Distributed Cognition
In
Interpersonal Dialogue

Grant Blair

I, Grant Blair, declare that

Distributed Cognition in Interpersonal Dialogue,

Is my own work and that all sources have been appropriately referenced.

Durban

July 2003

Abstract

The study of cognition has suggested different views of what a system needs to perform computations. A strong computationalist approach aims at producing and preserving true statements through syntactic recombinations of elements. Alternately a more action-oriented approach stresses the environment in which the system is placed and the structure that this may provide in performing computation. What is at issue is that the strong computationalist view depends on a particular view of symbols that are decontextualised and function primarily syntactically, in the service of pragmatic goals. It is argued that some of the lessons learned from embodied cognition, in the form of epistemic and strategic action, can aid the ways in which symbols are supposed to function in dialogue. In so doing, attention is turned away from the reification of speech both in the form of text manipulation and transcript and begins to look at the environment, situatedness, function and properties of speech. Consequently, language comes to embody some of the ways in which we manage our interaction with the world and other agents, making it a little more than an artefact but a little less than a complete and disembodied picture of rational cognition.

'We are interested in the concept and its place among the concepts of experience.'

— Wittgenstein

Acknowledgements

Table of Contents

1. Introduction

The primary aim of Cognitive Science and by extension AI has been to understand the ways in which human agents solve problems. From the inception of AI, the dominant approach has been one of computation understood as the ability of Turing (and later von Neumann) machines to solve cognitive problems. Exemplars of the Classical Approach are Chomsky (1965) and Fodor (1975) where the theoretical abilities of Turing machines are strongly identified with the mind. From the eighties a surge of interest in connectionist architecture (Clark 1989, Elman 1991) has served to put some pressure on this approach to understanding the mind. In addition, there have been more action and activity inspired approaches in the shape of research based on biologically plausible models of insects. In the early nineties Brooks (1991a,b) solidified this approach with two papers detailing some reasons for approaching cognition from a more activity driven perspective. Critically Brooks' account assumes an embodied architecture situated in and responding to an environment, an approach that differs radically from the disembodied cognition carried out by von Neumann-type architectures.

These two different approaches have suggested two opposing views about what constitute systems. Von Neumann style cognition simply is internal computation. The Von Neumann system is predicated on distinct divides between the world and the internal computation of the system. The definition of 'system' is an input/output cognitive model where all the computation takes place internal to the system. Action by the system is dependent upon full derivation of that action within the system. The system senses, computes and is able to effect actions in the environment only as a product of a strict sequence. Thus, the only action that the Classical approach is capable of is pragmatic. Pragmatic actions are those designed to achieve a goal according to a plan based on the information the system has gained from the environment. A further corollary of the divide between sensing and acting and the resulting linear sequence of cognitive stages, is the necessity for the provision of all information of the environment relevant to the task, before the system can act effectively. It is clear that von Neumann machines are only capable of pragmatic actions, due to the fact that all actions have to be centrally planned based on a representation of the task space in an abstract code.

1

Away from this theoretical approach, there has been a move to see systems as part of wider environments, that is as situated and embodied, where the environment and the physical capacities of the system make crucial differences to what the system can achieve cognitively. (Brooks 1991, Kirsh 1991; 1995; 1999) Crucially, to Brooks in particular (Section 3), the systems he builds can interact with an environment without the need for either centralised representation or control. The key features of Brooks' work, are the fact that external interactions effectively help lighten cognitive load and that the need to respond quickly to changes in the environment (either to avoid damage or to capitalise on fortuitous circumstances) counts against "representationally heavy" implementations of cognition in that they are simply too slow. In addition, the Classical demand for coding of abstract task representation leads to an emphasis on abstract problem solving amenable to these abstract representations, inconsistent with the everyday demands of systems operating in a world, (insects, humans).

Pursuing the active cognition line more strongly, Kirsh and Maglio (1992b, 1994, 1996) make a strong case for a category of action, called epistemic action, that cannot be accounted for by traditional models, directed not at pragmatic ends, but at changing the environment, and thereby lightening cognitive load (Section 4). Through their research on the computer game Tetris we can begin to clarify and define the concept of epistemic action in the Tetris task domain, as well as begin to clarify ways in which epistemic actions may be generalised to more interesting cases. In attempting to relate their work to the capacities of humans, Kirsh and Maglio show how human players can utilise epistemic actions in the service of higher-level policies. The ability of human players to use epistemic actions to control higher-level policies shows up important differences between the skills of Brooks mobots[1] and the functioning of human players in the Tetris domain. Architecturally however, the cognitive model generated from the Tetris research shows Brooks and Kirsh and Maglio to have much in common.

---

[1] Brooks mobots are mobile robots. They are designed to operate in a real and dynamic environment in strong contrast to Classical AI systems whose main aim is disembodied computation.

Pursuing interaction with the environment (and other agents) a little further, we turn to look at emotions, examining an empirical project (Section 5.1, Breazeal and Scassaletti 1998) that effectively tries to solve the problem of situated social infant learning (in robots) by combining emotion modules within the broad tenets of Brooks' architecture. Emotions here (which, it is granted, may bear only simplified and limited relation to human emotions) are used to prompt a naïve subject interacting with the robot to fulfil the drives associated with learning. The architecture is important in showing again the lack of centrally localised and abstract representation, as well as the micro-time domains (approximately 500msec) in which the interactions take place. This example provides ways in which to begin to think about the relation of emotions to cognitive systems construed as the ability to solve problems. In addition, I show how emotions narrow the search space of a problem (Section 5.2, e.g. Evans 2002[2]) and regulate and coordinate interaction amongst parties (Section 5.3, Ross and Dumouchel, forthcoming.)

As can be seen, the dissertation, while ultimately concerned with aspects of dialogue, is concerned for the first five sections in gaining a theoretical foothold on some of the aspects of what action can achieve in the service of cognitive and later interpersonal goals. This will allow an approach to language and dialogue in the latter sections (Section 6), showing the different views of language and their implications for cognition. From an embodied and situated perspective, language and symbols have been seen as artefacts that are part of action loops. Clark (1997; 1998) shows that language can aid the pursuit of cognitive goals by using some of the same strategies that Brooks' mobots and Kirsh and Maglio's work has done. However, most of these observations are made with respect to texts, (books) or to signs, (like restaurant signs), assuming the prior appearance of word-based forms. In contrast, I start to look at possibly the most active use of language, dialogue, which may not depend entirely on the prior appearance of structured language in the form of either text or spoken symbols.

With an emphasis on the epistemic and environmental features of language, we should not however go on to say that words are just sounds. My working hypothesis is

---

[2] See also Evans (2001), Frank (1988), Damasio (1994)

that speech is not like the production of a symbol string, nor is it the mere manipulation of external artefacts, (the ultimate artefact according to Clark (1997: Ch.10). It is rather that language and symbols can come to be seen as not only serving pragmatic ends, but also epistemic ones, in action loops, concerned with coordinated and joint activity that serves to change our own cognitive states (Section 7), thus filling an explanatory gap between internal computation and how language is a part of the scaffolding of our environment. Language becomes an activity by which we manage the interactions and tasks around us.

The paper does not claim to give an answer to what symbols are, nor specifically to how they aid cognition (e.g. Clark 1998), but to the question of how embodiment and situatedness can help symbols. I propose to look at the ways in which language can involve epistemic and strategic action and how this kind of action can come to shape and fix the context or circumstance of dialogue. In so doing I operate under the assumption that dialogue is primary and active and argue that dialogue serves epistemic ends in addition to pragmatic ones. In the process, language becomes something at least a little more than an artefact but a little less than a complete and disembodied example of rational cognition.

2. Action, the Classical Approach and Language

The power of the von Neumann style architecture is that it centrally represents the information about the problem it is trying to solve. The same ability is also its greatest limitation. For any problem in any task space, the system must build an explicit internal description of the environment (Brooks 1991b) before any cognitive activity can take place. If the system has less than full Classical representational information[3] then its decisions for actions become more or less random (Kirsh and Maglio 1994). Since this approach relies on a coding of the environment quite apart from the specific problem within a task or the potential actions to be carried out in an environment, it is called 'representationally heavy'. The model implies a complete and static representation according to the relevant abstraction of code, continuously refreshed as changes happen, whether the information is significant for the task at hand or not. Now this need not be a problem when the task domains are either highly abstract or highly circumscribed or both. However, it does become a problem for systems that are to be embodied and perform tasks where they have to extract information from an uncircumscribed task environment. In short, von Neumann type architectures are not good at tasks that include physical goals, due in part to the inability to extract salient information timeously from an environment that is not circumscribed, or that allows many different possible actions or tasks in the same space.

In contrast, the way in which Brooks mobots (1991a,b) operate is that they have perceptual systems and activity systems wired into the same module (and wired between modules). For example, a rudimentary robot, Allen, (Brooks 1986) uses 'sonar readings to keep away from people and other moving obstacles, while not colliding with static obstacles'. (Brooks 1991b:18) Thus, when an object or obstacle is encountered, even the very rudimentary coding of 'obstacle' is not necessary as a representation. Through continuous sensing of the environment, the robot will simply stop and by activating other modules that sense direction move away from the obstacle. The significance of this approach is that there is less computation going on than in a Classical computationalist sense, for there is no representation with which

---

[3] Brooks rejects representation in the form of 'explicit representations of goals within the machine' (Brooks 1991b:19), but does not reject the representation of aspects of the world at particular layers of his machines. Further representations can be introduced as *active-constructive representations* (Brooks 1991b:19, italics in original). For an example, see section 3.3.

the system can begin to compute trajectories. Nonetheless, the success of such systems has far exceeded the performance of more formal systems. Brooks mobots use information directly gained from the embodiment of the system in order to solve problems like navigation. In contrast, classic von Neumann architectures are typically not embodied, and peripheral systems, it was thought, could be added to the basic computational architecture later. However, this has proved difficult as far as the bodies of agents seem to have effects on the kinds of computation of which they are capable.[4]

A number of terms become important in keeping track of how the embodiment of systems can influence the ways in which we go about cognitive tasks. The first of these is scaffolding exploited through the second: action loops. The idea of scaffolding as an aid is not entirely new and has been well covered in Rumelhart et al. (1986:45-6) with the example of long-division. Rumelhart et al. use scaffolding to explain the way in which we manipulate symbols to perform calculations which could not otherwise be performed easily internally, or without the aid of external props.

> These dual skills of manipulating the environment and processing the environment we have created allow us to reduce very complex problems to a series of very simple ones. This ability allows us to deal with problems that are otherwise impossible. This is *real* symbol processing and, we are beginning to think, the primary symbol processing we are able to do. Indeed, on this view, the external environment becomes a key extension to our mind. (1986:46, italics in original)

Scaffolding is the structure present in the environment that we use, through action loops, to solve problems. For example, when doing puzzles, we typically turn pieces this way and that trying to match them through physical manipulation as much as through internal computation: matching the right colours. In contrast, a classic system trying to solve such a puzzle will represent the entire problem space, do some computations and pattern matching exercises and then execute the plan, i.e. place the

---

[4] For some wonderful examples and depictions of insect worlds see von Uexkull (1934) in Clark (1997:26,27). For instance, ticks hanging on trees are sensitive to butyric acid given off by passing horses. When a sufficient concentration of acid is sensed, they drop from the tree onto the horse. Their world, in some sense, consists of factors not present in our own, as a robots world also consists of factors not present in our own, due to 'their own distinctly non-human sensor suites'. (Brooks 1991a:3)

piece in the correct place[5]. This is patently not we do, nor, it would appear, are we typically very good at this kind of computation.

In contrast, the von Neumann style model of computation or Representational Theory of Mind (Fodor 1987) relies on the syntax of natural languages to account for the truth preserving properties of thought. (Also called The Syntactic Image, (Clark 1993) and with an emphasis on action, more generally the Classical Approach in this paper.) Syntax preserves the laws of logic and thus guaranteeing logical truth. Logical transformations between elements (commonly identified with words) are provided by syntax. The symbol system in which the syntax operates allows for *semantic compositionality* (Clark 1993:11, italics in original). The division between syntax and semantic compositionality allows for transportability, such that an element of thought in the current syntactic combination is the same element in a different combination. Thus, there has to be a level of translation between the sensing and the computation, complete with a logical/syntactic representational structure which Brooks has called a 'representational bottleneck'. (Clark 1997:21) An emphasis on more action-oriented models is precisely in contrast with this level of interpretation and representationally heavy view of cognition.

Symbols become prominent here in providing content to the computation. Moreover, syntax preserves systematicity[6], or logical transformation between propositions, where symbols are context-free bearers of content.[7] The problem with this sort of breaking down of objects, as well as reference to the objects themselves, is that they

---

[5] Other examples include re-arranging Scrabble tiles to prompt possible letter combinations (Maglio et. al. 1999) and re-arranging cards in hand to lighten computational load. Maglio et al show that physical manipulation aids participants in strong domains of traditional AI, ie. physical manipulation aids search, a trivial task for a von Neumann machine.

[6] In the classic example, 'humans who can think that Mary loves John can also think that John loves Mary.' (Fodor and Pylyshyn 1988, in Clark 1997)

[7] Context-free bearers of content are elements of objects, or other folk-psychological solids, which do not (and cannot) apply to any particular context as they must exhibit the property of transportability, and are perfectly general and recombinable. Thus 'In such a system we will find atomic symbols and *molecular* representations.' (Clark 1993:11, italics in original) such that a combination of A and B producing AB 'correspond to real physical structures in the brain'. (Fodor and Pylyshyn 1988:13) exhibiting syntactic and hence combinatorial properties. It is noted that there is no obvious in principle objection to 'the idea that a context-free content-bearer could be 'intrinsically' associated with action, just as long as it's the same type of action in each context'. (Don Ross: personal communication) This assumes that identical actions can be carried out in multiple contexts. An action-oriented approach denies this. The action-oriented approaches discussed here assumes that action is 'intrinsically' context-dependent.

have no intrinsic reference to action or the possibility of actions to be carried out with the object; not even for instance, whether it could be thrown or not, or other features which could be directly related to embodiment. In addition, while it is well known that spoken language is more context sensitive, the analyses proposed still rely heavily on categories taken from representationally heavy accounts, effectively dismissing the effects of the environment on an analysis of language.

It is here that the action modelled perspectives seem to provide alternatives to the Syntactic Image (Clark 1993) in that they demonstrate that cognition does not always need internal symbols (or tokens) in the form that the Syntactic Image would seem to require. From the point of view of an action oriented model, symbols or tokens seem superfluous, for we can build systems that perform tasks without the need to fully represent the world. As Clark (1993) and others (Kirsh 1991, Elman 1991) show, finding these context unbiased atoms is not necessary for their conception of cognitive systems. Situated and embodied systems do not need to rely so heavily on context-independent code in order to function[8]. Some researchers have found that, while symbols of the type demanded by Classic AI provide a great deal of computational power, they are unnecessary when accounting for the sorts of well documented cases of action that we use to lighten cognitive load. In relation to language, symbols taken as words are central to an internal computationalist approach. In contrast, the examples of active problem-solving we have covered so far do not use words or symbols and while these have not been ignored, they have been displaced from the centre of cognition.

