Chapter 2

Thought and Language

2.1 Good Old Fashion AI

Why do we want artificial intelligence, and particularly one that provides the possessors of this intelligence with the gift of language? Clearly the language is necessary if we want to interact with machines so that they can follow our instructions and also for them to interact to question us for information they may need and provide responses where we require them. But I shall argue in this chapter that language is important to intelligence for other more fundamental purposes. We shall see how *nouvelle* AI, that broke away from the traditional techniques developed from the 1950s through to the 1980s, can use the development of initially primitive language like behaviour to get back some of the advantages gained in the more traditional symbolic AI. The term Good old fashion AI or more commonly GOFAI refers to traditional approaches of AI in the light of *nouvelle* AI.

Whilst being a topic of interest for over a century with interest shown by Lady Lovelace as far back as 1842 [Lovelace 1961], AI as we understand it today came about with the invention of the electronic computer and was pioneered at the outset by Turing. The term "artificial Intelligence" was first used by McCarthy in 1956 at 'The Dartmouth Summer Research Project on Artificial Intelligence'. Since then the field has had two component objectives, these broke AI into what is known as *strong* AI and *weak* AI, terms coined by Searle:

> "I find it useful to distinguish what I will call "strong" AI from "weak" or "cautious" AI. According to weak AI, the principle value of the computer in the study of the mind is that it gives **us** a very powerful tool. For example, it enables **us** to formulate and test hypothesis in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states." [Searle, 1980].

The objectives of weak AI have been to provide a method to help us understand components of our own intelligence. This has been achieved either by attempting to directly model components of our own brain or through a more circuitous method of trying to increase the dataset of examples of a particular behaviour (which our own intelligence will provide one example of) to generalise about the particular behaviour. This second approach could eventually prove particularly useful with language where human natural language is the only example[1].

Strong AI has focussed on the actual development of intelligence, seeing our own behaviour as a subset of this. This side of the discipline is more aligned with control engineering and is concerned with how intelligent agents can be produced to perform specific tasks within normally tightly defined environments such as heart disease diagnoses. However, an overriding goal for this work is to be able to produce general intelligence where the agent, or if it is physically instantiated then robot, is able to deal with all aspects of the real world in which we are quite comfortable as humans with our own intelligent behaviour.

The direction of research in both of these branches has undergone a change within the last 15 years. By taking a look at some of the suggested failings of the original methodology we can get a clearer picture of what issues this new direction aims to address.

2.2 Problems with Traditional AI

Early AI was dominated by a symbolic approach. The problem space that the particular AI system was to be applied to was represented by a description of the world as a symbol system. Objects within the world would have a list of properties associated with each of them and also relationship with other objects. Newell and Simon proposed the *Physical Symbol System Hypothesis* which claims that a physical symbol system is all that is required for general intelligence where a physical symbol system

[1] An alternative route to generating further sets of language behaviour has been with the ape language programs of Savage-Rumbaugh [1978,1986,1994] though there is much controversy over the amount of language that can be and has been developed in this way. In the next chapter we shall discuss this further.

> "… consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity call an expression (or symbol structure). Thus, a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instance of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction and destruction. A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider that just these symbolic expressions themselves."[Newell & Simon, 1976].

For example an intelligent entity used to play a game of chess might have symbols to represent the different pieces that would have a position, represented by more symbols, relative to the other pieces and the board. The pieces would also have more symbolic properties such as type and colour. In this example the pieces might also have properties that dictate the range of moves they can make or this could be encoded into the axioms or physics of the world. These properties determine the processes referred to by Newell & Simon. When this is extended to real world problems, the idea was for as much information about the world to be encoded as possible that was deemed to be necessary for the particular problem domain that the intelligent entity was to be applied to. However, in most real world problems it is very difficult to limit the scope of a particular problem and as a result, the system needs as much world knowledge to be encoded as possible.

Besides systems to play restricted games such as chess by Prinz in 1951 and draughts. Strackey also 1951, the first attempt to use this approach was with the General Problem Solver [Newell, Simon & Shaw, 1957]. Given that heuristics were programmed into the machine and it was these that were used to make decisions, it is not clear that the intelligence is coming from the machine rather than the programmers.

