mouths to speak.

The vast complexity of this system is breathtaking; but its underlying principle is quite straight forward. A continual flow of information is sent from an array of sensors, through a complex maze of connections, and eventually out to an array of actuators.

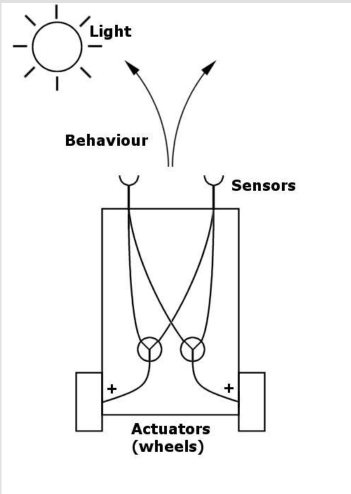
This basic principle is delightfully highlighted by Valentino Braitenberg in his book "*Vehicles*" [6]. In it, he describes a number of vehicles (agents) with light receptors (sensors) and wheels (actuators) directly connected through a mixture of excitation and/or inhibition (connections). Figure 1 shows three of his examples.



(a) **Fear** (b) **Aggression** (c) **Love**

(**Figure 1**) A selection of Braitenberg vehicles [6]

[4] This may simply be a proportional response to the strength of the stimulus or it may involve a more complex encoding through "cross-linking" or dampening as in the case of the eye or ear.



(**Figure 2**) A switched Braitenberg vehicle

We see that each of the connections results in a different set of "behaviours". To add a layer of complexity, we could add a timed switch to the connections (see figure 2), making the agent capable of alternating between two different behaviours (Braitenberg's fear and aggression). Continuing to increase this complexity further increases the behavioural repertoire of the agent.

This is the basic principle of behaviour-based systems (BBS); complex behaviours emerging from the purely reactive body of an agent.

Rodney Brooks, one of the earlier pioneers of this viewpoint, defines these notions more formally in terms of "*embodiment*" (an agent requires a body as it is the body itself which is the information processor) and "*situatedness*" (an agent requires an environment as it is the environment itself which is the information being processed) [7, 8].

This is in stark contrast with the Knowledge-based systems (KBS) described earlier, there is no need for any internal representations of the world, and there is no need for an internal "planner" working out how best to act. All actions are simply the causal response of a particular physiological system placed within an real environment.

This viewpoint proposes a completely different philosophical stance regarding the definition of intelligence. It states that intelligence is the *external* measure of the agent's behaviour rather than an *internal* measure of knowledge and reasoning. The more complex and appropriately "*adaptive*" this behaviour, the more intelligent the agent can be said to be. The implications of this altered philosophy are vast. It closes the Cartesian gap between mind and body by simply stating that there is no gap. The act of cognition (of thinking), is reduced to simply be the process of functional behaviour.

"To see" is simply the capacity of an agent to react to light. "To hear" is simply an agent's capacity to react to sound. These terms are scalar quantities; the greater the subtlety of reaction (behaviour) to light, the more the agent has been able to "see". Now imagine a Braitenberg vehicle with receptors connected to its wheels which are stimulated by damage or harm. A stimulus which highlights "bodily harm" and causes an agent to react adaptively (to reduce this harm) is considered to be exactly the same as the stimulus that we call *pain*. The subjective nature of these experiences isn't important here as they are regarded as epiphenomenal side effects to a purely reactionary system. Maybe "information" has two aspects, a physical and an experiential [20], but it doesn't really matter, as only the physical is considered to have any causal efficacy[5].

… in practice…

Unlike KBS, it is the practical results of the BBS approach which are probably its strongest advocate[6]. Recently, I had the great pleasure of building a Lego version of the simplest Braitenberg vehicle (figure 1a) myself. Unfortunately, the light sensors that I was going to use weren't working properly, so I instead used "proximity sensors"; rather than "fearing" light, my robot would "fear" obstacles. The robot (we affectionately named Thelma[7]) wandered around the laboratory merrily avoiding anything that got in its way. The whole build took about 30 minutes and required no complex programming. There was no vision parsing or planning mechanisms; the robot simply reacted to its environment, and with incredible adaptiveness. If a fellow robot builder tried to sabotage the results by sticking his foot out, or jumping into her path; Thelma would simply react by changing her course and wandering past them.