We now have two approaches. The first, the Classical Approach is a syntactic engine, whose gains in generality and power make it difficult to act efficiently in a world. Problem solving in such a model relies on top-heavy environmental modelling requiring complete abstract task information in which the system can then search for possible solutions and compute trajectories. Language is implicated in such a model due to the fact that the representational coding systems are extracted from natural

---

[8] Attempts have been made to address this gap between Classical representational structures and more contextual models of representation, through the use of connectionist networks that can code for contextual elements of classes of objects through the use of prototypes. (Clark 1993:88-91) The networks are not programmed, but through variations in the training data set, extract stable sets or clusters of co-occurring properties from many examples of say, dogs.

languages. Thinking, on this view, requires language, without which there is no computation. In contrast, more action-based approaches to cognition have provided alternatives to this model in at least two ways. The first is that common examples of external symbol manipulation in terms of scaffolding and action loops combined with the inability of human brains to search in the powerful ways Von Neumann machines are capable of, appears to undermine the generality of the Syntactic Image. Further, as we shall see, actions carried out in the environment can be directly constrained by that environment, effectively alleviating the need for a complete representation of the task space at hand. On an activity-driven perspective, language is typically conceived of as an adjunct to human computation, used more effectively by external manipulation than internally as the basis of cognition. Language is not seen as fundamental to cognition, but merely a particularly powerful artefact or instrument for cognition.

3. Brooks' Challenge to Disembodied Cognition.

Rodney Brooks' approach to robotics has eschewed the traditional routes of sensory input, representation, problem solving and planned action in strict sequential ordering. His contention has been that the division of the problem into these functional domains is not only artificial but creates further problems (integration and interaction) that are more difficult to solve and, it is argued, counterproductive. He has thus begun with what would appear to be low level intelligences, revealing how action plays a role in the ways that they achieve the cognitive tasks they are set, often with resources far below what one would expect, due to the exploitation of scaffolding and action loops. The two main points that are argued, are first, cognitive systems are aimed primarily at controlling action, and, secondly, the internal system need not have central representational capacity construed as a central planner. Following from these two points, it would seem that agents interacting with a world do not need internally complete syntactic and semantic systems.

3.1 Division of Labour

Following Brooks, Andy Clark has noted that AI research has typically been circumscribed into narrowly defined and highly specialised problem domains such as producing the past tense of English verbs (Clark 1997:58). Brooks takes this slicing up of cognitive space and outlines the problems and misconceptions which are attendant on it. Classical approaches typically focus on what Clark has called vertical micro-worlds: expert systems solving problems in highly specialised domains such as playing chess or planning a picnic (Clark 1994:13). Brooks approach proposes a way not to generalise or integrate these systems, but focuses his efforts on creating autonomous agents which operate in an environment or horizontal micro-world. A horizontal micro-world is based on activity-based decompositions rather than specific task domains. The focus on horizontal micro-worlds replaces specialised tasks (the ability to plan a picnic, produce past tense verbs) with goals (navigation, soda can collecting). The significance of this is that the goal environment is not circumscribed in the same narrow and abstract ways a task environment is. A focus on the environment in which goals are pursued introduces a number of key terms which guide the architectures of Brooks' robots.

Brooks, in his discussion of his mobile robots (mobots) and the rationale behind them, makes use of several key terms that will be useful. These are situatedness, embodiment, intelligence and emergence. Situatedness is merely that a system is located in a world. The idea is meant to act against a tendency to abstraction (specifically task abstraction and abstract tasks) and rather allow the proclivities of the immediate environment to influence the robot directly. Embodiment is that the robot or system has its own sensor suites that are directly linked to the control system. Obviously the notion of embodiment and situatedness go hand in hand, but the distinction is important, especially later in trying to draw the line between sensing, reacting and planning. Intelligence is used here to denote something that is largely in the observer, such that the source of the intelligence is not exclusively the property of a 'computational engine' (Brooks 1991b:3). but is linked to emergence, where intelligence is understood as the ability to interact directly and sensibly with a dynamic world. The importance of the concept is that the manifestation of action may not be linked to a computational state or discrete event in the system. The complex that these concepts make is the push to spread both the computational load and the system beyond the boundaries of the physical implementation of the system.

The attraction of Rodney Brooks' approach lies partly in his physical implementation of mobile systems, and partly on his radical approach that dispenses with representation. I will cover some of his arguments against representation in the engineering methodology of system design. Brooks approach is not explicitly philosophical, his concern being to build complete systems (autonomous agents). However, while there are independent reasons to be wary of commitments to a strong language-based representational theory, we should still be wary of throwing out representation in the form of integration altogether (Clark 1997:32; Wheeler and Clark 1999; Wheeler 2001).

Brooks objects violently to the maxim of early AI that 'Good representation is the key to AI.' (Brooks 1991a:3) His objection rests on two bases; the first is the question of task decomposition and valid interfaces between sub tasks or different task specific modules; the second is the degree of abstraction associated with the different task decompositions. Brooks approach is to practise with lower level intelligences: 'Creatures' (Brooks 1991a:4) being able to move about in a world with purpose, before attempting higher-level intelligences.

Given the two requirements of complete autonomous systems and real world sensing and acting, there are four requirements for Brooks' creatures.

- A Creature must cope appropriately and in a timely fashion with changes in its dynamic environment.
- A Creature should be robust with respect to its environment; minor changes in the properties of the world should not lead to total collapse of the Creature's behaviour; rather one should expect only a gradual change in the capabilities of the Creature as the environment changes more and more.[9]
- A Creature should be able to maintain multiple goals and, depending on the circumstances it finds itself in, change which particular goals it is actively pursuing; thus it can both adapt to surroundings and capitalize on fortuitous circumstances.
- A Creature should do something in the world; it should have some purpose in being. (Brooks 1991a:4)

As far as the building of systems is a function of decomposition into sub-systems, Brooks favours the approach of activity decomposition as opposed to functional decomposition. In functional decomposition separate modules are responsible for perception, cognition and action in a sequence. In contrast Brooks' commitment to subsumption architecture is conditioned by a choice of activity decomposition. Brooks' subsumption architecture is such that every system is self contained; perception and action are immediate within a single module[10]. Thus the entire system maintains a purpose, whereas the sub-systems maintain an activity, e.g. sensing proximity to objects.

Brooks' activity decompositions are meant to alleviate the system's need to abstract away from the task space in the form of representations. In the Classical approach, the pursuit of systems suited to individual and abstract tasks combined with the idea of good representation as the key to problem solving, lead each individual research team

---

[9] For an account of the brittleness of the visual system of a classic representational architecture see Brooks (1991:8-9). The system was not sufficiently robust to cope with changes in light intensity, and navigated in bare rooms of sparsely distributed blocks and wedges, where edges and planes were specially painted. No system with a comparable architecture has managed to operate in an environment more complex than this example.

[10] Subsumption architectures are layered and interconnected systems of such modules. Input from the environment is output either to a motor or another module.

to perform the abstraction of salient information specific to the task. Environmental features are coded in a symbol system prior to the system computation. Due to the disembodied nature of the system this is both necessary, and encourages abstruse task domains in addition to a pre-occupation with categorisation. The need for representational information combined with the requisite code abstraction leads to an emphasis on circumscribed and abstract tasks. The task of symbol abstraction and encoding being left to programmers, the system has no reliable way in which to interact with the world: either in responding to it, or in being able to use it in accomplishing the task it has been set. For example, an expert chess system encodes board positions and computes possible continuations. Humans on the other hand are able to use the ability to move pieces in order to test positions, as well as moving their own position in order to get a different perspective on the board, particularly when learning to play. Brooks' contention is that abstraction removes the difficult part of the problem, leaving the system to do what it does best, which is search.[11] Further the abstraction, in a language convenient to the task, begs the question of valid interfacing between two different systems (say visual systems and actuators) that neither use the same abstraction processes nor have commensurate representational data stores. It is in part because of those abstraction and translation processes that central representation is necessary to a Classical approach as well as computationally costly and slow.

As Brooks is quick to point out, from his perspective, representation in the Classical mould, as a requirement for a system, comes under increasing pressure. Firstly subsumption architecture does not allow any notion of central representation. There is no translation of features of the world into any code whatever upon which the system is then able to reason. Information is transmitted according to transducer signals rather then symbols. Neither is there a sense in which there is a module to which all information present to the system is available. The importance of this consideration is that no central representation is necessary due to the fact that there is no need for a further stage or central executive module with which to make decisions. Secondly there is no representation in the sense that no module represents any feature of the environment to itself for the purposes of reasoning about it. Rather the module or layer reacts directly to

---

[11] The approach in general has been sustained by the continued increase in computational technology and hence the power of searching and data retrieval. (Brooks 1991b:6)

13

the environment; because perception and action are contained in the same hard-wired module.

It has been charged that Brooks' systems might be merely reactive or instinctual. Brooks (1991b:18-19) responds that mobots do carry out planning in so far as some layers are programmed to take on long term goals. For example in a mobot which is designed to collect soda cans around a laboratory ('Herbert', see Brooks 1991b), the successful identification of a can at some location not within the immediate vicinity, may lead that plan to take precedence in the arena of competing behaviours. Subsumption architecture allows for the fact that a long-term goal may be maintained in conjunction with lower-level tasks. Thus Brooks' idea of planning is the maintenance of long-term goals, not the explicit formulation of plans. A module that is programmed to do a lower level task may be sending signals to other layers which in turn supersede the signal by not reacting to it. The basis for planning thus rests on the fact that some modules may send signals up to ten times before they are received by the modules to which they are connected. There is no explicit planning as no internal state is maintained for longer than three seconds so that, if a can is moved, the robot has no memory a can was there. Thus moving a can will not destroy the mobot behaviour, but will destroy the plan construed as the goal of the robot's behaviour.

The morals as I take them, (in deference to Brooks' stern anti-philosophising) are two-fold. The first concerns his warnings against abstraction and the second the pre-occupation with activity. Brooks' systems do not utilise anything that could remotely be called a symbol. All signals are numeric. These numeric signals are not numbers, but rather an indication of magnitude (transducers). The basis for a lack of symbol processing is that the system is doing the perception (and consequent salient feature extraction) rather than the researcher building the system (code). The approach avoids begging the question of the link between cognition, perception and action. The second moral follows from the first. In so far as there is no representation and no symbols, 'the world is its own best model' (Brooks 1991b:15). The decomposition by activity places pressure on the idea both of central representation, and central control. Complete and explicit representation is computationally expensive and does not allow for quick reactions to environmental changes in the microsecond domain. Brooks' creatures provide instructive examples of how an autonomous agent, with an

architecture that has no representational capacities or central executive capabilities in the traditional sense, can, nonetheless, give rise to coherent behaviour out of 'a system of competing behaviours' (Brooks 1991b:6).

3.2 Brooks and Classical Semantics

To pursue the first moral, the abstraction of cognitive code, semantics will have to be revisited, as illustrated in the following examples. Formerly success was defined by how well a system could represent the world in terms of static entities and trajectories; encoding typically involved recourse to a 'uniform and simple semantics' (Brooks 1991) and search techniques which were then used in planning strategy. Classical systems designed to execute actions usually operate in a highly constrained and carefully built environment, e.g. a world made entirely of specially painted blocks with carefully controlled light (Brooks 1991b). The semantics involved employed reductive techniques in order to render the relatively complex environment in terms of 'pertinent facts' or salient features. For a contrast case, let us look at the navigation system of Herbert in a system of blocks and the navigation of a classic robot. On the one hand, the Classic robot (e.g. 'Shakey', Nilsson 1984) needs continual and computationally expensive updates to its internal models in order to navigate. Herbert on the other hand, even in the absence of visual apparatus, would find the task trivial, partly because he does not need to encode the visual field in order to move, and partly, he is far more robust with respect to the environment as a result of activity decomposition, as he would be able to get around in the dark, even with a visual system.

For more commonplace environments, in the order of physical objects, a Classical coding of a chair, for example, is defined by two characterisations, both true:

(CAN (SIT-ON PERSON CHAIR)), (CAN (STAND-ON PERSON CHAIR)).
(Brooks 1991:3)

This concept however is manifestly missing large numbers of other features that we would also want to include in the concept of a chair. All the relevant and manifold detail has been reduced to a single simple definition. Now while the definition does detail how the object is to be used, it does not specify the way it is to be recognised. Further, the

15

recognition itself requires modelling, rather than immediate relation to the object by a system. The upshot of this approach is that the difficult part of the operation is being cut out of the picture. The abstraction has already been done by the researchers. Whereas 'Psychophysical evidence suggests they [recognition, spatial understanding, dealing with sensor noise, partial models] are all intimately tied up with the representation of the world used by an intelligent system.' (Brooks 1991a:3) Whatever representative techniques are used then have to factor in all these details. Perception, which Brooks relates to abstraction in the Classical approach, is not cleanly divided from reasoning in a real world.

Brooks goes on to what could be seen to be a critical assumption of his work. There is no neat division between perception and reasoning. Perception may well not be a process of abstraction, but may actually use some of the detail which had been previously cut out from the problem-solving activity. The abstraction with which Classical AI models work accounts for their 'brittleness' when introduced into a real-world environment. The difficult part of the abstraction process has already been done by humans in giving the system the 'right' kind of data. An environment in which a system should operate has typically been reduced to a simplified conceptual basis through task-specific abstraction. In the order of goals for Brooks' research, this neither allows the system to deal robustly with an environment, nor make use of fortuitous occurrences to lighten the task load.

Brooks' systems employ modularised perception and action. However, there is still a sense in which action is a response to perception, rather than there being a possibility of sensing activity. The system design does not allow action to be used in the sense that it can be deployed to discover or re-arrange the world for cognitive purposes. Thus a robot may navigate reliably, but will not be able to move an obstacle. Of course nothing stops the addition of further modules, however I suggest that the arrangement or layering of modules themselves will not be sufficiently flexible to cope with multiple purposes in these examples of activity producing robots.

3.3 Active Representations in Subsumption Architecture

Before pursuing projects more closely associated with humans and considering the emphasis on action, arguments so far biased in favour of behaviour at what appears to

be the expense of representation, I turn to an example of action-oriented representation. Mataric's (1991) mobile rat 'Toto' (Brooks 1991b) provides an example of representation which builds on Brooks' activity decompositions. In the Classical approach the rat would have to survey its surrounds and, once a map had been built as a representation of the environment, it could then begin to execute actions within that environment. The Mataric robot on the other hand is built of three layers. The first is a layer attached to sonar that senses the distance from objects like walls and furniture. The second codes for direction against a standard (North), so that it can always tell in which direction it is moving and how far away it is from objects as it does this. The third layer uses a combination of these two types of information, which can be gained as the robot wanders around, to code for points in a network that are a combination of the two factors from the previous two layers. The robot is effectively able, using its sensors, to learn its way about, as it moves. The significance of this is that at the top layer, the map is not static, but made of pairings of lower level features, which not only indicate a position, but also indicate which directions are more easily followed, for the map was built by moving and not by looking and then deciding what to do. The map itself does not require further cognition, but simply is the cognition.

> 'The map is its own user, and its knowledge is both descriptive (of locations) and prescriptive (it represents the relationship between two locations as the sequence of movements that would carry the robot from one landmark to the other). The robot is thus a perfect example of the idea of *action-oriented representations*: representations that simultaneously describe aspects of the world and prescribe possible actions, and are poised between pure control structures and passive representations of external reality.' (Clark 1997:49, italics in original)

Given a location, the rat can, by knowing where it is, connect in the shortest way possible, but the connection itself also tells it how to get there. Here we begin to have something that looks representational, except that it is not symbolic, but rather aimed at controlling activity, built from what are initially active sensory parameters. This is not to deny that there are symbols, as well as symbolic representations, but that they are not specifically and by definition context-independent. The symbols and representations as well as words that we do use, contain elements in their

representational structure of how we can and do use them. Further these elements, as actions, are primary, aimed not at changing the world, but our own cognitive states.

This is not to deny that we have such things as maps and that we can think about them and within them without the aid of moving around. Rather the practices which support the use of maps, are based in action, both from an historical learning perspective and also from micro-temporal perspectives, (twisting the map this way and that, tracing lines with fingers.) The second point is that it is likely that the representational structures and capacities of our heads do not contain the map as an object in our heads, but contain something like a map, superposed on all sorts of other experiences of map-reading as well as experiences of finding our way about. We look at the map for particular land-marks, so that we will know where we are when we actually find them.

