A Perspective on Multi-Agent Systems

Iain D. Craig
Department of Computer Science
University of Warwick
Coventry CV4 7AL
UK EU

February 3, 1995

Abstract

People do not live in isolation from others: they live in societies. This paper is about Multi-Agent Systems and their relevance to cognitive science and to theories of social behaviour. We examine some of what we believe to be central issues in the theory of Multi-Agent Systems and relate them to issues in Cognitive Science. Our aims are to introduce Multi-Agent Systems and to show how they can be illuminating as modelling devices in the production of cognitive theories that relate the individual to social context.

To appear in: Proc. First Bulgarian Summerschool on Cognitive Science.

Acknowledgements

This paper summarises (and in some places expands on) a series of five lectures given at the First Bulgarian Summerschool on Cognitive Science, Sofia, Bulgaria, 12–24 September, 1994. I am grateful to Dr. Boicho Kokinov for his invitation to teach this course, and to the students who took the Multi-Agent Systems course for their stimulating questions. I am also grateful to EC TEMPUS project SJEP-7272 for financial support in connection with the Summerschool.

1 Introduction

People typically do not live in isolation from others. The solitary castaway or the prisoner in solitary confinement are comparatively rare. In any case, even when isolated, the castaway and the prisoner often return to more populated environments. The isolation that castaways and solitary prisoners experience is, in any case, not absolute: they are the products of a society.

Cognitive Science deals with the mental processes of subjects considered as individuals and in isolation from all other subjects. It constructs theories of an individual's mental processes in terms of a computational metaphor. Multi-Agent Systems, on the other hand, can be interpreted as an investigation of the behaviour of collections of individuals embedded in an external context. In Multi-Agent Systems, the relationship between an individual's cognitive processes and the social context within which they act form the central topic of investigation. For Cognitive Science in its traditional interpretation, social context and histories are of little or no interest.

This paper is about Multi-Agent Systems as models of human society (we concentrate on this and do not consider Artificial Life). Although there are many fascinating and important issues raised by the modelling of animal groups, we prefer to focus on people and society in the hope that we will come to understand cognition and society better. Our aim in this paper is to discuss some of the issues raised by Multi-Agent Systems and show how they relate to various putative processes of cognition.

We strongly believe that Multi-Agent Systems constitute a powerful tool with which to examine the relationship between cognition and society. A human society is composed, ultimately, of individual human beings who communicate with each other and who act in various ways in order to navigate their way through the world and through society. People are brought up in social settings and much of their behaviour is socially oriented (for example, conversations at bus stops between strangers). We are all immersed in society from the moment we are born: how we view the world (cognitively) is related (probably closely) to our social interactions and how we view the society in which we live. Furthermore, this can be seen as a two-way process: cognition and society interact in complex ways. Collectively, we construct society and individually we interpret it. However, we can also argue that we construct society individually and we collectively interpret it: this is the paradox of social organisation.

Social organisation can change as a result of intentional engineering—the Soviet and Nazi régimes showed how this could, for them transiently, be done: cognition (beliefs about how society should be) was turned into organisation. However, we do not need to rely upon such formidable examples. As we argue below, our perceptions of society, and, hence, beliefs about society, about values and about others, can be strongly influenced—indeed formed—in various ways, advertising being one (and what is the point of advertising without an audience that belong to a society and that shares the values of the society's culture?) While reading this paper, the reader should bear in mind that we are interested at all times in the relationship between cognition and society.

It is our belief that society and cognition are strongly interdependent and that cognition in (the human sense) *cannot* occur without a social setting that is situated in the physical world (sadly, we do not have sufficient space to argue this in detail).