30 minutes with a few blocks of Lego produced an adaptive robot capable of "obstacle avoidance" and in any dynamic environment you cared to place it in (stairs were admittedly a bit of a problem). Years of KBS research would never have produced anything that could have even come close.

The precise physiology of an agent plays a crucial role in the behaviours it is capable of performing. For Thelma, the inappropriate positioning of the "proximity" sensors would result in her only being able to avoid objects within her "forwards" field of vision. Angling the sensors a little meant that she had a wider field and was able to react to obstacles that were slightly to her left and right.

Barbara Webb highlights the importance that "precise physiology" plays within natural systems too [9]. She examined why crickets are so unswervingly attracted to the sound of

[5] This is not my own view necessarily, but is the natural conclusion of the BBS approach

[6] It is perhaps interesting to ponder how much the practical success of the approach had on bringing about the acceptance of some of the its philosophical underpinnings

[7] The anthropomorphism of these robots is exceedingly difficult to avoid

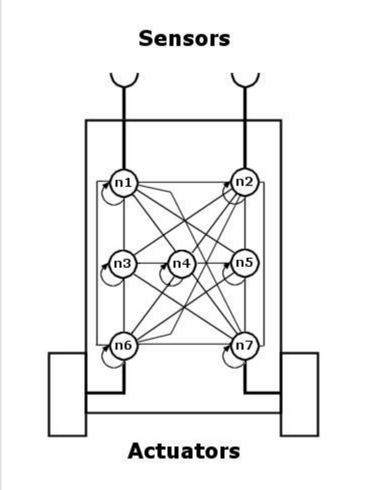
mating chirps and found that the precise form of a cricket's "ears" play an essential role. The physical shape of the ear is "tuned" to only transmit a neuronal impulse when presented with a mating chirp (all other sounds are dampened). This signal was found to be naturally strongest when the source of the sound (the mate) was directly in front of the cricket. In other words, like our Braitenberg behaviour of "light attraction", the cricket's physiology is such that its body will naturally perform the behaviour of "mating chirp attraction". This biological evidence of the BBS approach is something that the KBS approach severely lacked and is further evidence of its plausibility and legitimacy.

… but is it complete

Unfortunately, the BBS approach is no panacea. David Kirsch points out in his provocatively titled "*Today Earwig Tomorrow Man*" [10] that there are many aspects of intelligence (particularly human intelligence) which cannot be performed in a purely reactionary system. He suggests that "concepts" and "reasoning" are an essential aspect to more comprehensive tasks which require planning and scheduling. He cites chess playing, mathematical problem solving and replying to non obvious questions as a few examples; but actually the list is far longer than that. For example, the Jumping Spider, has been shown to scan the environment surrounding the web of potential prey before moving to capture the prey sometimes choosing an indirect route – this behaviour seems to indicate some sort of capacity for insight learning [11]. In Kirsches view, this behaviour would require some kind of an internal world of representation and logic.

By way of a response to this important critique, we need to introduce a further element to our Braitenberg vehicle; a Continuous Time Recurrent Neural Network (CTRNN). A CTRNN is a particular type of artificial neural network (ANN) and is meant (particularly in this case) to closely model the workings of the central nervous system. Each neuron (node) has a weighted connection to each of the others[8], and also to itself. Any node in the network may or may not be fed a signal from some sort of external input (sensor) and any node may or may not be used as an output (to trigger an actuator). Figure 3 below shows a Braitenberg vehicle with receptors and actuators connected via a CTRNN.

[8] This is to make the mathematics easier, and doesn't detract from the plausibility as a model of the brain. A weighted connection of ZERO is effectively the removal of a connection.