3.4 Pursuing Human Interaction

Much of Brooks' work has been done on robots which have tried to emulate insect level intelligence. Even the social situations that have been modelled have mostly been of the insect persuasion. However there is suggestion that in the modelling of human level intelligence we will have to build systems which can, amongst other things, detect faces, distinguish human voices, make eye contact, follow gaze, understand where people are pointing, interpret facial gestures and respond appropriately to eye contact. (Brooks 1996:1)

Connected with this leap in a system's ability to maintain more than one (call it) global goal, are the questions of coherence and self-adaptation. The robot will have to judge for itself with respect to various sensors and to the intensity of motivation, whether it should re-direct it's attention or not. For example in playing with an infant robot, the robot cannot be either too distracted by the things going on around it, nor can it be so intent on the task at hand that it is not monitoring the goings on about it at all. The final bar seems to be that humans interacting with robots which displayed these behaviours would not be able to treat them as human, and would probably surmise that they were ill or that something was wrong with them, or alternately that they acted like aliens. Thus a naïve subject may be confused by a robot that is interested in shiny spots of light of a particular

source intensity, but not voices or its own hands. The challenge as Brooks sees it 'is to identify the appropriate signals that can be extracted from the environment in order to have this adaptation happen seamlessly behind the scenes.' (Brooks 1996:3, Section 5.1)

In response to these challenges Brooks has begun to build a humanoid robot, COG, that currently consists of head, torso and two arms as well as hands. The purpose behind building Cog is to provide it with the ability to learn rather than be programmed, and thus the emphasis is on building perception and activation mechanisms. The experiment with Cog has been not to encode specific knowledge but to try and build active ways in which a robot may be able to find out things about the world it is in. Thus the focus of the project is not on the brain, but on the so-called peripheral systems, 'perception and motor skills' (Brooks 1991a:2), like hands and eyes. Systems and units are built that are able to follow general human like strategies (e.g. Kismet, Breazeal and Scassellati 1998). For example the robot is able to foveate, to localise sound and then to integrate the two on the assumption that noise is usually accompanied by movement. (Brooks 1996:5) Neural networks have been used to correlate particular aural locations with visual locations, such that an event which is heard is mapped onto the appropriate place to look for that event. The idea is not to programme the robot but to provide it with sufficient degrees of freedom and the ability to follow what is going on around it, so that it can learn, rather than be programmed. The emphasis then is not on programming but on strategy, and in the following case we examine some novel and active strategies used by human players in the game Tetris.

4. Kirsh and Maglio play Tetris

I have begun by considering how machines develop rudimentary intelligence. Following Brooks, it has been shown that, in designing robots, it is dangerous to begin with *a priori* forms or process idealizations. To increment the capabilities of intelligent systems, it is argued, one must abandon distinctions between action, perception and reasoning. This is because, working from the bottom up, there is no easy way to ensure that posited pieces or interfaces will be valid (Brooks 1991a:1). To achieve the robust responsiveness needed by robots in the environment, Brooks rejects traditional views of representation as using the 'wrong unit of abstraction.'(Brooks 1991a:1). In his 'mobots' the central executive of serial, digital computers is replaced by a subsumption architecture. This allows them to function, non-serially and without explicit coding. The machines achieve their engineered goals by exploiting real-time dynamics arising from how what is sensed affects sub-agent systems that control action. Not surprisingly, these robots differ from symbol-manipulating machines in that, like us, they show flexibility.

To expand on the notion of flexible action and a lack of representation we begin to look at the ways in which human agents make seemingly odd moves within the specific task domain of Tetris. Kirsh and Maglio (1994) argue that a certain class of actions, called epistemic actions take place, and the fact that they do, affects the ways in which we traditionally conceive of cognitive modelling, suggesting a cognitive model which looks more like Brooks' subsumption architecture. Firstly, I will clarify the new class of actions: what they are and how they work. Building on some of the definitions and clarifications gained here, I will look at some of the ways in which the principles associated with epistemic action (essentially non-sequential models of cognition that accommodate more than pragmatic goals) can be exploited in generalising across cases outside of what is a neatly and experimentally described task domain.

Kirsh and Maglio's paper 'On Distinguishing Epistemic from Pragmatic Action' (Kirsh and Maglio 1994) is an attempt to recognise a further category of action which is not pragmatically motivated and also does not fall into the perceptual category. Perceptive actions are those actions which have been construed as the motor part of a control of gaze. (Kirsh and Maglio 1994:5, Kirsh and Maglio 1992b) These have been divided into two domains: control of attention within an image and control of gaze; 'the orientation

and resolution of the sensor... to create a *new image*.' (Kirsh and Maglio 1992b:1)[12] Pragmatic action is action which is carried out in order 'to bring one physically closer to a goal' distinguished from epistemic action which is 'performed to uncover information that is hidden or hard to compute.' (Kirsh and Maglio 1994:2) The importance of this is that motor action, rather than the motor part of visual attention only, is not only seen as the product of cognition, but may aid or facilitate cognition.

4.1 Playing Tetris

To get the flavour of scaffolding, action loops and the prominence as well as priority of epistemic action, let us look at a detailed example from the research of David Kirsh and Paul Maglio on the computer game Tetris. Tetris provides examples of action that are not the product of plans produced from complete sensory information even in a game as pragmatically orientated as Tetris.[13] Tetris is a game played on a computer in a field of thirty squares by ten. Shapes, called tetrazoids by Kirsh and Maglio, composed of combinations of four blocks, result in seven different 'zoids': a square, a long piece composed of four squares in a row (|), a square composed of four blocks (□), a T-shaped zoid, an L shape and an S shape zoid, each of which also have reverse formations (mirror images). These shapes drop from the top of the field, and can be moved either right or left (translated), or rotated 90 degrees in a clockwise direction. Thus the block has one orientation, the long piece and the S shaped pieces two, and the L and T shaped pieces four orientations. As they drop from the top of the screen these zoids are fitted together at the bottom of the screen. Using translations and rotations they have to be placed so as to reduce the number of free spaces between the blocks such that on the completion of a horizontal line of ten blocks in the field, (composed of different zoids fitted together) it disappears allowing the game to go on, leaving more space (and consequently time) in which to place the blocks. Scores are awarded for the number of complete rows and the game ends when no more zoids can enter the screen because the blocks have been poorly fitted, so filling the field and no further zoids can be placed. The randomly selected zoids enter the top of the screen in random orientation. The game also speeds up as it proceeds, giving the player less and

---

[12] See Maglio et al. (2000) for some applications to user interfaces.
[13] Tetris was originally created to keep cosmonauts occupied in space. It has proved highly addictive to many humans.

less time to identify a zoid, decide where to place it and then to place it (execute the decision). The game is an effective research domain because the criteria for what counts as a good placement is determinate, (least number of open blocks) and it is easy to judge a better or worse game, (more or less completed lines).

Now how should a system play such a game? Kirsh and Maglio use a fully sequential model taken from classic AI, 'RoboTetris', as a source of contrast with human performance. An actual system that plays Tetris, has been implemented as a contrast to the human player. The model has four steps. First, an early bitmap representation is created. Second, the bitmap is encoded. Third placement is computed and fourth, moves are planned to achieve goal placement (Section 2)[14]. On this model no step can happen until the previous one has been completed without being random. Now, given the time sensitive nature of the game, speed would seem to depend entirely on the internal computational capacities of the system; the faster the internal computation, the faster the placement at the end of that computation. The above model generates a prediction in which one would expect a zoid to drop from the top of the screen and then, after a delay for processing, be swiftly and correctly placed in the minimum number of moves. However, among human players, 'This is patently not what we see in the data. Rotations and translations occur in abundance, almost from the moment a zoid enters the Tetris screen.' (Kirsh & Maglio 1994:15) Human players' behaviour does not fit the predictions offered by a sequential model ('RoboTetris'). Rather players make moves (perform rotations and translations) before the zoid is identified, or a candidate placement has been selected.

The model based on classic AI cannot predict the above outcome and consequently cannot account for actions that do not seem to be planned and therefore are not pragmatic. A model that only takes into account pragmatic goals rests on a fairly strong distinction between mind, body and world. The world is apprehended by perception, cognition takes place and then action is undertaken in order to effect a change in the world. Action here can only have one function. How then do we begin to account for these types of actions, which Kirsh and Maglio call epistemic actions? One way to do so is to look for the benefits of these 'superfluous' actions. For from the data, expert

---

[14] The model bears relation to Brooks (1991b) model of Classical Robotics; *'the sense-model-plan-act framework, or SMPA for short.'* (Brooks 1991b:2, italics in original)

players make more rotations and translations and start making them earlier than either RoboTetris or novice players. It is not only that there are more actions, but also that they are characteristic of better players. The benefits have been cashed out in various ways:

- Unearth new information very early in the game,
- Save mental rotation effort,
- Facilitate retrieval of zoids from memory, (Kirsh and Maglio 1994:20)

To take one example of the several ways in which Kirsh and Maglio make the case for the existence of epistemic actions[15], I will look briefly at how rotation is able to facilitate the disambiguation of zoids with mirror images (S and L shapes).

The experiment conducted used two screens in which zoids that could possibly be mirror images of one another were dropped from the top of the screen. The subject needs to identify whether they are the same zoid in different orientations, or mirror images of one another. They should then press one button if they are mirror images of one another and a second if they are not. When asked to identify the zoids without the aid of pressing a rotate button, identification happens in the region of 800-1200ms (1994:24) With the aid of the rotate button, this naturally can happen faster. The task has been simplified. A zoid that does not appear congruous through rotation cannot be the same zoid. The economic reasoning based on this experimentation is simple; the rate at which a figure can be disambiguated without the aid of action (i.e. internally, without pressing 'rotate') is in the range of 800 to 1200 milliseconds. The rate at which a figure can be rotated by physically pressing a button (even allowing time for the selection of the button of 200ms) and thus disambiguating it is in the order of 300 to 500 microseconds. Generously it is twice as fast.

---

[15] The extensive survey and experimentation provide other possible functions for epistemic action. Early Rotations for Discovery (pg. 21-23) Rotating to save Effort in Mental Rotation and Mental Imagery (pg. 23-26), Rotating to Help Create an Orientation-Independent Representation (pg. 26-30), Rotating to Help Identify Zoids (pg. 30-34), Rotating to Facilitate Matching (pg. 34-35) Translation as an Epistemic Action (35-38). Some of the experiments suggest that epistemic action either reduces the need for memory of different orientations of zoids, or aids the building of such memory. See also 'Epistemic Action Increases with Skill' (Maglio and Kirsh 1996, Section 4.3)

Analysing the situation we see that there is a tight coupling between the cognitive system of humans and the screen on which the game is being played. Based on an experience of playing the game, certain actions are found to have benefits that are other than pragmatic. (Few players, even expert players, could tell you why they rotate so early.) The rotations carried out are not geared primarily toward placement, but toward changing the cognitive state of the agent, to better enable placement. The screen, as part of the game, is also used to play the game. So in rehabilitating the original sequential model, placing action in the first stages shows that the stages cannot be neatly separated sequentially. For the screen is being used as scaffolding in order to change the cognitive state of the system, through action loops, which is distinctly faster and more efficient than trying to do it exclusively internally, which, incidentally, is what novice players try to do. Here, the world features as part of the cognitive system, rather than being external to it.

4.2 Epistemic Action, Scaffolding and Environment

What makes Kirsh and Maglio's paper (1994:3) different from other accounts of scaffolding (Rumelhart et al. 1986) is that they are conducting research in an area which is not explicitly symbolic. Further it is action taking place in a very time sensitive context or task-space and is not related to an expert activity, where special knowledge is required, (although there are expert Tetris players). This is assumed to be a closer reflection of the every day demands which are made on human cognition: conditions that commonly reflect a dynamic environment and a need for time sensitive action and information. Secondly, due to time constraints, the scaffolding which takes place will not be of the type convenient to traditional AI search methods. This is because usually scaffolding takes place in the form of external and structured storage and manipulation of explicit symbols. It is scaffolding attached primarily to memory storage. (Diaries and wall planners are a good example of structured and external memory storage.) So in the Rumelhart et al. (1986) example of long-division, external scaffolding is a structured storage mechanism of simpler steps in a longer sequence aimed at solving a more complex problem. Epistemic action begins to describe scaffolding in ways that are sensitive to the internal state of the agent and the possible actions that the agent may carry out in order to uncover salient information.

The distinction between pragmatic and epistemic action can be seen as goal relevant. Pragmatic actions are those which are effected in order to carry out physical goals; this would include reaching for a pen, or switching a kettle on; thus actions which cannot be construed as physically or pragmatically goal relevant appear unmotivated and superfluous on a sequential model that tries to find the shortest possible way or least number of moves toward a goal. It assumes that all cognition is internal. It is this gap in the understanding of action that epistemic action is supposed to fill. Kirsh and Maglio's thesis is that epistemic actions may well overlap with pragmatic action[16] and that action may in some cases even precede cognition. The main point is in revealing that agents are not at all passive in extracting information, but that they actively extract and manipulate the information which is most salient. Epistemic actions 'are not performed to advance a player to a better state in the external environment, but rather to advance the player to a better state in his or her internal, cognitive environment'. (1994:38) In the research they show that there are very few unambiguously epistemic actions, but that they are logically distinguishable from pragmatic action and further that they only appear motivated when we take into account the existence of goals other than pragmatic ones.

4.3 Epistemic Action in Skilled Tetris Players

Now it might seem that epistemic actions are only needed when we are learning to perform a particular task e.g. chess. However a further claim arises from the research in Tetris, namely that 'Epistemic Action Increases With Skill' (Maglio and Kirsh 1996). Intuitively, when looking at what counts as skill in Tetris, expert players should make fewer mistakes and therefore should make fewer backtracking and redundant moves. However Kirsh and Maglio present evidence to show that while following the power law of practice (Newell and Rosenbloom 1981)[17], in Tetris, '(a) Certain sorts of backtracking increase as skill develops; and (b) despite this increase, Tetris skills resembles other skills in following the power law of practice.' (Maglio and Kirsh 1996:1) These results

---

[16] For example, Rotation that aids discovery of the orientation of a zoid may at the same time enable placement. The fact remains that it is still faster to physically rotate zoids than to try and rotate them internally.

[17] 'Typically, practice improves performance in accordance with a power function of practice time or practice trials (Newell and Rosenbloom 1981) either by decreasing the time to react to stimuli by taking a single action (Seibel 1963), or by decreasing the overall time it takes to perform a task that requires a sequence of actions (Crossman 1959).' (Maglio and Kirsh 1996:2)

support the claim that more, rather than less, epistemic action is involved in skilled behaviour. Certain backtracking moves are not mistakes, but lead to better play.

The data presented is a longitudinal study of subjects acquiring the skills needed to play Tetris. Three types of data were recorded: the time taken to place a zoid from entering the screen, the time taken between actions and the time taken for the first action to be executed. 'In all three cases, our data indicate that performance speeds up according to the power law of practice.' (1996:2) Intuitively as players become more skilled in the placement of zoids one would expect the number of redundant moves to decrease. However, *the number of apparently extraneous actions increases with practice.'* (1996:2, italics in original).

The domain of Tetris allows for neat definition of what counts as a backtracking or redundant move. Backtracking moves are ones which have to be undone later in order to place the piece. So backtracking can be defined as the number of moves taken to place a piece less the minimum number of moves necessary to place the piece. This definition and data allows two further observations. First backtracking cannot be a product of simple motor errors. The data suggests that extra rotations are not made randomly for different zoids, but are made in the case of particular zoids. The pieces rotated more often seem to be those that have mirror images (S and L zoids) and is supported by data from previous experiments (1994) as these are the most difficult to disambiguate. This allows a second observation. Extra rotations are made *before* the decision to place the zoid has been made. This is established due to the fact that the correct orientation happens earlier in a rotation cycle than the decision to place it can. The decision to place a zoid does not generally happen before 1130ms (1996:2) and therefore the desired orientation is normally passed 400-500 ms earlier (1996:2), necessitating extra rotations.