In the next section (section two), we introduce the concept of an agent, although for reasons of space, we leave the concept of agency undefined and expect the reader to supply an appropriate definition; we consider agents to belong to societies or communities. We expect agents to behave in ways that are conditioned by social setting (including culture). Section three is concerned with the topic of communication. Multi-Agent Systems traditionally rely upon communication to gather and exchange information. Communication is a particularly important social behaviour (cf. [18], p. 21) and, if the Sapir-Whorf hypothesis is accepted, shapes our view of the world and of the society in which we live. We consider communication from the viewpoint of *meaningful and situated* action, and we also consider some of the limitations on communication. In addition, we briefly discuss some issues relating to information gathering and knowledge of other agents. In section four, we examine some of the problems posed by distributed state (state information must be exchanged between agents for a number of reasons). We consider the ways in which state information can be exchanged between agents and the limitations that are imposed by nature on these processes. Section five considers the problem of control within Multi-Agent Systems and contrasts some traditional approaches with methods of social control and regulation found in human societies. Section six contains discussions of how various traditional aspects of Cognitive Science are involved in Multi-Agent Systems, and, more importantly, in constructing theories about cognition and society.

2 Agents

In this section, we consider the nature of agents[1]. Unfortunately, for reasons of space, we cannot examine every aspect of the concept of an agent in as much detail as we might like. In particular, we will rely upon the reader's intuitions about the concept of agency rather than giving a definition. We will assume, that is, that the reader has an adequate conception of what it is to be an agent and to have agency in various situations. The kind of agent we consider here is more or less autonomous. In the limit, we will expect agents to be autonomous in their decision making and in their acting in the world: that is, we consider our agents to have the same degree ofautonomy in their behaviour that is exhibited by people. People exhibit different

[1] For a variety of reasons, we would prefer to refer to what we call agents as 'actors'. For technical reasons, we will use the standard terminology.

degrees of autonomy depending upon time and context. In certain situations, for example in military organisations autonomous decision making is discouraged; in others, autonomy and independence are highly prized. However, an off-duty soldier is freed from the constraints that apply at other times and can exhibit more degrees of autonomy.

For the sake of simplicity, we will assume that agents exhibit as much autonomy as people: this is not necessary for a study of Multi-Agent Systems, but, as we have noted, we are particularly interested in constructing theories about the interaction between cognition andsocial setting and structure. Unlike the kinds of agent found in Distributed Artificial Intelligence—in particular, in distributed problem solving—we expect our agents to be capable of doing more than simply solving problems. In distributed problem solving, agents are first and foremost construed as problem solvers, and agents that must communicate with others only second. This strongly contrasts with our view of how an agent behaves. In particular, we view problem-solving as a possiblyimportant aspect of an agent, but regard it as just *one* activity that an agent can engage in.

We consider agents as being able to exhbiting a number of qualititatively different kinds of behaviour: problem solving, reflecting and introspecting, communicating and interacting with the environment and with others in rational ways, reasoning or reflecting on others. For us, agents exhibit rich behaviours that are not necessarily tied to any particular problem-solving context: indeed, behaviours may be exhibited in response to social setting or to social events. This richness of behaviour is important because we are interested, *inter alia*, in how the cognitions of an agent condition and are conditioned by the social context in which it finds itself and the culture in which it lives.The context (in particular, the social context) in which an agent interacts with others and the culture in which it has grown up and in which it lives are important to our view of Multi-Agent Systems. In Multi-Agent Systems, the agents that comprise a system are in social relationships of one kind or another with each other.

We view a Multi-Agent System as a model of a social organisation. There are relationships between the agents and their behaviours and the social organisation. For people, actions are constrained by society and culture and by the social context (or 'situation') in which they currently find themselves. For example, some acts are forbidden by society or culture in general (murder and incest are two examples)[2]. In a given situation, there will be actions that are forbidden or considered abnormal, while there are other actions that are expected. If a lecturer removes his or her clothes while giving a lecture, we would normally consider this as abnormal: under the *normal* social rules for giving a lecture, such behaviour is not permitted. However, if the lecturer shows photographic slides or a video tape, the behaviour is

[2] It should be noted that these cultural or social restrictions can differ across time and culture. While murder is almost invariably forbidden, restrictions on incest vary: the ancient Egyptians, for example, permitted incestuous marriage.

considered as normal or expected. We are not free to behave as we would wish in any situation we care to mention: there are always constraints on what we can and *should* do[3].