At the start of this thesis we saw an example of a dialogue produce by Weizenbaum's Eliza system. [Weizembaum 1966]. This system did have incorporated into it some world knowledge such as the ability to parse English sentences and some very limited information about relationships between words such as family, mother, father etc. However there was considerably little knowledge present for it to perform the task it appeared to be doing of acting as a human therapist and the program relied on the users to provide information, and then exploited their willingness the put personal and relevant interpretations onto the very neutral responses regurgitated by the system.

Another system that interacted with users of a similar period but with considerably greater intelligence was SHRDLU created by Terry Winograd [1972]. This system allowed a user to question it about a *blocks world*. This blocks world was a simulated environment in which various coloured and shaped blocks existed. The physics of the world allowed the blocks to be placed relative to each other in different ways by the intelligent system. The system would interact with a user via natural language both typed in as input and written to a screen as output. The user could issue instructions for the system to interpret and act upon such as "place the red ball on the green block". Furthermore, the user could question the system about the state of the environment, for example "Do you need to touch a blue object to lift the green block?" and the system might reply " yes, the blue pyramid.". The system was able to parse these interactions, both syntactically and, importantly, semantically (this was a major difference to the ELIZA system).

When we look in a little more detail at the semantic model employed by the system a flaw appears. In one of the examples we saw the system being able to differentiate colours. It seemed to have knowledge of what was blue when questioned about objects this colour. However, there is no correspondence between the system's notion of a particular colour and that colour. The system does not experience qualia because it does not interact with the world. If some property of red in not encoded into the systems knowledge base then this information will always be inaccessible to the system. This problem is referred to as the *frame problem*. A strong attack of traditional AI with respect to the frame problem is made by John Searle's thought experiment "The Chinese Room" [1982].

Given that any particular system might not have sufficient symbols to describe a particular world state at any time, the systems must be able to create new symbols at any stage that relate to these newly encountered world features and also relate to their existing symbol system.

A further problem besides adequate levels of description for a symbol system is the computational burden of using world models. With the top down approach to intelligence, complex representations had to be manipulated that captured all the details of the environment that might have some impact on the behaviour of the system. As a consequence of this, any real world problem resulted in a very large computational burden on the intelligent system which the system could not deal with in real time. Where the system was controlling a robot such deviance from the consequences of not working in real time was clearly not acceptable. An example of such a problem was encountered with SHAKEY [Nilsson, 1984] that was restricted to very simplified environments. Another approach had to be found to reduce the computational load on the system.

Work in the 1980's went some way to addressing the issue of complex representations with micro features through the use of neural networks. This approach was particularly championed by the PDP (parallel distributed processing) research group headed by Rumelhart and McClelland [1986:I,II]. They were able to show that an initial objections to neuron inspired computing with perceptrons on account of difficulties of scaling and an inability to deal with Xor type problems [Minsky & Papert 1969] could be overcome with back propagation algorithm when neurons with semi-linear activation functions were employed. These objections had almost completely stopped research in this area during the 1970s. However, even though the networks broke down the high level representations, they were still seen as requiring such high levels of symbolic abstraction for input and output. An example of this is portrayed with Rumelhart and McClelland use of networks for past tense language acquisition [1986:II].

2.3 Nouvelle AI

A new approach to cognition started to emerge in the late 1980's born out of an engineering desire to create functional autonomous robots whilst avoiding the shortcomings we saw in SHAKEY. Inspiration from the dynamic and navigational capabilities of insects with very small control architectures lead to a more reactive approach with the creation of robots such as Allen [Brooks 1986] and Herbert[2] [Connell 1989]. Herbert maintained no state for very long (3 seconds) with a task of finding drink cans in a working laboratory rather than a specially prepared clean, clear room.

The basis of this new approach was to attempt to produce AI behaviour from the bottom up, starting with small units of control, working toward integrating them to create ever more complex systems rather than the previous top down method of design and investigation. The notion of a bottom up approach is made clear in the editorial of an early conference

[2] Allen & Herbert were named after Newell & Simon

proceedings, to bring awareness to the new movement by Steels and Brooks [1995]. This was further reinforced by an allegiance with biological sciences.