Kirsh and Maglio's conclusion is that the combination of the fact that Tetris players follow a power law of practice, as well as make more backtracking moves, 'support our hypothesis that redundant actions are epistemic actions which both simplify perceptual computation and play a natural role in skilled behaviour.' (1996:3) The evidence above provides argument for the fact that because Tetris follows the power law of practice (like most other skills), and exhibits a greater reliance on epistemic action, epistemic action can be expected to play a greater and more general part in normal human cognition than was formerly expected.

4.4 Epistemic Action in Reflection

Now it might be argued that epistemic action can only really function in a more or less reactive way, in an environment where the outcome for any given action is reasonably determinate. Kirsh and Maglio (1992b), do however offer some suggestions as to how their category of epistemic action can be coupled with reflection. The suggestions are useful in trying to show how systems designed to cope with real-time dynamics and complexities, such as Brooks' robot systems, can be combined with what we know of human Tetris playing strategies 'because truly reactive systems tend to be immune to top-down interference.' (Kirsh and Maglio 1992b:1)

Brooks' systems present a case that looks purely reactive. And although Brooks has argued against this view, his systems are composed of hardwired input-output modules (Brooks 1991a). The claim that they are not purely reactive rests on the basis that 'input to a layer may be suppressed, output inhibited or augmented.' (Kirsh and Maglio, 1992b:1) Essentially the output signal of a module can be ignored by other modules, or the input signal not taken into account in reaction. 'The processes occurring inside each activity layer are well insulated and modular, sealed off from the computations occurring in other layers.' (Kirsh and Maglio, 1992b:1) The signals then can only have effects in the modules to which they are wired, and while limiting cross-talk to the periphery, does not allow for variable effects within the system, but only in the behaviour of the robot as it interacts with the world.

Similar claims of 'information encapsulation' (Kirsh and Maglio 1992b:1) have been made for skilled behaviour in humans, positing the possession of a skill as the product of largely automatic processes which are relatively immune, or robust with respect to interference. These have been explained with reference to chunking, an explanation which works well with a linear and pragmatically driven model. Evidence from typists, (Gentner 1988) supports these claims. The claim is that behaviour which is skill driven, is under this definition, running parallel to other systems in a largely automatic fashion and 'insensitive to outside information' (Kirsh and Maglio, 1992b:1). Brooks' robots provide examples of largely automatic processes. Now it should be reasonably clear that, given the evidence for epistemic action in playing Tetris, we need a model which can

integrate the information provided by epistemic actions into skilled and practiced behaviour.

Kirsh and Maglio define a skill as 'an error reducing control mechanism built on a statistical model of the environment.' (Kirsh and Maglio, 1992b:1) Thus on the basis of practice in an environment the agent, through building a model (unconsciously and implicitly or explicitly[18]) can make (imperfect) predictions of the 'statistical structure of the inputs it confronts and the effects of the actions it can take.' (Moray, 1986)

> 'Central to this notion of skill is the idea that behaviour is perceptually driven – since errors are perceptually discernable – and goal specific – since the goal of reducing differences is intrinsic to a skill. Thus skills do not set goals, they adaptively carry them out. They rely on their perceptual representation of the current situation and their implicit model of the domain to respond adaptively. Activity layers in mobots qualify as skills according to this definition.' (Kirsh and Maglio, 1992b:2)

As before, Kirsh and Maglio make use of RoboTetris as a contrast case to human players. Despite the fact that there are 17×2^{270} possible Tetris states, RoboTetris 'represents a board [field] situation by a vector of six features: number of holes present, total board height mountain height, covered holes, filled rows, and local fitness.' (Kirsh and Maglio, 1992b:2) These features are based both on the verbal information gained from expert players and data gathered from their performance.

RoboTetris does outperform any human player, but there remain some interesting discrepancies. For instance RoboTetris is far more inconsistent than intermediate and expert players. Further the profile of highest peak to lowest valley is also unnaturally high. Kirsh and Maglio admit that these performance profiles may be adjusted by a better set of weights on the six features, but as it stands, due to RoboTetris unreliable behaviour, they take it 'as an indication of the limitations of a skill based approach.' (Kirsh and Maglio, 1992b:2)

---

[18] Kirsh and Maglio consider the explicitness or implicitness of the model to be an implementation detail, although they assume that the model is implicit. (Kirsh and Maglio, 1992:2)

What becomes important in trying to understand the reliability of human performance is the report of expert players of an overall strategy in the form of a '*set of concerns*' (Kirsh and Maglio 1992b:2, italics in original). It is not however immediately apparent what an expert would mean by a concern for 'flatness' (in the profile of the placed zoids in the screen). Nor that the relative importance of flatness is maintained throughout. For example, based on a current estimate of ones ability, an expert may go mountain building, completely disregarding the concern for flatness. '[W]e take the varying nature of concern with flatness to be a key empirical indicator of higher level control.' (Kirsh and Maglio, 1992b:2) Thus flatness which should be a perceptually discernable state of the game, in line with the definition of a skill, turns out not to be always perceptually identifiable, or reducible to perceptual properties, given that it seems to depend on the state of the agent.

The variable nature of the discernability of the policies of the agent in task-space appears to provide some argument against a theory of skills which makes them hostage to automatic processes that can be explained by chunking. In coupling reflective systems to reactive systems it is commonly assumed that the high level planner can either suppress input or override the output of the reactive layers, (Kirsh and Maglio, 1992b:2) 'Little attention is paid to actively redirecting the sensors of the system to bias the input stream.' (Kirsh and Maglio, 1992b:2) The inability to bias or redirect attention appears to create a distinction between the skills that humans display in playing Tetris and the skills typists have, or that Brooks' mobots display in their robust and reactive activity in an environment.

An attempt can be made to account for the disparity between stated concerns or global strategy and the inability to discern these in the playing field of Tetris. The account would seem to need recourse to a set of high level representations reflecting 'a global perspective' (Kirsh and Maglio, 1992b:2)

> 'Policies identify how things ought to go in the course of a game. The job of the planner is to monitor for significant differences between how things ought to be and how things are. These significant differences are concerns – high level descriptions of respects in which the reactive agency is falling short of implementing the

system's policies. Once concerns are identified they are translated into directives for changing the focus of attention.' (Kirsh and Maglio, 1992b:2)

Skills then are linked to flexibly deployable policies. These flexibly deployable policies do hinge on the state of the game at the moment, but must also take into account the state of the agent playing the game. Skills here are not automatic processes, but reflect higher level policies and concerns, where concerns are deviations from policy. For example, if the profile of the placed zoids is not sufficiently flat, then attention is redirected to solve this concern; seeking to place pieces in such a way as to reduce the contour. This provides at least one way in which human Tetris players are different from Brooks' mobots. Brooks' robots cannot undertake actions to repair the failure of a higher level policy, i.e. the behaviour or purpose remains in tact, but there are no variable policies within the behaviour. This is in part due to the fact that because the sensors of the robots are fixed to detect particular things, they cannot redirect there attention according to different policies or sets of concerns within the performance of its purpose. Epistemic actions used to aid the execution of a task cannot be utilised by Brooks' mobots. Human agents on the other hand use actions repeatedly in the execution of tasks. The argument for higher level policies can in part be made by a greater reliance on epistemic action, and therefore skills, when related to policies, need not be always and entirely automatic. Epistemic actions can reflect the higher level policies and concerns of the agent, which is in turn a reflection of variable higher level control.

4.5 Models and Planning

'Epistemic actions are actions designed to change the input to an agent's information-processing system.' (Kirsh and Maglio 1994:38)

The definition of and arguments for epistemic actions have several implications, not least of which is the inability of a sequential model of cognitions to account for the abundance of rotations and translations in playing Tetris. A fully sequential model of cognition aimed entirely at pragmatic actions simply cannot account for the types of moves that aid cognition in Tetris, making expert players faster and therefore better. In order to accommodate these types of moves we have to begin to look at models which are not feed-forward only, but are capable of featuring recourse to the world through

action at any stage in the process. The models generated by Kirsh and Maglio's arguments for the category of epistemic action result not in a simple input-output system, but a system which in a single task can make repeated calls to the world through action in order to augment and facilitate cognition.

Kirsh and Maglio present a functional model (1994:42) based on Minsky (1986) which begins to look a lot like a model generated by Brooks' subsumption architecture. Naturally given the fact that it is not purpose-specific in the ways that Brooks' layers or modules are, it is potentially more general than Brooks' 'horizontal microworlds' would allow. The main feature of this model is to break down the input-output model of sequential reasoning resulting in pragmatic action which is normally assumed to be the only type of action possible. Here problems neither occur nor are they solved primarily within the agent, but are solved by repeated actions in the world, which result not in the accomplishment of an immediate goal, but in the solving of epistemic problems which facilitate the continuation of the task.

'[The model's] chief novelty lies in allowing individual functional units inside the agent to be in closed-loop interaction with the outside world.' (Kirsh and Maglio 1994:38) The use of functional units do not undermine Brooks' proposal, but provides a potentially more general case of subsumption architecture underlying his mobots. Further it exposes some of Brooks' assumptions. It is not necessary that the sensing and actuation be in the same module, although this may be better for his behaviour based mobots, but rather that the functional units have multiple outputs that result in non-sequential cognition, where non-sequential means the ability to exploit the external world through repeated actions loops and redirections of attention, rather than process insulated information automatically or entirely internally.

One major reason for the apparent success of Kirsh and Maglio's arguments is

'The property of Tetris that makes such a strategy pay off is that the local effects of an action are totally determinate. There are no hidden states, exogenous influences, or other agents to change the result of hitting the rotate key. There is a dependable and simple link between motor action and the change in stimulus. Consequently, a

31

well-adapted attentional mechanism might incorporate simple *calls* to the world as part of its processing strategy.' (Kirsh and Maglio 1994:40, italics in original)

32

The determinacy of the system interaction provides evidence for epistemic action even in the face of pragmatically oriented goals. How then do we begin to generalise the category of epistemic action to task spaces that do not allow for such determinate results to any given action? Tetris results are characterised in part by the determinacy of the results of a key-stroke. Here epistemic actions are used in the pursuit of pragmatic goals. Now if we begin to examine task spaces where the result of actions is not determinate due to the influence of extraneous factors as well as other agents, the need for more timeous information grows. This will lead to more general applications of epistemic action aimed not ultimately at pragmatic goals, but at maintaining epistemic goals, where epistemic goals are construed as being able to gain sufficient information to carry out a task.

5. Emotions in Interaction

Thus far we have covered some of the ways in which action can be used to carry out cognition in the presence of insufficient information. We have also looked at some of the ways in which epistemic action can be used in the service of flexibly deployable policies in the domain of Tetris that argue for higher level control of cognitive goals. What I propose to do is show ways in which emotion can aid in delimiting search space. In doing so, I will be dealing with emotion in a rather narrow fashion, that is in so far as it effects the pursuit of cognitive goals. This approach leaves out, amongst other things, the evolution of emotion and the experiential quality of emotion. Stated positively, emotion will be discussed with respect to its ability to highlight salient information without the aid of a search algorithm or decision tree. To demonstrate some of the progress being made in this area in relation to active cognitive systems, I will start with a more empirical project in the form of Cynthia Breazeal's 'Kismet' (1998). Kismet is a robot that is based on Brooks' (1991a,b, Brooks et al. 1998) guidelines for architecture, but involves recourse to emotions, in the form of facial expressions and vocal tones, that are basically driven by higher level policies that can aid communication with a human interlocutor, even in the absence of phonemic processing. Further these can be shown to explain ability to scaffold cognitive and interpersonal space. Secondly I will use Dylan Evans 'Search Hypothesis of Emotion' (2002), to provide some general clues to the ways in which emotions, whatever they may be, can be construed as positively aiding cognition by narrowing the search space in human problem solving. Lastly I will look at Ross and Dumouchel's 'Emotions as Strategic Signals' (forthcoming), arguing that emotions rather than narrowing the search space, positively highlight significant trends in interaction adding to the context of the conversation, such that the participants are able to negotiate the limits of the goals of dialogue.

5.1 Teaching Kismet

One example that begins to highlight some of the complexities of interaction among cognitive systems is Kismet (Breazeal and Scassellati 1998). Kismet is an autonomous robot system designed to model the interactions of caretaker-infant dyad. Building on and complimentary to the system built for COG (Section 3.3), Kismet makes use of a schematic face that is able to respond in real-time to a naïve subject in social interaction.

Kismet's architecture is heavily indebted to Brooks subsumption architectures. Further, emotions are used here to solve the problem of socially situated infant learning.

The architecture implemented is an agent-based architecture. As we have seen each module is not responsible to a larger integrated/integrating module, but rather separate modules are responsible for well-defined tasks. The inputs and outputs are therefore contained within a single module and affect other modules. Thus 'the process is active when its activation level exceeds an activation threshold. When active, the process may perform some special computation, send output messages to connected processes, spread some of its activation energy to connected units, and/or express itself through behaviour.' (Breazeal and Scassellati 1998:7)

Kismet is composed of five distinguishable subsystems: the perception system, the motivation system, the attention system, the behaviour system and the motor system. (Breazeal and Scassellati 1998:7) Aside from the implementation details of the project, what is important is the way in which the emotions are used to satisfy drives within the Motivation System. Here emotion is used in the form of an expressive state, to constrain and structure behaviour in the interaction with a human subject.

The motivation system is composed of two subsystems: drives and emotions[19]. Drives serve three purposes:

1. They influence behaviour selection by preferentially passing activation to some behaviours over others.
2. They influence the emotive state of the robot by passing activation energy to the emotion processes. Since the robot's expressions reflect its emotive state, the drives indirectly control the expressive cues the robot displays to the caretaker.
3. Third, they provide a learning context; the robot learns skills that serve to satisfy its drives. (Breazeal and Scassellati 1998:8)

Currently the robot has three basic drives: Social Drive, Stimulation Drive and Fatigue Drive. The drives come in three basic regimes, underwhelmed, overwhelmed and

---

[19] As in Breazeal and Scassellati (1998), courier type words refer to system implementations, rather than the general uses of the words eg. 'emotion'.

34

homeostatic. So long as the system `drive` is in the homeostatic range of the robot, the needs of the robot are being met. If the drives are not being met, the robot can use emotional cues to bring the interaction back in range. If the social drive is overwhelmed, the robot will begin to avoid face-to-face contact, becoming `asocial` and if it is underwhelmed, it will become `lonely` and try to seek out and establish face-to-face contact. Similarly the stimulation drive operates between `bored` and `confused`. The cues offered by this drive encourage the caretaker to try new and novel things when the robot becomes `bored`, or too slow down when the robot becomes `confused`. The fatigue drive does not operate in a range, but slowly moves toward `exhausted`. Exhaustion means that it can no longer regulate its activity with the world and goes to 'sleep'. In the long term project this will, hopefully, be the time in which the robot is able to integrate and consolidate what it has learned in the interaction. When the drive returns to a homeostatic range, the robot 'wakes up'.

In addition to basic `drives`, the robot has `emotive and expressive states`. The emotions serve two functions:

1. They influence the emotive expression of the robot by passing activation energy to the face motor processes.
2. They play an important role in regulating face-to-face exchanges with the caretaker. (Breazeal and Scassellati 1998:9)

The connection between the `drives` and the `emotions` is one in which the `drives` establish the `emotional` state of the robot, which is directly shown through its facial expressions. The legibility of the emotional state of the robot constrains and provides clues about how the human caretaker should go about fulfilling the `drives` of the robot.