There is a further dimension to the issue of permitted or licenced action: this is the explicitly *cultural* dimension. We are all born into a culture and we grow up in a culture. Culture influences our behaviour and provides a context within which we act. As we noted above, culture determines, in general, which acts are permissible and which are impermissible. Our relationship to our actions and to the actions of others is shaped, at least in part, by the culture in which we live. More particularly, the culture in which we grow up exercises a strong influence on our attitudes and behaviour. Such attitudes may change as a result of our individual (in some cases) or (more commonly) collective rational reflection upon them and their justification and worth, but they are, nonetheless, instilled by our culture.

The historical perspective on an individual agent is important, therefore, because agents bring their past to their interactions and other kinds of behaviour. Their past experiences and, more generally, the strictures imposed by the surrounding culture influences the behaviour of an agent in any particular situation. The general cultural milieu can also exert an influence over a longer time frame.

For example, in some cultures education is a worthy and respectable objective, so actions that relate to formal education are considered in a positive light—agents are encouraged to obtain an education and to obtain the best possible results in examinations. In other cultures, money is what matters, so making as much money as possible is considered the highest possible achievement for an agent. Agents growing up in each of these cultures will have different attitudes to money and to education.

Cultural influence reaches back to before we were born: we not only have our own personal memories, we also have memory for our culture. Apart from the influence that social context and culture exert on agents, it is also necessary to observe that people live in an external, physical world[4]. The world is external to our cognitions and presents objects to us. Some of the objects are within our control while others are not; some of the objects in the external world are created by us for our own purposes and some are not; some of the objects in the external world are amenable to manipulation by us and some are not; some of the objects in the external world are animate and some are not.

Unlike conventional Cognitive Science, with Multi-Agent Systems we must take

[3] The reader might ask about the relationship between situation, society and thought. It must be noted that some societies have attempted to control the ways in which its population thinks: Nazi Germany is one infamous example, as is Stalin's Soviet Union. Propaganda, like advertising, is a method for controlling the attitudes (and hence thoughts) of the populace. Orwell's novel *1984* [20] has thought control as one of its subjects.

[4] For reasons of space we can say nothing here about the status of the claim that there is an external world that can be known by us. We will also defer any argument about the status of our perceptions and of the statements we make about the world.

full cognisance of the fact that there is a distinction between that which happens inside an agent (cognition, sensation and affect) and that which happens outside. In our theories we need to make distinctions between the mental and the physical, but we also need to relate them (in the simplest case because some of the physical objects with which we interact are similar to us—they are people). The relationship between internal and external must be taken seriously, particularly because people can make use of the external world to make manifest our internal states, intentions and memories. When we write a note on a piece of paper, we are using that paper as a *social object* (for example, diaries), and we are using it as an *extension* to our memory, an extension that can outlast us in certain cases. The paper has a usefulness because it has been endowed with an additional function: that of recording some information that we wish to retain until a later time. The paper is social because, unlike a memory that is in our heads, the paper can be used by someone else in inferring our aims or purposes or in recording our actions at a given time: the use of the paper is not necessarily restricted exclusively to the author of the note, but can be used by anyone who picks it up and reads it (and puts the message in context). Other external objects can be used in similar ways: a mark cut into a tree denoting a path, a flag denoting water safety, an open door denoting the occupant's availability to students. In each case an object is used in aparticular way: that way is meaningful to those who understand the code.

In each case, also, the object is used as a means of communication. Furthermore, we also use some of the animate objects in various ways: we can ask people to remind us of things, to do things for us, and so on. In summary, we make use of the external environment and the objects that populate it in various *meaningful and useful* ways: it is important to observe that utility is often related to *meaningfulness*.

One particularly important area in which we find meaning is that of social relationships: relationships between agents, in other words. Unlike the social and cultural background which usually change relatively slowly, relationships can change within the space of a very short time. Some relationships, say those between siblings or parent and child or between spouses, usually change, at least on one level, only relatively slowly. Other relationships can change very rapidly. Even within a relationship that seemingly changes only slowly, less noticeable changes occur with time.

To take an example that shows both the speed of change in a relationship, consider a couple who have been married for some number of years. Outwardly, their marriage is happy and, even in private, the behaviour of the spouses to each other is everything that one would expect from such a relationship. However, one day, the wife returns home unexpectedly to find her husband engaged in an act of adultery. All of the wife's beliefs about and attitudes towards her husband can change in what seems, subjectively, an instant. The relationship changes within an extremely short time.