> "These researchers have built up some strong alliances with recent work in artificial life because of the strong biological bias in their work The link with biology is not in terms of modelling biological phenomena but rather in exploiting principles underlying living organisms." P.1 Steels & Brooks 1995

Rather than relying on complex internal models of the world that would be used for planning, this new approach saw the world as its own best model and relied on systems having a tight coupling between the environment and the agents. Clark & Chalmers [1995] called this extended cognition with the need to offload the cognitive load onto the environment suggesting that in biology, evolution has tended to favour the ability to take advantage of immediate information rather than bearing the cost of storing and retrieving information.

Whilst models were very simple to start with on account of the bottom up approach, the focus was always on how they could be part of a larger program to create ever more complex intelligent behaviour.

> "These robotic agents are not built with the prime goal of automating parts of sensory processing or action control, but as a first step towards the study of full cognitive agents."[Steels & Brooks p2 1995]

In dealing with notions of symbol grounding Varela[1995] argues that representations arise through interaction with the world rather than from a priori Cartesian categories and these are for control rather than a means to recreate the environment in which the agent is based

> "For the dominant computationalist tradition, the point of departure for understanding perception is typically abstract: the information processing problem of recovering predetermined properties of the world. I contrast, the point of departure for the enactive approach is the study of how the perceiver can guide its actions in its local situation. … the overall concern of an enactive approach to perception is not to determine sow some perceiver-independent world is to be recovered; it is, rather, to determine the common principles or lawful linkages

> between sensory and motor systems that explain how action can be perceptually guided in a perceiver-dependent world." [p.16 1995]

This idea is further supported by Simon's [1969 1981] suggestion that complex behaviour could be exhibited without complexity of control but as a reflection of a simple set of interactions with a complex environment. Brooks makes explicit the claim that internal representations that caused such problems with previous endeavours are not necessary with this bottom up approach

> "The post-Cartesian agent manages to cope with the world without necessarily representing it." [Brooks 1997 p.428]

A sub-discipline within this new bottom up approach that has looked to biology for inspiration is evolutionary robotics. Evolutionary robotics has borrowed techniques developed in the 1970's for genetic programming [Holland 1975] but applied it to non symbolic control architectures. Early examples of this include Viola [1988], Garis [1991], Harvey & Husbands [1992], Beer & Gallagher [1992], Cliff, Husbands & Harvey [1993].

Simple control architectures are evolved through a method of variation and selection. Rather than developing a single controller, a population of many is normally used (thought there are cases where a population of 1 is acceptable), and these might be used in turn within a single robot or spread across a population of robots. Selection occurs by means of a fitness function used to grade the performance of the robots with each of the controllers. Some portion of the controllers that create what is considered to be nearest performance to the goal behaviour are selected and used to create new controllers. These new controllers vary from their *parents* typically by processes similar to those observed in Darwinian evolution such as mutations and combination of parent controllers.

Evolutionary robotics has had much success at many reasonably low level control tasks such as complex coordination for multi-pod motion [Jakobi 1995], Ijspeert, Hallam & Willshaw 1998], action selection [Seth 1998] and typically navigation tasks [Husbands Harvey & Cliff 1995] but in common with other bottom up approaches, and not surprisingly, does not compete effectively with traditional AI techniques at higher order cognitive tasks such as planning and scheduling.

It seems as though the benefits achieved by both the top down and bottom up approaches would provide an ideal solution, allowing for the possibility of grounded symbols necessary for agents to perform a particular task. Even Brooks [1997] entertains this notion and claims that one of the most serious questions about the limitations of the bottom up approach concerns the level of complexity of behaviour that can be achieved without the assistance of central representations. This is supported further by van Gelder [1997]

> "..in fact there is nothing to prevent dynamical systems from incorporating some form of representation. Indeed, an exciting feature of the dynamical approach is that it offers opportunities for dramatically reconceiving the nature of representation in cognitive systems, even within a broadly non computational framework." P.444.

van Gelder goes on to suggest that linguistic structures might be used for such representations. We shall now consider how language can help bridge the two approaches.

2.4 Language

When we use language we are making use of a similar type of system to those used in traditional AI. Our language is essentially a grounded physical symbol system and our cognitive functionality to deal with it is a symbol manipulating system. There is nothing new in the idea that language can act to link our various sensorimotor activities in order to perform higher order cognitive functions, This idea was espoused by Vygotsky in the 1920s, who suggested that language acted as a cognitive scaffold that assists us to learn associations at a sensorimotor level

> "By means of words children single out separate elements, thereby overcoming the natural structure of the sensory field and forming new (artificially introduced and dynamic) structural centres. The child begins to perceive the world not only through his eyes but also through his speech. As a result, the immediacy of "natural" perception is supplanted by a complex mediated process; as such, speech becomes an essential part of the child's cognitive development" [Vygotsky 1978].