The robot currently has eight emotive and expressive states. These include five primary human analogue emotions: anger, disgust, fear, happiness and sadness. (Breazeal and Scassellati 1998:15) In addition the robot has three states that are not typically thought of as emotive in the human case, but are none-the-less important responses to events in human social interaction. These are surprise, interest and excitement. These states are

considered important in regulating social interaction as they provide clues as to how to go on in the interaction.

Emotions are typically expressed in response to the homeostatic range of the robot's drives. Thus when the robot is in an acceptable range, positive emotions such as happiness and interest are expressed. When the robot is either underwhelmed or overwhelmed it will display emotions with negative connotations. These are taken as a sign (by the caretaker) that something in the interaction needs to be changed in order to make the robot happy. Particular emotions give additional cues to the caretaker as to what to do. For instance, when the robot is socially overwhelmed, it will begin to show signs of disgust, which, interpreted as asocial behaviour, is a cue to relax the intensity of the interaction. In non-social interaction, overwhelming motion may result in the robot showing fear, a typical response in an infant who is confused by the amount of motion in its environment. Note that the expressive responses are determined in relation to drives. For instance if the robot is tending toward the exhausted end of the scale of the fatigue scale, interaction which would normally have provoked a positive response, may now provoke a negative response, like disgust, which may in turn descend into anger because the caretaker is essentially not allowing the robot to go to 'sleep'. The caretaker in turn may interpret this as the robot 'acting "cranky" because it is "tired".' (Breazeal and Scassellati 1998:16) One necessary property of the system is the ability to show emotion relatively quickly, for instance surprise at an event in the robot's visual field needs to be shown strongly and quickly.

Some early experimental data on short interactions have yielded quite interesting results. For example in an interaction with a person waving, the robot initially shows interest and will continue to show interest within a homeostatic range. However if the waving becomes to strong or vigorous, the robot begins to show fear, signalling the person to regulate the intensity of the motion to bring the robot back into homeostatic range. Due to the robot's ability to distinguish between faces and non-faces, in interactions with a slinky, the robot does not only show fear, but eventually begins to get angry if the intensity of the fulfilment of the drives is too high. Alternately in interactions with faces, the robot will at first show disgust, becoming angrier as the interaction is continued too long at too high an intensity. Generally the robot is 'happier' playing with people, than

with its toys, as it fulfils more drives, although it is still 'interested' in its toys. (Breazeal and Scassellati 1998: Figures 13-18, for graphs showing activation levels of robots `drives` and `emotions` in interaction.)

This project is one manifestation of the way in which a fast and responsive architecture is being used to provide an analogue to human social interaction. Again there is no need for representationally invested accounts of the system, as the system makes use of transducers, rather than code-like categorisations. Even in the absence of phonemic processing, the robot is able to interact legibly with a human subject. `Emotions` are used to fulfil `drives` of the robot, essentially maintaining a learning environment. Subjects in that environment are not passive parts of the environment, but are actively sought out in order for the robot to ultimately learn. The expressive acts of the robot aim at being legible to human subjects, effectively giving the human subject cues as to how interact with the robot. The expressive states here are not pragmatically motivated but are aimed at maintaining a state of affairs congenial to the robot's `drives` being fulfilled. The `emotional states` of the robot are not arrived at by a decision, but are in direct response to what is going on in the world, and more importantly what is going on in the interaction with either a human subject or its toys. Kismet's emotions serve to signal ways in which to interact appropriately in order to fulfil its drives which are instantiations of the pragmatic goal of social learning. The goal of Kismet's `emotions` is however epistemic, aimed at maintaining a suitable level of interaction to facilitate the goal of the current system.[20]

The drives themselves can be see as goals or policies expressed through emotional states. Emotional states then provide cues to the fulfilment of these goals. When the interaction is proceeding according to a normative policy, then the robot merely encourages the interaction. When the interaction begins to deviate from the fulfilment of the robots policies, this creates a concern which is addressed by the providing emotional and

---

[20] The Kismet system does not as yet learn anything. It, like Brooks' mobots, does not have a memory. The architecture differs from a classical architecture in that the interactions do not have to be sequential, nor do data stores describing emotions have to be programmed. The emotions expressed by humans and legible to the robot do not have to be categorised. The robot is able to function without any representational architecture as traditionally understood. It interacts legibly because of sophisticated peripheral systems. It is still an open question as to whether any learning in a system of this kind can take place. In line with Brooks' programme, the system begins to provide ways in which a learning environment, approaching the natural learning environment of an infant, can be maintained.

affective clues to the subject. This in turn creates expectations that the behaviour that produces such a concern will in fact be altered. As we will see these kinds of clues provide insights into the function of emotion in interaction.

5.2 The Search Hypothesis of Emotion

Turning to a view that details way in which emotions can aid problem-solving, Dylan Evans' 'Search Hypothesis of Emotion' (2002) is an attempt to clarify at least one way in which emotion can aid cognition, effectively trying to free cognition from mere computation in the face of *a priori* emotional subjective utility value. In the process he distinguishes between what emotions are supposed to do and what they are. For the purposes of this paper, what emotions are supposed to do is more important than what they are. According to Evans' reading however, we do need some idea of what emotions are in order to evaluate his theory and so I will try to show what emotions might be in the context of this theory, using restricted and affective definitions of emotions, showing what they might be able to accomplish in interpersonal dialogue.

From the perspective of rational decision making and planning, emotions have essentially been viewed in a negative light. Human reasoning is better as far as it is unaffected by the emotions. Evans calls this the 'negative view of emotions.' (2001) In contrast Evans, with Frank (1988) and Damasio (1994) have begun to argue for a positive view of emotion that affects reasoning for the better. 'The positive view suggests, that other things being equal, humans will be *less* rational to the extent that they lack emotion.' (Evans 2002:498, italics in original).

As Evans explains, from Hobbes onward, emotion has been seen as the way in which the desires or ends of agent are fixed, quite aside from rational decision-making. Rationality is limited to finding the most efficient means to an end, but cannot fix ends. 'Reason is thus reduced to computation, in true Hobbesian fashion.' (Evans 2002:498) The assumption in rational choice theory has been that reason is able to fix the possible outcomes in any given scenario, to which the emotions can attach a utility value and reason can then set about calculating the most efficient path to utility satisfaction and maximisation. However it turns out to be a non-trivial task to be able to fix the possible

outcomes of a scenario. Given that the effect of a decision can have progressively burgeoning outcomes based on this outcome and the decision to pursue it, the task quickly turns into too many alternatives and possible scenarios from which to choose. Nor can we solve the problem by assigning a utility value for time, because this decision would itself require a decision leading to an infinite regress.[21] Due to these problems in what appears to be the inability to fix ends, Evans claims that 'emotions,..., play more than one role in rational choice. Not only do they assign subjective utility to each outcome; they also delimit the range of outcomes to be considered.' (Evans 2002:500)

Now it may seem that emotions can be the answer to the search problem, where problem solving is construed as the ability to search (Newell and Simon, 1976). Rather than generate a complete set of possibilities, we open the search space step by step, looking for acceptable solutions at every stage. If the solution is acceptable, the search is terminated. If the solution is unacceptable, the search continues to expand. What this means is that 'Emotions prevent us from getting lost in endless explorations of potentially infinite search spaces by providing us with both the right kind of test and the right kind of search strategy for each problem that we must solve.' (Evans 2002:503) The search strategy employed to make decisions is simply that I imagine, one at a time, the possible outcomes of a particular decision. If the outcome meets the test, which may be as simple 'as meeting some aspiration level.' (Evans 2002:502) or simply being too much bother, then the search is terminated.

The search hypothesis of emotion is not a claim about what emotions are, but what they do. Further, as noted by Evans, it is vacuous claim unless we have some independent idea of what emotions are, against which it can be evaluated, but first, some possible objections. The recourse to the search problem at all seems to suggest some commitment to a computational theory of mind; however, as a claim about what emotions do, rather than what they are, it is not necessary to assume that emotions involve 'rule-governed transformations of syntactic representations,' (Evans 2002:507). The search hypothesis does not then stand or fall with computationalism or the Classical Approach. What it does say is that emotions can be used in ways that aid cognition and decision making,

---

[21] This problem has been solved through the use of stochastic cut functions in many classical expert systems e.g. logic theorem provers (Don Ross: personal communication). I think the problem Evans is trying to highlight, is that the vast range of possibilities, considered aside from emotion, may lead humans either to confusion, or the inability to make a decision.

rather than hindering it. Rather than fixing ends amongst possible options, emotions can cut down the search space so as to reduce computational load. When ends are not even considered, emotions can begin to be seen as a positive way of stating the possible outcomes of a search space. Evans is assuming the confrontation of agent with a task, rather than decision-making amongst agents. And so we begin to examine ways in which emotions may play a part in the interaction amongst agents which, rather than only narrowing the search space of a potential decision tree, is used to highlight salient points of interaction that will in turn create expectations of the participating agent.

5.3 Emotions as Strategic Signals

Ross and Dumouchel (forthcoming) take up a complementary point of view to that of Evans (2002), in criticising "'Hume's chasm' in contemporary accounts of emotion" (Ross and Dumouchel forthcoming: footnote 16); arguing that reason need not be a "slave of the passions". The 'negative view of emotions' (Evans 2001) has been undermined in recent years by Frank's (1988) book *Passions within Reason*[22]. Ross and Dumouchel's paper is an attempt to refine Frank's thesis, claiming that emotions rather than being a disturbance to reason, are an aid in avoiding mutually uncongenial situations *among rational agents*. Here, in the context of arguments about how economic game theory can be applied to emotions, Ross and Dumouchel ask 'What *interactive* purposes do emotional signalling systems evolve and stabilize to serve?' (Ross and Dumouchel forthcoming: 19 emphasis added.) Ross and Dumouchel argue that emotions are used as a tool for negotiation between agents to signal unacceptable outcomes (e.g. prisoners dilemmas) in a game-theoretic tree, and hence to try and avoid them.

Ross and Dumouchel analyse 'agents using emotions to avoid getting into games that are dilemmas.' (Ross and Dumouchel forthcoming: 7) This formulation is in distinction to Frank's formulation of 'agents using emotions and other commitment devices to *get out of* social dilemmas' (Ross and Dumouchel forthcoming: 7) The burden of the reformulation rests on the fact that strategic signals can be used to encode preference intensities, which are difficult to infer, rather than preference

---

[22] The first statement of the thesis, 'which used a combination of evolutionary psychology and Smith's moral philosophy to endogenise emotional influences with the model of the rational economic agent', was made by Hirshleifer (1987), although Ross and Dumouchel address their criticisms and revisions to Frank (1988; 2003) in particular.

orderings that are more easily inferred. Furthermore, unlike standard commitment devices, strategic signals need not be explicitly 'constructed in advance of strategic interaction.' (Spurrett and Cowley forthcoming: 15) The use of emotion allows agents to credibly signal preference intensity. This credible signal can then in turn be used to (sometimes) avoid games that are mutually unbeneficial to the agents concerned, not by getting out of games like prisoners dilemmas, but to avoid these intractable situations altogether.

Ross and Dumouchel (forthcoming) introduce the distinction between *control-system mechanisms* and *conventionalised signalling systems*. Ross and Dumouchel suggest that there is no reliable way to distinguish these through the use of a folk-psychological term. The distinction is useful in sorting out, for the purposes of game theory, games and signalling happening amongst a group of people, and legible both to the party giving the signal, the interlocutor and other people within the group, from neuropsychological and bodily responses to events. Accordingly, Ross and Dumouchel argue that 'the culturally variable, conventional symbols by which groups of people sort out their behavioural patterns into emotional types are fundamentally strategic' (Ross and Dumouchel forthcoming: 10). Further,

> 'Emotions in the sense of neuropsychological responses... are strategic responses only occasionally and accidentally, whereas conventionalised emotional signalling systems are essentially strategic systems, in the sense that but for this function they wouldn't exist.' (Ross and Dumouchel forthcoming: 11)

Emotion, as understood in Ross and Dumouchel (forthcoming), simply is the conventionalised signalling that happens amongst groups of people, and does not refer to more automatic responses of neuropsychological states, that may be an appropriate agent response to an event (whether it involve other agents or not).

Emotions understood as conventionalised signalling systems are by definition public, arising amongst groups of agents. This goes some way to explaining the 'enormous variation on the input side, that is, with respect to their causal aetiologies as regards mechanisms, stimuli, and degrees of cognitive penetration of emotional responses, relative to the degree of output variability, that is, cross-cultural disagreement in typing

and understanding of emotional responses' (Ross and Dumouchel forthcoming: 18). This type of evidence is used by Ross and Dumouchel (forthcoming) to motivate the distinction between control-system mechanisms and conventionalised signalling systems, a distinction that Frank (1988) fails to make.

Thus, guided by some of the considerations explored in Clark (1997) Ross and Dumouchel's use of the term emotion is not based in neuropsychological response, and therefore need not depend on physical processes confined to the biological body of the agent. Thus, emotion, in reference to conventional signalling derives its 'meaning' not from physical processes of an agents body, but is at least in part 'a function of wide environmental, socially governed, [and] contextual factors that are exploited by decoders.' (Ross and Dumouchel forthcoming: 14) On this understanding, the fact that signalling is public means that an 'individual "doing the meaning" relies on these socially governed factors, in just the same way as here interlocutors do, to understand herself.' (Ross and Dumouchel forthcoming: 14) Emotional signalling fits well with a distributed view of cognition, that posits causally complex aetiologies of events, not confined to the internal physical processes of an agent's body.

One important point is that 'According to Dumouchel (1999), emotional signalling is produced by systems of expression which are continuous, in which those events to which we give names of particular emotions are salient moments in uninterrupted processes of affective expression' (Ross and Dumouchel forthcoming: 18). For example, in the simpler case of Kismet, even though different affective states can be labelled at different times, for instance Kismet may grow angry, this does not show the presence of a digital system, but rather an analogue system, that is continuously active, but might for the purposes of signalling, be labelled by agents.

Emotional signalling makes use of affective states to establish expectations. 'Through such things as bodily posture, muscle tone, pitch of voice and facial expression we *negotiate* reciprocal intentions into tolerably stable sets of expectations *within which* our base level games are well-defined' (Ross and Dumouchel forthcoming: 21, italics in original). Emotions are used as signals to establish a ground in which the relationship (conceived here as a game amongst agents) is to be played. The importance of these

emotional signals is that they do not typically function to exchange information about objectives or goals, but *dynamically influence and determine* each others intentions through exchanges of affective expression.' (Ross and Dumouchel, forthcoming: 21, italics in original.) We do not communicate about the objectives of interaction, but through the exchange of information about preference cardinalities, communicate the base-level meta-games in which subsequent games will be played.

Crucially, the reason that we have to do this, is that prior to any interaction, in the form of negotiation, 'there is often no fact of the matter as to which game we're playing. This process of affective exchange usually rapidly evolves towards some (more or less) fixed point of coordination that 'frames' or 'tropes' our relationship: distaste, pleasure, love fear, anger, confidence, disappointment, etc. – the Smithian sentiments.' (Ross and Dumouchel forthcoming: 21) Rather than emotions being used to narrow a potentially infinite search space, amongst agents, emotion is used to negotiate and coordinate a space for that interaction. The negotiation sets a base-line sentiment for the relationship. Importantly this is not information that can be decided on prior to interaction, but is a result of coordination.

Emotional signalling then is not primarily aimed at specific goals, and in the absence of these goals, the signals will be epistemic and strategic actions aimed at creating expectations about the goals to be pursued in this and successive interactions.

> ' 'Coordination,' here should not be taken as referring only to the focal points of well-defined conventions in particular games. Rather, affective coordination underlies all types of social interactions. People not only play repeated games with each other, but interact with (and expect to interact with) the same individuals across ranges of different games over time. [...] More generally, it leads to the fact that in a social context what has to be taken into account is not only the value of the objective pursued but also the value of the relationship itself for the organisms involved' (Ross and Dumouchel forthcoming: 20-21).