Less drammatically, consider the marriage in which the spouses become dissatisfied with each other because each has appeared to the other to change: such change has apparently taken place over a long period of time, but the spouses have failed to notice this. In summary, relationships are *dynamic* and they are also *meaningful* (see [5]).

At this point, it is worth making a methodological observation. This observation will also serve to reply to a possible objection to our approach. We are assuming that the agents under consideration are highly complex. People are also highly complex. Our interest is very much in producing *theories* about the relationship between cognition and social organisation—i.e., theories about *people*. That the subject matter of this enquiry is complex should not deter us. We can carefully select those aspects of greatest relevance to our current interests and study them in isolation. This is similar to the way in which we normally do cognitive modelling and, more generally, the way in which we do science. Although the subject matter is complex, we are always at liberty to narrow our scope and to select phenomena of interest. When we construct computer models of agents, we must always be aware that we will never construct one as complex as a real person: instead we select processes and other phenomena and model them.

Before ending this section, it is essential to observer that agents, like people, are finite. People have only a finite capacity to act: there are physical as well as social limitations that restrict behaviour. Although it is not usually noticeable, and is controversial, at least to some, human memory is finite (it must be finite because there are only a finite number of neurons in the Central Nervous System, and only a finite—althoug very large—number of synapses on those neurons). In a similar fashion, artificial agents are finite: their finitude comes, at least in part, from the fact that they are implemented on finite machines. Because agents are finite, they only have finite resources. Resources cannot continually be exploited: for example, it is not possible to divide attention more than a relatively small number of times before attention is lost: cognitive overload soon becomes problematic (think of driving a car along an unfamiliar road while talking to a passenger and listening to a piece of complex music—this is all but impossible). Because of their finite nature, agents cannot be required to perform additional tasks *ad infinitum*: there comes a point at which an agent become overloaded and cannot do anything else. The finite nature of agents and that of the resources they possess and can utilise must be remembered at all times.

3 Communications

Agents, as we have seen, do not operate *in vacuo*: they form social organisations, and they can use the world for their own purposes. However, agents communicate with other agents: they do this for a variety of purposes, some of which we shall now see. Amongst other things, agents communicate with other agents in order to perform the following actions:

Informing Agents inform others of things that they know. When informing another, an agent will generally believe that the others do not know what is to be communicated to them. The informer also believes (unless the act of informing is untruthful) that the information to be communicated is true. The informer also believes that the information being communicated will be of interest or of use to those to whom it is directed.

Requesting Agents make requests for information and for actions to be performed. The requester is either ignorant of some particular fact or facts, or is unable to perform the requested task for themselves. A request obliges the agent to whom the request is directed to provide the requested information or else to perform the requested task. Requests can, of course, be turned down. The effect of a request is to extend the requester's capabilities by making the agent to whom the request is addressed act on their behalf. This is a characteristic of a number of linguistic acts: they elicit behaviour of a kind that is useful to the performer of the linguistic act, this behaviour being, in effect, an extension of the performer's capabilities and knowledge. Such actions are also intended to satisfy goals that the actor has at the time the linguistic act is performed.

Commanding Agents command other agents to perform specified actions. A command can only be issued by an agent that is *empowered* by society or by the social context in which the command is issued. The society of social context must exhibit a hierarchical power structure. A command is issued in order that another agent perform some task. The task can be useful or useless (there are many examples from the military of orders that are of no use whatsoever—for example, painting coal white or cutting a lawn with scissors or with a knife).

Asserting and Denying An agent can assert a proposition or deny it. Assertion serves the purpose of informing another agent that the speaker believes that the proposition being asserted is true (we assume that the asserter is truthful). Furthermore, an act of assertion can also inform the audience of the proposition *and* inform them that it is, as far as the speaker is concerned, true. The audience in the case of assertion might know beforehand the proposition that forms the propositional content of the act of the assertion. It is also possible that the proposition in question was unknown to the audience. However, the central function of an act of assertion is to convey the information that the speaker believes the propositional content to be true. Denial is the converse of assertion.