(**Figure 3**) A CTRNN Braitenberg vehicle

The value of each node changes over time using the below function:-

$$\dot{y}_i \equiv \frac{1}{\tau_i}\left(-y_i + \sum_{j=1}^{N} w_{ji}\sigma\left(y_j + \theta_j\right) + I_i\right)$$

The value of each node is affected by the changing values of its neighbouring nodes and thus, a CTRNN is (as well as being an ANN) also a type of *dynamical system*. Just like a Cellular Automata, or a Random Boolean Network it therefore possesses the potential processing capabilities of stable states, cyclical attractors and chaotic attractors [12]. In fact, dynamical systems which are "*tuned*" to a certain complexity are able to perform emergent computations of the same universality as the Universal Turing Machine [13]. In other words, there is no function that a computer can perform, which a (properly tuned) dynamical system can't.

Incorporating this new dynamic connection system into our Braitenberg vehicle vastly extends the complexity of behaviours which can be produced from it. A great example is shown in the results of an experiment performed by Evolutionary Roboticists Dario Floreano and Francesco Mondada [14]. They used a genetic algorithm (GA) to evolve the weights and biases for a robot's internal CTRNN[9]. The robot contained a rechargeable battery with a life of about 20 seconds and was placed within a simple arena containing a "charging plate". The robot's task, the fitness criteria for the GA, was to travel as far as possible without dying; in other words, the experimenters hoped to evolve a robot which would keep moving around the arena, whilst "keeping itself alive" by returning to the charging plate for energy. The resulting robot, after 240 generations, did just that. The

[9] It is not clear whether this was a CTRNN as I have defined it, or another slightly different style neural network.

robot, no matter where it was in the arena, would return to the charging plate with just seconds to spare. On investigating the neurons within the robot, it was shown that the firing of one neuron (n7) "*depended on the position and orientation of the robot in the environment*". The evolved dynamical system had found its own "*spatial representation of the environment"* [15].

Looking back to Kirsch's criticisms what does this tell us of the capacity of the BBS approach to perform the more complex planning and scheduling tasks? Well from a behaviourist's perspective, that's exactly what the robot did. Consistently, just in the nick of time, it changes its behaviour to return to the charging plate; a perfectly planned change of behaviour.

We can again call on nature for further evidence of this "emergent planning". A bee colony, looked at as a whole unit, forages for food in an exceedingly efficient manor. Given a number of sources of food the colony will concentrate its efforts on the most profitable sources first and then move onto the next profitable and so on. One could observe this behaviour as scheduled or planned but it is shown to "emerge" purely from the dynamics of the complex local interactions of the individual bees [16]. From a behaviourist's perspective, the intelligent action of "scheduling" and "planning" has been performed.

Whether or not these behaviourist responses fully answer Kirsch's critique depends on your particular viewpoint. Personally, I don't believe that notions of cognitive maps and emergent efficiencies are enough to cover the full dept of his concerns pertaining to the "logical manipulation of concepts" and in the remainder of this paper I aim to highlight a number of more recent ideas which might.

4 New Concepts

In theory...

Carlos Gershenson [17] recently published an intriguing paper in which he introduces the term "*Behaviour-based Knowledge Systems*" (BBKS). In it, he establishes the idea of a behaviour-based agent with the capacity to learn "concepts" about its environment through its embodied interaction with it. "*A kitten might play with a ball of paper to explore what can be done with it. Once the kitten [has experienced] the possibilities of sensation, perception and use, a concept representing the ball of paper shall have been created, so that the animal will behave accordingly in future presentations of balls of paper*". He further proposes that evolution has found a method by which these agents are able to perform a kind of internal "reasoning" based on these concepts; although how this "logical manipulation" is manifested is not made clear. This is a far cry from Brooks' view of "*Intelligence without Reason*" [8]. Although the agent is behaviour based, it is also able to act based on its own internal representations, or knowledge; a combination of external and internally driven behaviours.