The creation of expectation is important in so far it serves to bring about coordination amongst agents, as well as to maintain it. 'The important point is that public emotional-state categorization creates expectations in others, including third-party observers of

negotiations, and the very existence of these emotions will then tend to constrain agents' behaviour' (Ross and Dumouchel forthcoming: 22-23). Emotional, strategic and epistemic signalling, when it is converged upon by agents in interaction, is not concerned only with the value of a goal, but also with the value of the relationship itself.

> 'Our basic claim, then, is that conventions governing emotional expression can constrain interactive behaviour by creating expectations to which agents hold one another responsible in systems of self-enforcing equilibria. They can play this role equally well not only at the ground level of face to face interactions, but also in more abstract and temporally extended bargaining contexts where agents never actually meet' (Ross and Dumouchel forthcoming: 24).

Ross and Dumouchel provide ways of understanding emotion as active signalling amongst agents that has epistemic benefit in that it uncovers information about the relationships involved and in so doing changes the cognitive state of the agent, both by the very act of signalling and by the response of interlocutors.

5.4 Emotions as Epistemic Action

In any given case, what is the goal of emotional signalling and how can it be compared to epistemic action? The aim of emotional signalling is coordination, where sentiments are coordinated to avoid games that are mutually unbeneficial to the agents involved. The purpose of signalling includes the value and maintenance of a relationship. Emotional signalling, understood as socially and publicly conventionalized, places constraints and creates expectations for future behaviour. In comparison to the examples of epistemic actions so far examined, all of which have been in a very short time domain, signalling systems can be applied both to longer term strategic negotiations as well as to short-term face-to-face encounters. They are like epistemic actions in that they do not need to be explicitly planned, neither need the agents involved fix on emotional-state labels for the purposes of signalling. Further they are not digital mechanisms that can be switched on and off, but continuously present in interaction. Emotional signalling is particularly interesting in that it does not primarily concern pragmatic goals or objectives, but signals the best way of going about attaining those goals, or avoiding paths that will not lead to the

satisfaction of those objectives. Hence, they are like epistemic actions in that they provide information that is not easily known or discovered, and by acting, preference intensities are signalled to other agents in interaction and third parties to the interaction. Emotional signalling, then, is aimed, epistemically, at being able to continue interaction, by avoiding games that may end the interaction.

Ross and Dumouchel's (forthcoming) account can be equally applied to both short and long term interaction. However given the emphasis of specific actions on the interlocutors present, I will be dealing primarily with short-term interactions. Thus while specific policies or objectives could potentially be interpreted in the final sections in an examination of real-time conversation (section 7.2 and 7.3), I will not be taking up this speculative account. Turning now to views of language in particular, I will be examining Andy Clark's (1997, 1998) account of the cognitive benefits of language.

6. Language and Artefacts

So far we have taken two opposing views to cognition, the Classical Approach and the Activity-based Approach and explored ways in which actions can be used to solve cognitive problems externally in the micro-time domain, (i.e. faster than a fully representational system.) Along with external actions in the environment, I have tried to show how emotions can be used as epistemic actions to lighten cognitive load by constraining the search space and possible responses of subjects in interaction, in the service of higher level policies (concerned ultimately with pragmatic goals or objectives). What I propose to do now is to lay out some of the views of language from cognitive science in order to try and show where dialogue, taken as an extension of activity, might be made to fit in: in particular, using insights gained from the activity driven approaches above to clarify some of the functions of dialogue. As a proponent of the activity driven approach, I will be examining Andy Clark's (1997, 1998) view of language as the ultimate artefact.

6. 1 Language in Cognitive Science

The standard view of language in Cognitive Science is in some shape or form strongly associated with internal representation. Following Chomsky (1965) and Fodor (1975, 1987) language has been seen as a decontextualised string of code, composed of units with syntactic properties. Here language is identified with the 'abstraction amenable' (Spurrett and Cowley forthcoming: 1) elements of language[23]. Language so understood leads to a view of internal representation that seems to exclude important features of language and interaction given the contextual constraints of embodied and situated agents. Further the stringent conditions imposed on language by the requirement of decontextualised units makes it difficult for them to be plausible options for an active approach to Cognitive Science. Against these standard views, language as an object of study has been identified with transcription of speech

---

[23] See Spurrett and Cowley (forthcoming) for a brief response to two major objections to dropping the 'abstraction amenable' approach to language: namely, the power of a sophisticated digital model and the view that language is arbitrary. In contrast, the emphasis is turned to utterance activity, a term of art used to refer 'to the full range of kinetic and prosodic features of the on-line behaviour of interacting humans.' (Spurrett and Cowley, forthcoming: 1)

(Garrod and Pickering 2003, section 7.1) where the usual linguistic categories[24] still apply. In contrast, Clark (1997, 1998) follows a more activity driven perspective, stressing the external nature of language.

By now it should be clearer that cognitive processing is not circumscribed either by language, or dependent on language in the form of representation (construed as a classical string of syntactically recombinable and context free units.) Brooks' various activity decomposed systems do not require any representations to operate robustly in particular task domains. One of the ways in which we can start to bring language into the picture is to treat language as an artefact, that, similar to pen and paper methods for performing long divisions, augments human cognition in various ways.

Andy Clark in 'Magic Words: how language augments human computation' (1998) seeks to extend the idea of human cognition being augmented by the tools and structures that we find in the environment. The idea is that like other things that we use to lighten computational load, it is a 'public ... tool ... a species of external artefact' (Clark 1997: 162) that allows us to restructure computational space in order to solve problems. He is not denying that language is 'an instrument of interpersonal communication.' or that it plays a role in 'processes of information transfer' (Clark 1998: 162) but seeks to point out some of the other computational benefits that it may have.

Clark's main claim is that language serves purposes other than communication. This view falls under the rubric of what he calls 'supra-communicative views of language' (1998: 163) What this posits is that language is more than a communication device for information transfer. These are not new claims and follow precedents set by Vygotosky (1986) and followed up in developmental studies by Diaz and Berk (1992) and also in philosophical work by Carruthers (1996). Clark's view is basically 'the idea of language as a computational transformer which allows pattern-completing brains to tackle otherwise intractable classes of cognitive problems' (1998: 163). Focussing on the external aspects of language, Clark argues for the computational benefits of language that are primarily external, rather than internalisations of text-like language.

---

[24] Garrod and Pickering (2003) use the following representational categories, phonetic, phonological, lexical, syntactic, semantic as parts of their model of dialogue.

6.2 Cognitive Benefits of Language

Clark's vision is that language is largely the 'ultimate artefact' (1997), and because of this emphasis on externalisation, relies quite heavily on the idea of scaffolding. 'Scaffolding denotes a broad class of physical, cognitive and social augmentations - augmentations which allow us to achieve some goal which would otherwise be beyond us.' (1998:163) This is taken from research by Vygotsky (1986) and Berk and Garvin (1984) into the role of private speech in children. In observations and recordings of groups of children aged 5-10, Berk and Garvin 'found that most of the children's private speech (speech not addressed to some other listener) seems keyed to the direction and control of the child's action. Subsequent study (Bivens and Berk 1990; Berk 1994) showed (a) that private speech increased amongst children trying to perform a difficult or complicated task whilst alone and (b) the higher the incidence of private speech or self-directed talk, the better the task was mastered. The emphasis here is that language is a tool for structuring and controlling action. The mind is not merely a manipulator but a controller. Language is not manipulated by the brain but used by the brain to control actions. Language is not merely a vehicle or medium of information transfer. This leads to the intra-individual view of language not as a representation or expression, but as a tool for effecting change in one's environment. What then is special about natural language is that it has an effect on an individual precisely because it is not in a hidden or more fundamental code. In line with scaffolding language can be seen as a sort of control loop or action loop which, amongst other things, reminds us what to do, controls what we are doing and provides a bench-mark against which to measure performance.

This leads to two related but distinct views. Carruthers sees language as a special kind of thought. Dennett offers a second version which depicts 'linguiform inputs as having distinctive effects on some inner computational device'. Thus one view is that language enables a sort 'public thinking'. This is not however tied to a particular language, and the brain is not being re-programmed, but that certain types of thinking are undertaken in public language and thereby enabled. Dennett's view is put by Clark to be a suggestion that 'conscious human minds are more or less serial virtual machines implemented-inefficiently- on the parallel hardware that evolution provided for us.' (Clark 1998: 166, Dennett 1991: 278) The difference here between Clark and Dennett is that Dennett seems to see language as the 'literal installation of a new kind of computational device

inside the brain." While it is not clear to what extent Dennett and Clark disagree, Clark (at least) sees language as primarily external and a complement to representation and computation. Clark contends that language 'use(s) the same old (essentially pattern completing) resources to model the special kinds of behaviour observed in the public linguistic world.' (1998: 168)

Clark looks at language as a transformer. Thus, it is part of material culture caught up in - but not constituting - cognitive looping. Given his concern with challenging cognitive internalism, it seems natural to highlight the artefactual nature of much language. However, while endorsing this part of his thinking, I wish to shift the emphasis away from so-called higher cognitive processes. For Clark (1997: 202; 1998: 169-173), the following are the important resources made possible by language.

- Memory augmentation (use of diaries, libraries etc.)
- Environmental simplification (e.g. use of road signs)
- Co-ordination and reduction of on-line deliberation (use of language in (joint) planning)
- Taming of path dependent learning (language allows previous learning to cross between agents and bear on 'unrelated' future events)
- Attention and resource allocation (given these resources, what do we prioritise?)
- Data manipulation and representation (especially in working with text)

Given the 'reasoning bias' (higher cognitive function) of the list, it is not surprising that, like Vygotsky (1986), Clark emphasises the self-directed speech that allows vocalizations to re-organize cognitive space. The re-organization of cognitive space allows children and adults to better direct and plan subsequent actions, allowing them to accomplish cognitive tasks that would otherwise be too difficult or time-consuming. Written and spoken language is used, at least in these cases, to augment the natural capacities of the brain, primarily by being manipulated externally.

6.3 Applications to Language in Dialogue

We started out by looking at some of the benefits which can accrue to cognitive systems that actively used the constraints of embodiment and situatedness to solve cognitive problems. These systems do not need internal representational or central control capacities as understood by a Classical Approach. Expanding on an active approach to cognition, an examination of human player's strategies for Tetris reveal a function of action, epistemic action, aimed at, through the deployment of action, discovering information that is either difficult to uncover or compute and thereby aimed not at an immediate pragmatic goal (completing the task) but changing the cognitive state of the agent. It should be clear by now that language in its symbolic form can have cognitive benefits other than information transfer in the service of pragmatic goals. However even in these views, we have been looking at language in ways that assume the appearance of word like forms, and is for the most part dependent on the manipulation of actual text and signs. What we need to begin to do is look at ways in which words can be used in an on-line immediate fashion, looking for evidence of what could be called epistemic and strategic action that happens not as manipulation of text or speech in the pursuit of a task, or the solving of a problem, but as interaction between people and the sounds that they make. For the Classical Approach excludes the kinds of emotional effects that language use can have in interaction and Clark's brand of externalism, whilst acceptable, does not address the effects that the speaking of words may have for ourselves as well as other agents in interaction For in on-line behaviour it is probable, that rather than relying on the appearance of the abstraction amenable elements of language, word-like forms and syntax, spoken language may have functions other than information transfer, namely the discovery of sufficient information about the relation of the agent to either a task, (assuming indeterminate and dynamic results of actions) or, more importantly, other agents in order to pursue further tasks as well as continue (or have the possibility of continuing) the interaction in which an agent is currently involved.

7. Distributed Cognition in Dialogue

Now that we have looked at some of the ways emotions can aid agents in approaching situations where the potential interaction space is not bounded, nor information complete, we will begin to try and apply some of these comments to dialogue as a case in which both action, context or situation and words are all central to the function of the event. I will examine Pickering and Garrod's (2003) mechanistic view of dialogue, which contends that the alignment (according to representational categories) in dialogue is automatic. Then, focussing on the phenomena of acoustics and actions in speech, I will try to show that sounds, apart from words, can have strategic and cognitive effects on agents in interaction. Alignment then does not depend on common ground based on representation, but on current and shared histories of joint activity. In contrast I will look at some of Cowley's data (1998) to show that what is important in dialogue and responded to by the persons involved, is not alignment of representations, but strategic and epistemic actions aimed at alignment and coordination amongst interlocutors.

7.1 A Mechanistic Psychology of Dialogue

As in the example of Kismet, social interactions can be regulated by cognitive looping and scaffolding that has no need of representations of abstraction amenable elements of language. Using architecture inspired by Rodney Brooks, Breazeal has been able to build a robot that is able to loop flexibly with people in interaction without direct planning. The actions and (emotions) that the robot exhibits are used to influence the subject in the interaction into fulfilling the current state (drives) of the robot. This provides some argument for more basic and autonomous mechanisms underlying human speech. Pickering and Garrod (2003) provide evidence and argument for alignment at the level of linguistic representation.

The following transcript is taken from a conversation between two players in a co-operative maze game (Garrod and Anderson, 1987).

'In this extract one player A is trying to describe his position to his partner B who is viewing the same maze on a computer screen in another room.

Example dialogue taken from Garrod and Anderson (1987)

1. **B**: ...Tell me where you are?

2. **A**: Ehm : Oh God (<u>laughs</u>)

3. **B**: (<u>laughs</u>)

4. **A**: Right : **two along from the bottom one up**:

5. **B**: Two along from the bottom, which side?

6. **A**: The left : going from left to right in the second box.

7. **B**: You're in the second box.

8. **A**: One up :(<u>1 sec.</u>) I take it we've got identical mazes?

9. **B**: yeah well : right, starting from the left, **you're one along**:

10. **A**: Uh-huh:

11. **B**: **and one up?**

12. **A**: Yeah and I'm trying to get to.....

[28 utterances later]

41. **B**: You are starting from the left, **you're one along, one up**? (<u>2 sec.</u>)

42. **A**: **Two along** : I'm not in the first box, I'm in the second box:

43. **B**: You're **two along**:

44. **A**: **Two up** (<u>1 sec.</u>) counting the : if you take : the first box as being one up :

45. **B**: (<u>2 sec.</u>) Uh-huh :

46. **A**: Well : I'm **two along two up**: (<u>1.5 sec.</u>)

47. **B**: Two up? :

48. **A**: Yeah (<u>1 sec.</u>) so I can move down one:

49: **B**: Yeah I see where you are: (Pickering and Garrod 2003:4)

Now it looks like there are very few grammatically well-formed sentences. This is taken as important evidence against an undue focus on complete and grammatical sentences that grows out of theories associated with Chomsky (1965) and Fodor (1975). Further there are differences between one speakers description of the same position (A: (4) "two along from the bottom one up" is quite different from (46) "two along, two up"). Garrod and Pickering point out that it can be seen as quite orderly if

'we assume that dialogue is a joint activity (Clark, H. 1996).' that 'involves cooperation between interlocutors in a way that allows them to sufficiently understand the meaning of the dialogue as a whole: and this meaning results from joint processes.' (2003: 5).

Part of the importance of looking at dialogue is that turn-taking is not hard and fast and further that many utterances can only make sense across a number of 'turns'. e.g. statements like (4) "Right, two along from the bottom two up" 'requires an affirmation or a query'. (2003: 5) The point is that a co-ordinated activity such as conversation cannot be broken down into neat turns. For the purposes of their paper, Pickering and Garrod define coordination as joint activity (like dancing or boxing) and alignment as when interlocutors share the same representation at a particular level. '[A]lignment occurs at a particular level when interlocutors have the same representation at that level. Dialogue is a coordinated behaviour (just like ballroom dancing). However, the linguistic representations that underlie coordinated dialogue come to be aligned.' (2003: 6)

Now it is posited that 'alignment of situation models is central to successful dialogue.' where a situation model is 'multi-dimensional representation of the situation under discussion.' (2003: 6) including space, time, causality, intentionality and reference to main individuals under discussion. Now it might be that the interlocutors represent situation models differently, but it is assumed that it is inefficient to maintain two models at once. Further if the situation models are in fact aligned, then it is unnecessary to model the representations of the interlocutor. In higher level arguments, of course the interlocutors need not align totally. They do however have to still align in terms of reference to particular parties or entities, whereas they may disagree about the value of these particular entities or parties.