We could, of course, cite many more examples of communicative acts. In addition to the relatively neutral examples given above, we could also mention acts such as comforting, admonishing, rebuking and praising. Each of these acts has an effect that is affective (emotional): the aim of performing such an action is to induce a change in the affective state of the audience. Part of what it is to rebuke another is to inform them that the speaker believes their action to have been inadequate or improper and also to make the hearer feel guilty or uneasy with respect to their previous action. Conversely, when one praises another, one wants them to understand that we approve of their actions and we want them to have a positive emotional relationship with their actions and with the reasons for their actions (their "motivation" in other words—we must be careful about admitting the concept of a "motive" without far deeper analysis). Before moving on, we should note that we have so far only considered *isolated* linguistic acts.

Linguistic acts always occur within a context and are often performed with specific goals in mind (whether the goals are explicit or implicit, conscious or unconscious[5] is irrelevant). Furthermore, linguistic acts are often performed as part of an act of communication that is more extensive in time than a single linguistic act. It is usual that linguistic interaction consists of a number of individual linguistic acts (as well, often, as extra-linguistic acts such as gestures and eye movements). Any analysis of interaction and communication must take these sequences into account (and note that more extended periods have the advantage that they afford more time for the construction of context and the analysis of utterances within an unfolding context). Such an analysis must also take into account the fact that there are uses of language that are *purely* social in role (talk about the weather in British English is very often purely social and intended not to convey information).

A further role for communication that we have not considered, one that involves the extended nature of communication, is that of describing some experience or situation for the benefit of others. This may not seem particularly remarkable or useful but it allows the audience vicariously to have the experience being described. This is important because it allows agents to have knowledge of episodes (hence events) and to have experiences without needing to be present. Second-hand experience can be used as a substitute for direct acquaintance: although it will never be as good, it will suffice as long as adequate information is provided in the description. For example, if I do not tell you that Rila Monastery is painted red and white, you will assume it will be stone-coloured (unless you have contrary evidence or knowledge).

Finally, it is worth noting (and, for reasons of space, that is all we can do here) that discourse and dialogue are important aspects of linguistic interaction. We briefly mentioned interaction in terms of temporally extended sequences of messages: we did not mention that interactions are structured in various ways. Such structure can be local in scale—e.g., the turn taking behaviour that is exhibited by conversation—or can be more global. An example of more global organisation is the proposal and rejection structure of negotiation. Negotiation is, in any case, an interesting and, arguably, important aspect of interaction: for example, we negotiate

[5]We intend by 'unconscious' the property that the goal is not *immediately* amenable to introspection at the time the action is performed. Thus we admit goals that have been "forgotten" into this category as well as goals that are either not presented to consciousness or are only implicit in the satisfaction of other goals. We do not intend any Freudian or other psychoanalytic connotations in our use of the term.

our terms and the meanings of those terms, we negotiate our conduct and so on.

For the most part in Multi-Agent Systems research, communication tends to be conceived in terms of speech or writing. Of course, language is of central importance to our (human) communication, but we should not be blind to other forms. For example, the yellow shell shape on a filling station identifies the brand of petrol as Shell Oil; an elegant woman with flowing wings represents Rolls Royce cars; a cross of a particular shape represents Christianity. There are many examples of such symbols. As we noted above, there are other kinds of phenomena that are used for communication amongst people: sometimes they require active manipulation of the environment, sometimes manipulation of the disposition of objects in the external environment, as in the case of the open door denoting availability. There are more "natural" symbols: dark clouds "mean" rain, smoke "means" fire, and so on.

This "natural" class of meaningful associations are of a different order from those examples we gave first: the concept of meaning is different in each case and there is a difference in terms of communication. In the first cases, there was, in each case, a deliberate attempt—an intention—to communicate something definite[6]: there is someone who wants to communicate and someone who interprets the message. In the second case, there is no person who actively wants to communicate—there is no intention on the part of the agency that creates the message. Although these other forms of communication exist, Multi-Agent Systems concentrates on the linguistic forms we have concentrated on. If we consider Multi-Agent Systems as being models of communities of people who interact, then it comes as no surprise that the most efficient and effective medium forms the central focus.