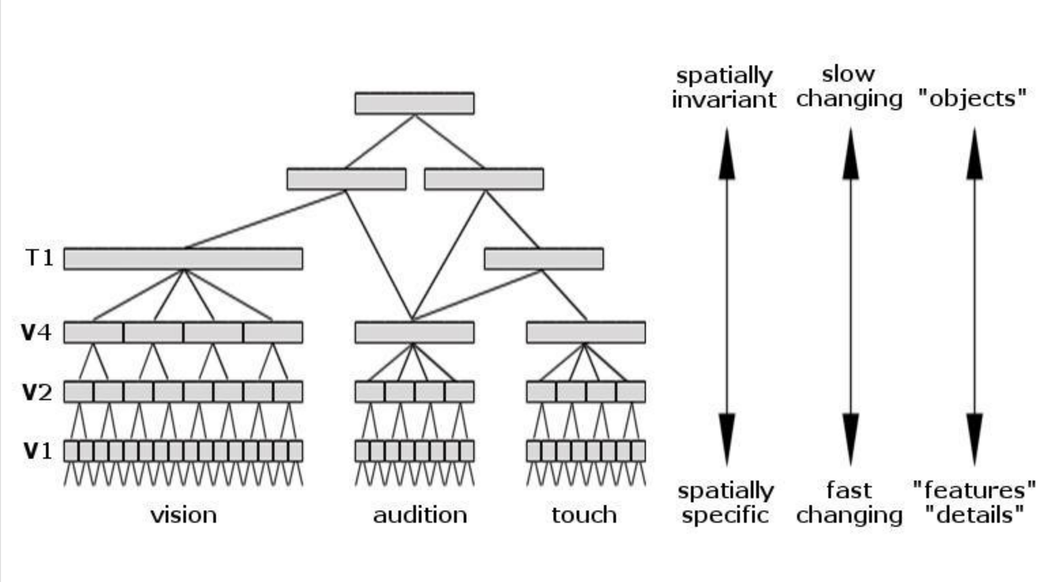
… in practice…

The concepts to which I believe Gershenson is (possibly inadvertently) referring are the "invariance representations" eloquently described by Jeff Hawkins in his book "*On Intelligence*" [18]. This notion is the controversial revival of the "grandmother cell" hypothesis proposed by the neurobiologist Jerry Lettvin back in 1967. At the time the idea was dismissed as "nonsense" by psychologists and it was regularly used as an example of "*one of the more ridiculous notions of intelligence*".

However recent work by neuroscientists Rodrigo Quian Quiroga, Gabrielle Kreiman, Christof Koch and Itzhak Fried has indicated that it may not have been so far from the truth after all [19]. Using electrical implants positioned within the brains of seizure patients, they found that some neurons fired only when the patient was presented with a rather specific visual stimulus (their tests used images of famous faces). One patient had a particular neuron fire exclusively to images of the actress "Halle Berry". Astonishingly, this didn't just include clear photographs, but included more obscure images[10], pencil-drawn pictures and even just to the written words of her name. "*This neuron is responding to the concept, the abstract entity, of Halle Berry*" (Quiroga).

Hawkins extends this work and gives a detailed explanation of how invariance representations are formed within the brain. The visual cortex (and in fact all cortical regions) is made up of hierarchically structured layers, with each layer being made up of a number of "columns" (figure 4 shows a very basic representation of this structure, taken from a number of diagrams from within his book). Each column learns (through Hebbian "wiring and firing") to recognise and respond to a particular stimulus. In the lower layers these stimulus are very simple visual representations (for example a "\" shape, or an "/" shape); but as we move up the hierarchy, the representations become more complex.

[10] Her dressed up as Catwoman, wearing a mask which covered most of her face.



(**Figure 4**) Hawkins' simplified view of the cortex

At the top of the structure, we would find our "Halle Berry" neuron, our "Bill Clinton" neuron and our fabled "Grandmother" neuron.

However, will we be able to find neurons for other (less noun-like) concepts such as an "underneath" neuron, or an "in an hours time" neuron or a "6.462713942" neuron? What about experiences like a "smell of coffee" neuron, or a "colour red" neuron or even a "love" neuron? These questions are so far unanswered, but I expect that most neuroscientists would accept that a "love" neuron is probably asking too much. Some concepts are not quite as simplistically "invariant" as a noun.

… and beyond

Let us think back to the Floreano robot for a moment. Floreano states that the special n7 neuron possessed a "*spatial representation of the environment*". Surely this can be said to be similar in some way to an "invariance representation". It is a representation (a concept) of "*where I am in the world with relation to the charger*"; but it differs crucially as it is a proportional representation of this concept. It is a "***variable** representation*".