Given that explicit achievement of alignment (e.g. agreeing on definitions) is unusual Pickering & Garrod propose that '"global" alignment of models seems to result from "local" alignment at the level of the linguistic representations being used.' (2003: 7) In achieving alignment of situation models, Pickering and Garrod 'propose that this works via a priming mechanism, whereby encountering an utterance that uses that representation makes it more likely that the person will subsequently produce an

utterance that uses that representation' (2003: 7). The processes involved are then much more automatic and multi-modal and less explicit than we would assume for a task that appears to be quite cognitively demanding. Neither are they representationally demanding in terms of a situation model, but rather rely on basic coordination of speech.

For instance in the dialogue presented at the beginning of this section, A and B assume slightly different interpretations of 'two along'. A is interpreting two along by counting the boxes in the maze, whereas B is interpreting 'two along' as the connections between the boxes. Thus in relation to A, 'two along' actually means 'three along'. The misalignment is recognised due to the fact that 'two along' does not easily fit into a current representation. The way in which it is repaired is to iterate or reformulate the utterance. For instance in (7) the formulation is simply repeated with a rising intonation, or a simple clarificatory request is made, as in (5) B 'Two along from the bottom, which side?'. Sometimes more radical reformulations are necessary as when 'Two along' is reformulated as 'second box' (6). These requests, sometimes in the form of requests and iterations, sometimes in direct reformulations 'reflect failures to understand what the speaker is saying in relation to the listener's model' (2003:9) The clarificatory moves are an attempt to make the interlocutor find a linguistic representation that is more suitable to the listener's model. For example B says 'you're one along, one up' (41) which B immediately reformulates as 'Two along' (42) This returns a clarification request of 'You're two along' until by utterance (44) the representations come to be aligned, evidenced by the fact that the interlocutors can now complete each others utterances.

The use of a transcript leads Garrod and Pickering to a heavy reliance on some notion of representational categories. To their credit though, they do not assume the decoding and encoding of representations through words like forms. Rather the representations are legible between parties without the necessity of encoding, and as such are not in a hidden or more fundamental code. The interesting case and argument that they make is that these alignments are more or less automatic. If these sorts of alignments are automatic, then they begin to look like the skills that we have in such pragmatically motivated domains as typing.

Now while it is granted that this particular experimental set-up is required for control purposes and is able to display some of the interesting alignments that happen in dialogue, this set-up is evidently not usual in most conversation[25]. The experimental subjects are excluded from a social situation and have been provided with a cognitive task that has definite criteria for what count as success or failure, (the ability to navigate jointly around a maze from different perspectives.) In contrast, most conversation is undertaken in face-to-face situations that for the most part involve social aspects of interaction. Thus, while the experimental set-up produces the reported results, it also excludes the possibility of social interaction where less definite goals may give rise to more strategic and epistemic interaction.

I propose now to look at a natural conversation where the situation is given, (in virtue of the fact that interlocutors are in the same place). Then we can assume that all these automatic processes are in operation. This has two consequences. First if the situation is given then the relation to the situation becomes more important (and that includes to people in the same situation). Secondly, differences in tone and prosodics, associated with the emotions that they convey, held against a situation that is the same for all interlocutors, tend to be more prominent.

7.2 Pieces of Italian Dialogue

I now turn to an example of dialogue that has what I consider to be several paradigmatic aspects. First the parties to the conversation are all present. Secondly it occurs in a domestic setting. Thirdly it is more interesting for our purposes since it involves what Cowley has called 'a bit of nagging'[26]. This means that the parties are well known to each other. What I propose to do is show the ways in which utterances can have emotional effects and how these emotional effects can be shown in phonetic transcripts of the conversation. The phonetic transcripts will, I argue, provide evidence of the ways in which people manage the relationships around them.[27] In so

---

[25] Garrod and Pickering do not deny more complex functions of conversation, e.g. the provocation of emotion, but maintain that these alignment mechanisms are primitive and underlie conversation more generally.
[26] Personal Communication
[27] To avoid any process view, the synchronic perspective is dropped for a focus on particular events. Drawing on various traditions, Cowley (1998) looks at talk with respect to voice dynamics (see, Abercrombie, 1967) and temporally embedded context (see, Kendon, 1990).

doing we see how in utterance activity, two or more brains contextualise activity that results not in internalised representations but public coordination. Drawing on epistemic actions, I will try to show that aspects of talk - both word-based and phonetic - are dedicated to the perception and discovery of information that, now, is relevant to the agent.

Let us consider events from an family conversation taking place in Italy (for fuller description, see Cowley 1993; 1998). The participants involved are the mother, Rosa, the father, Aldo, and an adult daughter, Monica. While eating an evening meal, talk turns to why a husband, Aldo (A) failed to cut his wife, Rosa (R), the pea-poles she wanted (it is several months into pea season), or, paraphrasing, that 'a certain person is too lazy to cut pea-poles'. In raising this, Rosa (R) is heard as complaining - in her husband's hearing - to their daughter Monica (M). Strategically, she is hoping to gain her daughter's sympathy by complaining of her husbands inadequacies. As it happens it soon comes to light that Aldo has in fact cut the pea-poles that she had requested. Not surprisingly, this comes nowhere near satisfying Rosa. Quite the contrary. Rather, while acknowledging her mistake, she changes tack:

English version   Italian original

(10) R: Too right, they were you should have   R: Affatti se vedessi le bacchette ah son piu'
seen the poles oh they're longer than this   lunghe di questa camera se non piu'.
room if not longer.
(11) A: Come o::n   A: Oeu
(12) M: Come o::n   M: Oeu
(13) R: (if) not longer   R: Non piu'
(14) M: No   M: Va

                                                        (Cowley 1998:548)

Rosa attempts once more to gain her daughters' sympathy by complaining that the poles, even though they had been cut, are far too long, and thus unsatisfactory. She is effectively saying that the poles are longer than the four metre room. Now replying to the assertion that the pea-poles were as long as the four metre room, would be missing the strategic significance of the utterance. Further, aligning according to linguistic categories will not solve a problem, for this is not what is at stake. Rather, words' failing

Aldo, the utterance evokes what Goffman (1981) calls a 'response cry'. The sound made is 'amenable to being glossed as "come on, you must be joking", and in the context is clearly legible as an action of gentle mocking' (Spurrett and Cowley forthcoming:13). What is transcribed as 'oeu' (and translated as 'come o::n') represents a non-standard sound that cannot be reduced to words, but is a 'vocal gesture... the duration of which can be stretched to that of a short sentence.' (Spurrett and Cowley forthcoming:14) As Spurrett and Cowley (forthcoming) comment, 'What is most striking, though, is not the internal prosodic properties of Aldo's "Oeu" but its relational properties in the context of the interaction, and the shared history of the three people present' (Spurrett and Cowley forthcoming: 14).

The talk exemplifies a social payoff that arises in managing family members. On this occasion, Rosa gets what husband and daughter regard as her due reward: they ridicule her. In so doing, they find that they share each other's attitudes and feel they belong. They get satisfaction from being 'on the same wavelength'. None of this, however, arises in what is said: rather, it depends on vocal (and visible) expression lacking any syntactic or semantic basis. Or that saying 'oeu' is a practice which, among other things, enables one to label something (or someone) as absurd. Even if no such word appears in dictionaries, the activity is constrained by, at times, saying 'oeu' while looking and gesturing in specific (Italian) ways. Cognitive action spreads across the environment even if, necessarily, it loops between individuals.

Rosa's initial utterances are strategic, aimed at getting Monica to coordinate emotionally with her, and against her husband. When the initial utterance fails to gain the desired effect, her next utterance tries again, on a different tack, to produce the desired effect. Importantly, what is being said (the words used) are not critical to the possible strategic effects of the action, and unlike Tetris, do not have determinate results to possible actions. Rather Rosa's utterances are epistemic action loops, that include other persons, motivated by a need for information (is Monica with Rosa against Aldo?) that will also serve to change cognitive states, Monica's and in turn Rosa's, by producing coordination. Her utterances, as strategic actions, create expectations amongst other persons, which as we see, may be disappointed. The lack of determinate response, as evidenced by Aldo and Monica's response to Rosa's second attempt, demands greater

flexibility from the agents involved in tracking the different vocal gestures and the emotional signals that they contain.

The moves in this family game reveal several aspects of epistemic action. Rosa saying that her husband cut four metre pea-poles is designed neither as part of a story nor as a factual claim. Rather, it is action designed to probe attitudes and create expectations. Without a determined response to an action, the goal of the utterance is, through epistemic action, emotional coordination amongst agents (and against others). Aldo instead of hearing what Rosa wants, rightly responds with what is transcribed as "oeu". Although stating nothing, saying 'oeu' alters each person's state of mind. Apart from anything else, as shown below, it prods the daughter into subtle response. Before examining microcognition in real-time, we can visibly see from the transcript (and following phonetic record) that talk can be irreducible to successive speech acts. Often what happens is simultaneously strategic, epistemic and perceptual action that primes and provokes further goings on.

## 7.3 Microcognitive Details of Conversation

Turning to how these abstract descriptions play out in microtemporal detail, we find, as in Tetris, that events depend on how the actions are modulated, attempting to show how 'three voices reaffirm family relationships' (Cowley 1998: 549). The temporal details are as follows: when Rosa prods her husband with the first utterance of 'non piu' (10) she speaks so that her voice falls to 220 Hz and, as it turns out, this influences his responding.

In the following, her 'non piu'' is represented iconically and measures are given for its acoustic peak and minimum fundamental frequency (the 'p' has no pitch).[28]

275

220

non   piu`

---
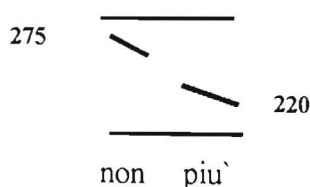
[28] All measures were made on a Kay sonagram. For details, see Cowley (1998).

58

The message and the dynamic features of her speaking prod Aldo into his response cry. Beginning between the 'non' and the 'piu'', instead of using his usual (male) pitch range, beginning at 180 Hz he raises his voice into a female domain, ending his 'Oeu' at the same frequency as Rosa's last utterance, 220 Hz (duration is about 760 msec.):

o:::: e:::::::: u:::::
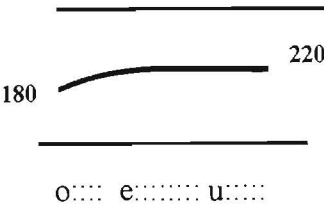
Given that Aldo starts speaking during the 'p' of 'piu'', the alignment cannot be planned. Rather, the overlap shows that some kind of perception-action mechanism[29] allows him to orient to the pitch of Rosa's voice. In producing 'oeu' his voice matches her final pitch level (to within about 4 Hz) and flattens out (220 Hz). In short, he latches on to or echoes the final note of his wife's utterance. Nor is this likely to be coincidence. Rather, it is likely that it is nonpragmatic action. The best evidence for this is found in his daughter's response. Approximately 300 milliseconds after he starts speaking, she not only comes in with a similar 'oeu' but, as the acoustic record shows, orients to the same target.

---

[29] The harmonisations present in this example are coordinated both according to the vocal production of the utterer as well as that of the other. This example of affective coordination, like strategic actions, 'involve integration of inputs from each participants's own behaviour and that of others.' (Spurrett and Cowley, forthcoming:15-16)

The voices are shown below,

Figure 1: O eu'
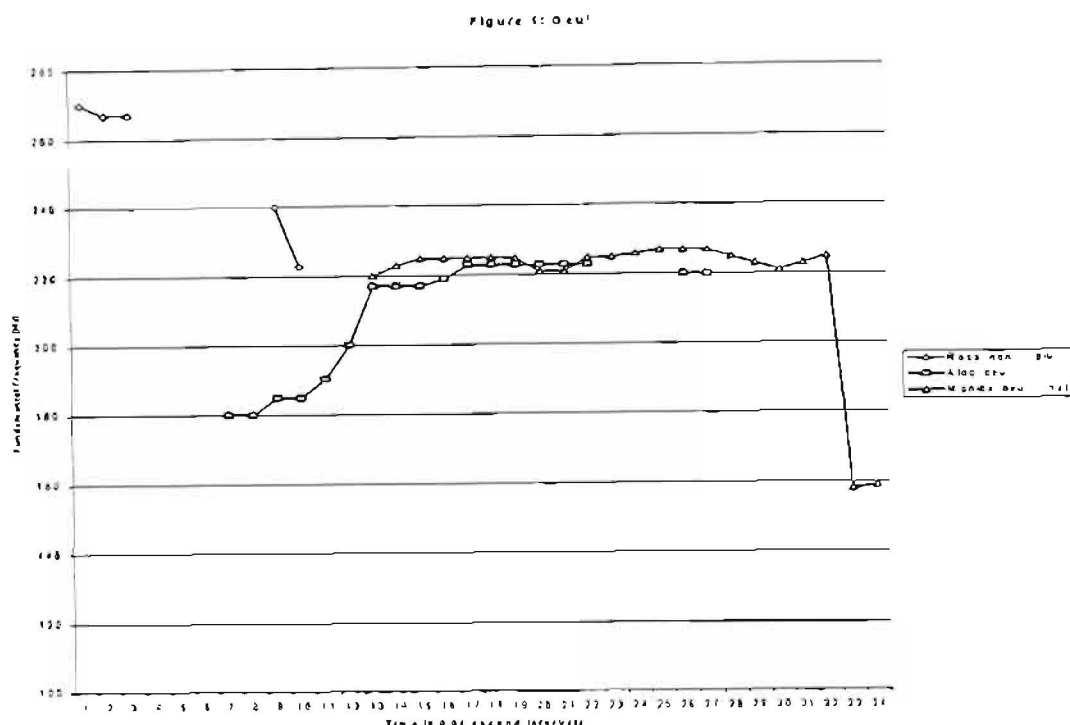
(Spurrett and Cowley forthcoming. 14)

Beginning at 227 Hz, Monica's voice matches Aldo's 0.16 seconds after she begins her 'Oeu'. Their voices, not merely converging, harmonise for a further +- 0.25 seconds, remaining on the same note as Rosa's last utterance. Finally, Aldo running out of breath allows the pitch of his voice to fall, while Monica continues for a further 0.2 seconds, at which point she 'signalled her enjoyment with a laugh pitched to the top of her father's range.' (Cowley 1998: 551)

From one perspective, this is Aldo and Monica 'ridiculing' Rosa good naturedly and in harmony. Saying 'oeu' as described, in these circumstances, is just *that*. 'At this moment, this is helped both by smiles heard in the concurrent oeus (smiles that are also seen) and, as is highly audible, in the harmony and simultaneity of the voices.' (Cowley 1998:549) They ridicule *Rosa* thanks to how they, so to speak, align their voices meaningfully in contrast to hers. The alignment produced is not representationally based, but understood immediately by all parties present, producing coordination in public space. The 'good nature' is physically-based harmony as well as Monica's little (father directed?) laugh. Far from relying on planning, this is spontaneous, public

activity spilling across persons. Such cases serve to make three points. First, the 200ms duration of Aldo and Monica's harmonisation allows the alignment or attunement to reach consciousness: given Monica's laugh, it seems likely that this occurs. Second, what happens is too fast and too responsive for central planning. Evidence of this occurs as Aldo's 'oeu' begins between Rosa's 'non' and 'piu''. Third, while the example is useful because it occurs on 'oeu', similar effects often exploit the words actually spoken. In utterance-activity, there are times when words are mere background to relationships that rely on the meshing of vocal and other expression and these meshings alleviate the burden for a full representational account either cognitively or linguistically.