Clark [1997], and Clark & Chalmers[1997]more recently suggest that language not only assists us with the development of our cognition but also acts to enhance it by allowing us to extend our cognitive processing capabilities beyond our bodies

> "Language appears to be a central means by which cognitive processes are extended into the world. Think of a group of people brainstorming around a table, or a philosopher who thinks best by writing, developing her ideas as she goes. It may be that language evolved, in part, to enable such extensions of our cognitive resources within actively coupled systems." Clark & Chalmers[1997]

Clark further argues that when we use language, the symbols that are a key component of language provide for us the ability to make 2nd order references (that is references about references) and this opens up a whole new level of abstraction particularly as the ability to make 2nd order references allows for any subsequent depth of reference. This notion of symbol, or symbolic signal, contrasts against iconic and indexical signals. These signal types being categories of referential association defined by Charles Peirce [1955].

An iconic signal captures similarity of physical features between an object with a sign so that there is a one to one relationship between them. A classic example would be a camouflaged moth that had markings similar to the bark of a tree that it lived off. The moth is conveying a signal "I am just another bit of tree so pay no special attention". Indexical signals capture similarity of temporal and physical variables between an object so the markings made by tigers scratching as high up a tree as possible to indicate their fighting strength as a consequence of their size would be an indexical signal. Symbolic signals capture a conventional or arbitrarily agreed link with no physical correspondence between an object[3] and a sign. When we use the word 'white' to describe something in our environment that reflects light with an even distribution from across the visible light spectrum, the word could as easily be 'blanco', 'blanche' or even 'black'. We simply have agreed that under certain contexts this is what this signal refers to.

Deacon [1997] stresses that all of the signals are key components to our cognitive processes and that mental representation which is key to our cognitive capabilities is simply a form of communication

---

[3] The Object does not need to be a physical object. In order for 2nd order reference we saw that the object in that case would simply be another symbol

> "Icons and indices are not merely perception and learning, they refer to the inferential or predictive powers that are implicit in these neural processes. Representational relationships are not just these mechanisms, but a feature of their potential relationship to past, future, distant, or imaginary things. These other things are not physically re-presented but only virtually represented by production of perceptual and learned responses like those that would be produced if they were present. In this sense, mental processes are no less representational than external communicative processes, and communicative processes are no less mental in this regard. Mental representation reduces to internal communication." [Deacon p.78].

2.5 Summary

In this chapter we have considered the history of AI and argued that the symbol grounding problem has acted as a stumbling block for the traditional physical symbol system approach. To replace this we have considered a bottom up approach particularly within the context of situated adaptive behaviour and evolutionary robotics, which is one manifestation of this. This approach is not without its faults however and systems designed in this way lack some of the less reactive capabilities enjoyed by physical symbol systems. To overcome this limitation we propose that evolutionary robotics should move to incorporate the power of a physical symbol system through the acquisition of linguistic behaviour. In keeping with techniques of evolutionary robotics, this will still be done in a bottom up manner ensuring that any symbols are grounded in the sensorimotor capabilities of the rest of the system. This position means that we accept that the development of language capabilities offers a way of extending cognitive development as the components are useful in both. There is always the possibility for debate about which follow the other but we shall be assuming within this thesis that by developing the capabilities of linguistic (or at very least communicative) function, they will be available for enhancing cognitive powers. Whilst we have seen that three types of signal (iconic, indexical and symbolic) are possible, it is towards symbolic signalling that we should strive as this offers us the ability to make arbitrary references and more importantly 2nd order reference that free us from the shackles of the merely physical.

For the next chapter we now need to look at a variety of other disciplines that share an interest with notions of signalling, communication and language and by looking at the ways in which these terms are defined, draw up a clear notion of the behaviours that we shall be expecting our agents to acquire. With this in mind, we will then look at existing research that more directly contributes towards the goals of this project. An extensive review is undertaken which enables us to perform a categorisation of the work in order to break the problem into a number of meaningful sub-disciplines.