There is also another distinction that needs to be drawn out. In what we might call "conventional" Multi-Agent Systems (e.g., BLONDIE-III [25] and ARCHON [27], and our own CASSANDRA[9, 11]), communication is an essentially simple process. One agent typically issues a request for information and then awaits a reply; alternatively, one agent informs another agent of some proposition. A (slight) parody of the kinds ofcommunicative action performed by these systems is:

1. Requester sends request message.
2. Request message received by corresponding agent.
3. Corresponding agent returns a reply (either containing the requested information or a refusal).
4. Requester acknowedges receipt of information.

(Note that item 4. above can be omitted.) This kind of communicative behaviour is primitive in the extreme and is inherited from computer science. We need to

[6] The reader should be clear that we consider that intention is important in communication of one kind. However, we do not believe that it is important for all forms of communication. In particular, body language is not intentional communication, but it can meaningfully interpreted and can be seento exhibit regularities.

emphasise that the kinds of communicative actions we have in mind for Multi-Agent Systems are considerably richer and more complex: indeed, we have in mind an analysis akin to that in contemporary linguistics[7].

There is an additional problem with the "conventional" (i.e., purely engineering) approach to Multi-Agent Systems: the conflation of the **knows-of** relation with communication. In a conventional Multi-Agent System, agents are able to communicate with other agents. All commmunication within these systems is direct in the sense that agents are not allowed necessarily to forward messages to other agents. It is also direct in the sense that an agent can communicate via some intermediary with another. Communication is direct in the sense that, if agent A_1 needs to communicate with agent A_2, there must be a direct communications channel between them. Without such a channel, A_1 and A_2 cannot communicate. In the absence of a channel linking them, the two agents *cannot* know of each other. In this kind of system, agents only know of other agents if and only if there is a *direct* channel between them. This would not be a problem if it were not also the case that communications paths are designed into the systems and are not permitted to be dynamic. In other words, once the designer has decided on the communication patterns (in terms of connections) between agents, the agents cannot alter this organisation at runtime. The communications structure (the structure of communication paths between agents) is fixed and inflexible. Consequently, if A_1 and A_2 are not given a direct communications path by the designer, it will never be possible for them to communicate, nor will it ever become possible for them to know of each other's existence, let alone of their respective capabilities, knowledge, and so on.

The reason that is often given in support of this inflexibility is that of cost. It would be too expensive to permit agents to establish and break communications paths at runtime (dynamically). There is the cost in terms of the underlying communications medium, but there is also a cost in terms of the reasoning processes an agent must operate in order to perform these operations. In particular, an agent must *reason* that a connection must be established and it must also *reason* that a connection must be broken. Furthermore, agents must also reason about the utility of establishing such links. Finally, it must be noted that a certain amount of reasoning is involved in engaging in acts of communication: where choice enters the picture, the costs increase (for example, an agent can need to determine those agents which will benefit from the reception of a message—we will see such an example below).

There are really two issues at stake here. The first relates to the plausibility of such inflexible communication structures. The second relates to the behaviour of people (ideally, we should pay particular attention here to how much reasoning is required by people, but we will not consider this because it will take us too far afield). We will consider each in turn. (There is actually a third issue, that of the limits of design, but we will leave that until later.)

[7] It is worth noting again that we consider Multi-Agent Systems as a tool for modelling human behaviour in social settings.

Human organisations vary to a considerable degree. In some organisations, communications are flexible and can be redirected at will. Some organisations are inflexible in the communication structures they permit. For example, the Soviet society or a civil service or military organisation can all be characterised by an inflexible communication structure. In all cases, there is a hierarchical organisation and communication must follow (not merely respect) the organisational structure. Messages must first be passed to immediate superiors (who may choose to surpress or censor them) who pass them to their superiors, and so on. There may be horizontal communication, but its effect within the organisation as a whole is of little or no consequence: what matters is the upward and downward flow of information. The opposite to this extreme is exemplified by some of the newer companies that are appearing. In these new companies, there are no hierarchical relations at all and communication can be directed to anyone who might be interested or who might have some opinion on the matter in question. Such 'open' communication is also a characteristic of some families and some research groups or departments. Between these two extremes, there are many examples of intermediate structure in which communication and organisation show some degrees of freedom while maintaining some degrees of inflexibility.