The actions and attunements present here show epistemic and strategic qualities. Rosa's assertion is aimed both at producing and discovering whether Monica is in accord with her estimation of her husband's inadequacies as well as prodding Aldo into some kind of response. However, she is disappointed as Aldo's response 'indexing [Rosa's] utterance' (Spurrett and Cowley forthcoming: 14) and Monica's harmonisation and the laugh that indexes his normal range shows. "Aldo and Monica are identifiably 'together' because their utterances harmonise, showing a brief allegiance in the same way as bodily orientation shows acceptance or rejection." (Spurret and Cowley, forthcoming:14) Nor was this short event without effect, for Rosa, rather than hearing the good humour, shortly avowed, that the next time she was in the village, she would get a friend to confirm the length of the poles. (Cowley 1998: 566)

The sense of the conversation is not contained in the formal or abstraction amenable features of language or conversation, but in the physical quality of *how* what is said is vocalised. The absence of the acoustic record would not be able to expose the subtle moves of a family game as they do here. The sense of this piece of conversation exploits how we co-ordinate practices through alignment that rely on bio-mechanical constraints. In talk, how we go on is often irreducible to word-based patterns and is better understood in terms of epistemic action and coordinated activity.

7.4 General Implications

The lessons we learn from this example of pitch matching are two fold. First, meaningful content is ascribed to parts of speech, particular sounds and harmonisations, which of

necessity happen *between* people. These typically are actions in very short time domains, that exploiting either response cries or actual words, are public and external. The fact that they happen between people and are not even discernibly word-based (as in this, perhaps extreme, example) is enough, to show that it is unlikely that speech production and language is concerned only with the generation of abstraction amenable elements of language, as a representational account would have it. For here there is no, and cannot be, the kind of formal aspects of language that a representational account of cognition demands. Secondly these are not pragmatic actions, but emotional and epistemic ones. This is in part because, given the short response time, they cannot be explicitly planned. Further they cannot be planned to produce a pragmatic effect, aimed as they are at eliciting information from persons about the relationships concerned. We use these features of utterances to impart information and pick up on information that creates expectations amongst parties that guide and coordinate subsequent interaction. The phenomenon of pitch matching is part both of the circumstances and the relations of the parties involved and as such is massively distributed. In order to elicit meaning, language can at times lean heavily on a coordination of events in the micro-temporal domain superposed on an environment of both interpersonal relationships and cultural norms.

Far from being an isolated incident, 'Similar forms of indexing can be found by looking beyond pitch, and attending to the ways in which, *inter alia*, accent, timing and loudness play out in utterance activity' (Spurrett and Cowley, forthcoming: 15). Further these features of utterance activities are all but ubiquitous at all ages (Spurrett and Cowley, forthcoming: 15). However the effects of these kinds of phenomena are not always as striking as the example presented here, as they depend on a shared history of joint activity. The physical manipulations of voices are based on the ability to perceive the tones of the person with whom one is conversing. This episode is, however, 'evidence of the ways in which prosodic patterns between people with histories of shared intimacy are modulated by that history, as they can be by cultural experiences' (Spurrett and Cowley, forthcoming: 14). Prosodic actions in speech operate as epistemic actions when legible by other parties with the aim, not of achieving immediate pragmatic ends, but coordination that is characteristic of relationships.

8. Conclusion

In this paper I have argued, that from the perspective of Cognitive Science, activity driven approaches are better at real-world problem solving because their architectures do not demand categorisation based on representation. Active systems are able to exploit the dynamics of real-world environments, allowing them to directly influence the achievement of their goals. In the process, problem-solving becomes less abstract and turns toward complete working systems accomplishing tasks in the real world with no reliance on either specific coding of environments or computationally extensive (and expensive) algorithms. Following a reliance on action to solve problems, there is a category of action, epistemic action, whose primary function is not to achieve pragmatic goals based on plans generated from complete representations, but, through continuous recourse to the world, through action, to change the cognitive state of the agent in pursuit of those goals. This ability alleviates the need for complete representation, thus being computationally less expensive, enabling the system to achieve a much faster response time. This demands a cognitive model that features recourse to the world through action at any stage of problem-solving. In addition, these actions are deployed more, not less, by skilled agents, under the guidance of higher level and longer term goals, providing evidence of higher level cognitive control.

In turning to how epistemic action may feature in human interaction amongst agents, it is shown that, in the absence of complete abstract and representational information, agents use emotional cues to prompt a desired response, placing expectations on their interlocutors. In addition to constraining the interaction amongst agents, emotional cues are used to signal preference intensity amongst possible outcomes, enabling the continuation of subsequent interaction. In examining a piece of natural dialogue, we find the sorts of alignments of attitude aimed at coordination that allow the continuation of interaction, quite apart from more or less arbitrary, (from the perspective of activity) formal representations. Moreover, in the example examined, no possible representational account can be given, for the utterances show no formal representational qualities, and in the most useful case for my purposes, do not exploit recognisable words. In this example dialogue is supported and sustained not by the presence of symbols, or of grammatical or linguistic representations, but by the

actions of the parties concerned, in epistemically and strategically attempting to align and coordinate externally through physical vocalisations and gestures in public space; thus, dialogue, while making use of what might be called symbols, is actively supported by mechanisms and strategies other than possible process formalisations.

I have tried to show how cognitive systems might be able to proceed in the face of a more or less constant lack of abstraction amenable information by paying attention to epistemic actions that occur in the microsecond domain. Far from resulting in degenerate forms of language use, this tends to be more efficient. Language in conversation is not primarily a form of scaffolding or artefact manipulation. The goals of conversation may include simultaneously the epistemic goal of alignment, (the eliciting of information that I do not have) and strategic goals (creating expectation and producing coordination). Epistemic and strategic actions include the ways in which we manage relationships. Conversation remains the most basic case of interpersonal interaction and language use and begins to show some of the ways in which language in terms of its formal properties can be meshed with views emphasising the activity of language and the activity surrounding language.

Reference:

Abercrombie, D. (1967) *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Berk, L. (1994) 'Why children talk to themselves'. *Scientific American* 271, no. 5:78-83.

Berk, L. & Garvin, R. (1984) 'Development of private speech among low-income Appalachian children'. *Developmental Psychology* 20, No. 2:271-286.

Bivens and Berk (1990) 'A longitudinal study of the development of elementary school children's private speech'. *Merrill-Palmer Quarterly* 36, no. 4:443-463

Breazeal, C. & Scassellati (1998) 'Infant-like Social Interactions between a Robot and Human Caretaker'. to appear in *Special issue of Adaptive Behavior on Simulation Models of Social Agents*, guest editor Kerstin Dautenhahn.

http://www.ai.mit.edu./projects/kismet-new/AB98.ps

Brooks, R.A. (1986) 'A Robust Layered Control System for a Mobile Robot'. *IEEE Journal of Robotics and Automation*, Vol. 2, No. 1, March 1986, pp. 14–23; also MIT AI Memo 864, September 1985

http://www.ai.mit.edu/people/brooks/papers/AIM-864.pdf

Brooks, R.A. (1991a) 'Intelligence without representation'. *Artificial Intelligence*. 47:139-159.

http://www.ai.mit.edu/people/brooks/papers/representation.pdf

Brooks, R.A. (1991b) 'Intelligence Without Reason' AI Memo No. 1293, MIT. AI Lab. (In Proceedings of the 12th *International Joint Conference on Artificial Intelligence*. Morgan Kauffman.).

http://www.ai.mit.edu/people/brooks/papers/AIM-1293.pdf

Brooks, R.A. (1996) 'Prospects for Human Level Intelligence for Humanoid Robots'. http://www.ai.mit.edu/people/brooks/papers/prospects.pdf

Brooks, R.A., C. Breazeal (Ferrell), R. Irie, C. Kemp, M. Marjanovic, B. Scassellati and M. Williamson, (1998) 'Alternate Essences of Intelligence', *Evolutionary Robotics: From Intelligent Robots to Artificial Life (ER'98)*, Vol. II AAI Books, Ontario, Canada, 1998, pp.307–326, also In 'Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)', AAAI Press.

Carruthers, P. & Boucher, J. (1998) *Language and Thought: Interdisciplinary Themes*. Cambridge: Cambridge University Press.

Chomksy, N. (1965) *Aspects of a theory of syntax*. Cambridge MA: MIT Press

Clark, A. (1989) *Microcognition: Philosophy, Cogintive Science and Parallel Distributed Processing*. MIT Press.

Clark, A. (1993) *Associative Engines, Connectionism, Concepts and Representational Change*. MIT Press.

Clark, A. (1997) *Being There, Putting Brain, Body and World Together Again*. Bradford. MIT Press.

Clark. A (1998) 'Magic Words: how language augments human computation' in P. Carruthers & J. Boucher, eds., *Language and thought: Interdisciplinary themes*. Cambridge University Press. Pg 162-183

Clark. H. H. (1996) *Using language*. Cambridge: Cambridge University Press.

Cowley, S.J. (1993) *The place of prosody in Italian conversations*. Unpublished PhD dissertation, University Of Cambridge.

Cowley, S.J. (1998) 'Of Timing, Turn-Taking, and Conversations'. *Journal of Psycholinguistic Research*, Vol. 27, No.5

Crossman, E. R. (1959) 'A theory of the acquisition of speed skill'. *Ergonomics*, 2, 153-166.

Damasio, A. R. (1994) *Descartes' Error: Emotion, Reason and the Human Brain*. London: Papermac.

Dennnett, D. (1991) *Consciousness Explained*. Boston: Little Brown and Company.

Diaz, R. and Berk, L., eds. (1992) *Private Speech: from Social Interaction to Self regulation*. Erlbaum.

Dumouchel, P. (1999) *Emotions: essai sur le corps et le social*. Paris: Synthelabo.

Elman, J. (1991) 'Representation and structure in connectionist models'. In *Cognitive Models of Speech Processing*, ed. G. Altman. MIT Press.

Evans, D. (2001) *Emotion: the Science of Sentiment*, Oxford: Oxford University Press.

Evans, D. (2002) 'The Search Hypothesis of Emotion'. *British Journal of Philosophy and Science* 53:497-509

Fodor, J. (1975) *The Language of Thought*. Crowell.

Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Fodor, J. and Pylyshyn, Z. (1988) 'Connectionism and cognitive architecture'. *Cognition* 28:3-71.

Frank, R. H. (1988) *Passions within Reasons: the Strategic Role of Emotions*, New York: Norton.

Frank, R. (2003) *What Price the Moral High Ground?* Princeton: Princeton University Press.

Garrod, S. & Anderson, A. (1987) 'Saying what you mean in dialogue: A study in conceptual and semantic co-ordination'. *Cognition*, 27, 181-218.

Goffman, E. (1981) 'Response cries'. *Language*, 54: 787-815. Reprinted in E. Goffman (ed.) *Forms of Talk*. Oxford: Basil Blackwell

Hirshleifer J. (1987) 'On the Emotions as Gaurantors of Threats and Promises'. In J. Dupre, (Ed.) *The Latest on the Best*. Cambridge, MA: MIT Press, 307-326.

Kendon A. (1990) *Conducting Interaction*. Cambridge, England: Cambridge University Press.

Kirsh, D. (1991) 'Foundations of AI: the big issues'. *Artificial Intelligence* 47: 3-30.

Kirsh, D. (1995) 'The intelligent use of space'. *Artificial Intelligence* Vol. 73, Number 1-2, pp. 31-68.

Kirsh, D. (1999) 'Distributed Cognition, Coordination and Environment Design'. in *Proceedings of the European Cognitive Science Society*. pg 1-11

Kirsh, D. and Maglio, P. (1992a) 'Perceptive Actions in Tetris'. In R. Simmons (chair) *AAAI spring symposium on control of selective perception.* Stanford University. March, 1992.

http://adrenaline.ucsd.edu/Kirsh/Articles/perceptiveactions/perceptive.pdf

Kirsh, D., and P. Maglio, (1992b) 'Reaction and reflection in Tetris'. In J. Hendler (Ed.), *Artificial intelligence planning systems: Proceedings of the First Annual International Conference (AIPS92).* San Mateo, CA: Morgan Kaufman.

http://adrenaline.ucsd.edu/Kirsh/Articles/reaction/reaction.ps

Kirsh, D. and Maglio, P. (1994) 'On Distinguishing Epistemic from Pragmatic Action'. *Cognitive Science* 18: 513-549.

http://adrenaline.ucsd.edu/Kirsh/Articles/cogsijournal/cogsci_journal.pdf

Maglio, P. & Kirsh, D. (1996) 'Epistemic Action Increases With Skill'. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

http://www.almaden.ibm.com/cs/people/pmaglio/pubs/cogsci96.pdf

Maglio, P., Matlock, T., Raphaely, D., Chernicky, B., & Kirsh D. (1999) 'Interactive skill in Scrabble'. *In Proceedings of Twenty-first Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

Maglio, P., Matlock, T., Campbell C., Zhai, S., Smith, B. (2000) 'Gaze and Speech in Attentive User Interfaces'. In *Proceedings of the Conference on Multimodal Interfaces.* Springer-Verlag. http://www.almaden.ibm.com/cs/people/pmaglio/pubs/mmui5.pdf.html

Mataric, M. (1991) 'Navigating with a rat brain: a neurobiologically inspired model for spatial representation'. In *From Animals to Animats* I, ed. J.-A. Meyer and S Wilson. MIT Press.

Minsky, M. (1986) *The Society of Mind.* New York: Simon and Schuster.

Moray, N. (1986) 'Monitoring Behavior and Supervisory Control'. In *Handbook of Perception and Human Performance, Vol II: Cognitive Processes and Perfomance*, chapter 40. New York, NY: John Wiley and Sons.

Newell, A. and Rosenbloom, P. (1981) 'Mechanisms of skill acquisition and the law of practice'. In J. R. Anderson (Ed.), *Cognitive Skills and their acquisition.* Hillsdale, NJ: LEA.

Newell, A. & Simon, H. (1976) *Human Problem Solving.* Prentice-Hall.

Nilsson (1984) 'Shakey the robot', Nils J. Nilsson (ed.) SRI A.I. Center Technical Note No. 323, April, 1984.

Pickering, M. J. & Garrod, S. (2003) 'Towards a Mechanistic Psychology of Dialogue'. *Behavioral and Brain Sciences.* Cambridge University Press.

http://www. Bbsonline.org/Preprints/Garrod/Referees/

Ross, D. & Dumouchel, P. (forthcoming) 'Emotions as Strategic Signals'. to appear in *Rationality and Society.*

Rumelhart, D.E., Smolensky P., McClelland. J.L., and Hinton, G.E. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol 1 & Vol. 2:*, ed. J McClelland et al. MIT Press.

Seibel, R. (1963) 'Discrimination reaction time for a 1,023 alternative task'. *Journal of Experimental Psychology*, 66, 215-226.

Spurrett, D. and Cowley, D. (forthcoming) 'How to Do Things Without Words: Infants, utterance-activity and distributed cognition'. to appear in R. Menary (ed.) *The Extended Mind: The Very Idea*, Amsterdam: John Benjamins.

Vygotsky, L. S. (1986) *Thought and Language* (trans. 1962 edition). MIT Press.

Wheeler, M., and Clark A. (1999) 'Genic representation: Reconciling content and causal complexity'. *British Journal for the Philosophy of Science.* 50 (1): 103-135.

Wheeler, M. (2001) 'Two threats to representation'. *Synthese.* Vol. 129, no. 2.