With invariance representations, you either are "Halle Berry" or you are not "Halle Berry"; there is no variability, this knowledge can be considered to be a **constant** representation.

This certainly takes us a long way to being able to conceive of Gershenson's Behaviour-based Knowledge Systems; but we still lack a clear, understanding of how these concepts can be "reasoned".

For this, I tentatively suggest a key feature of dynamical systems. Think back once more to our neuron packed Braitenberg vehicle. Gershenson proposes that it should be extended to be able to build its own set of concepts (**variables** and **constants**) purely by interacting with an information rich environment.

In so doing, the only place for these concepts would be as stable states within the dynamic system. A dynamic system which, "if tuned correctly", has the potential to perform the same universal computations of a Turing Machine. Variables and constants working as representations of knowledge within a system with universal computability; this sounds awfully familiar. Could it be that concepts in this sense are mere symbols within some form of emergent symbol manipulating system? This is obviously a purely speculative idea for now, but it appears, at least to me, to be a somewhat tantalising part of these exciting new *concepts of intelligence*.

5 REFERENCES

[1] **Searle J.** (1980) "*Minds, Brains, and Programs*", the Behavioural and Brain Sciences, vol. 3

[2] **Penrose R.** (1990) "*The Emperor's New Mind"*, London: Vintage.

[3] **Crick F., Koch C** (1990) "*Towards a Neurobiological Theory of Consciousness*", Seminars in the Neurosciences. vol 2 pp 263-275.

[4] **Torasso P.** (1985) "*Knowledge based expert systems for medical diagnosis*", Stat Med. 4(3): pp 317–325

[5] **Drurie B.** (2005) "*Senses special: Doors of perception*", New Scientist. 2484: pp 34

[6] **Braitenberg V.** (1984) "*Vehicles: Experiments in Synthetic Psychology*", Bradford

[7] **Brooks R.** (1991) "*Intelligence without Representation*", Artificial Intelligence Journal (47), pp. 139–159.

[8] **Brooks R.** (1991), "*Intelligence without Reason*" Proceedings of 12th Int. Joint Conf. on Artificial Intelligence, pp. 569–595.

[9] **Webb B.** (1996), "*A Cricket Robot*" Scientific American, Dec., 1996, pp 62-67

[10] **Kirsch D.** (1991), "*Today the earwig, tomorrow man?*" In Artificial Intelligence, V 47, pp 161-184

[11] **Tarsitano M.S., Andrew R.** (1999) "*Scanning and route selection in the jumping spider Portia labiata*", Animal Behaviour, Volume 58, Number 2, pp. 255-265(11)

[12] **Wuensche A** (1996) "*Attractor Basins of Discrete Networks*", D.Phil thesis, University of Sussex

[13] **Berlekamp E. R., Conway J. H., Guy R. K.** (1982) "*Winning Ways for Your Mathematical Plays, Vol. 2: Games in Particular*", Ch 25, London: Academic Press

[14] **Floreano D., Mondada F.** (1996) "*Evolution of Homing Navigation in a Real Mobile Robot*". Cybernetics, 26(3), 396-407.

[15] **Floreano D.** "*Evolutionary Robotics: A gentle introduction*", http://lis.epfl.ch/resources/documentation/EvolutionaryRobotics/er.php

[16] **Seeley T.D., Towne W.F.** (1991) "*Collective decision making in honey bees: how colonies choose among nectar sources*" Behav. Ecol. Sociobiol. vol 28 pp 277-290.

[17] **Gershenson C.** (2002). "*Behaviour-based Knowledge Systems: An Epigenetic Path from Behaviour to Knowledge*". Proceedings of the 2nd Workshop on Epigenetic Robotics. Edinburgh.

[18] **Hawkins J., Blakeslee S.** (2004) "*On Intelligence*", Owl Books

[19] **Quiroga R. et al.** (2005) "*Invariant visual representation by single-neurons in the human brain*", Nature, 435: 1102-1107

[20] **Chalmers D.** (1995) "*Facing up to the problem of consciousness*", Journal of Consciousness Studies 2 (3), pp. 200-219