There is another aspect to the flexibility of communication that is of relevance to the current discussion. The basic observation is that we meet new people and we lose touch with others. The set of people with whom we regularly communicate is not necessarily stable. In the context of a rigid organisation like those mentioned above, it is probably the case that people come and go only relatively slowly: the set of people with whom one might communicate can remain stable for many years. However, we meet new people, form impressions of them and to communicate with them in effective ways. There appears, however, to be very little, if any, overhead in communicating with others, at least in the sense we described above for "conventional" Multi-Agent Systems. We decide what to say and to whom to say it, and then we say it. Unlike simulations, we appear not to suffer from the problem of determining to whom a message should be addressed. The number of people whom we know does not interfere withour ability to engage in effective communication.

Now, it must be noted, there are situations in which considerable amounts of effort are put into communication: worrying about the best person to address a message, worrying about how to phrase what is to be said (or written), as well as worrying about what to include and what to omit. We pay considerable amounts of attention to communication when it is of a particularly important nature. During routine communication, however, such deliberation is typically absent.

Nomatter how inflexible the environment within which we are communicating, it is certainly not the case that people only know about those persons with whom they communicate. Certainly, *direct* acquaintance can only occur amongst intercommunicating (and interacting) individuals, but we can know of individuals with whom we have never communicated—indeed, we can know a great deal about people with whom it is, in every sense, impossible to communicate[8], the example of a biographer is clearly relevant here. An example will make matters clearer.

Assume that you need to have your car re-painted and do not know of a reliable person who will do a good job while not charging an excessive amount. One option that is open is to ask a friend to recommend someone, so let us assume that you do and your friend supplies a name. Before you ask for the recommendation, you have not heard of the car painter. After the recommendation, you know that there is someone with a given name who paints cars; you also make the assumption, given that you trust the judgement of your friend, that the painter will be reliable, do a good job, and will not charge too much. Then you contact the recommended person and discuss the job: you make direct contact, in other words. From this point onward, you are in direct contact with the painter. However, before making direct contact, you had knowledge of this person and, what is more, you had this knowledge without there being any direct communication with that person.

It is also possible that you had *indirect* contact with the painter, for your friend might have asked them for information such as when to telephone and what sort of price to expect. The example shows quite clearly that direct contact is irrelevant to knowledge of a person. Before direct contact is made, you will know of the existence of the car painter, the painter's name, as well as other information about them (some of this information being perhaps derived by inference rather than by being told). All of this is known *prior to* any direct contact between you and the painter. This is clearly at variance with the considerably simpler position usually adopted in Multi-Agent Systems. If we want to model human communities, we need to adopt more flexible approaches to the modelling process. In particular, we need to be able to cope theoretically and experimentally (in terms of computational models) with the introduction of new agents into a community, with the removal of agents from a comunity, and with information being passed via a third party.

The most significant property of successful communication is that it is *meaningful*. Unfortunately, we do not have sufficient space to investigate some of the factors that contribute to meaningfulness—see [3, 10] for more details. Sperber and Wilson [23] (p. 18-19) present a convincing argument against the possibility of shared *knowledge*, at least in its traditionally understood form and with its traditionally assumed infallability. They conclude that all such knowledge must be open to question in the sense that it can be mistaken, and can also be misinterpreted or misapplied in any given context. All the parties to the communication must *find* it meaningful: ideally, they should find exactly the same meaning in the communication—that is, there should be what one might call *commonly agreed meaning*.

[8]But note that we can *never* know everything about such a person: equally, we can never know everything about another person. The best we can do is to make approximations. It might, indeed, be the case that we do not know and cannot know everything about ourselves, but that is another story.

Consider the (artificial or experimental) case of an attempt at communication in which one party speaks English and the other utters randomly chosen English words. There can be no communication between these two people because the second party is not making utterances that make sense (have meaning) according to the conventions of the English language. Consider, next, the case of an attempt at communication in which one party speaks English and the other speaks, say, Bulgarian: neither party speaks the other's language. Communication is not possible between the two parties because there is no common code: they cannot extract meaning from the other's utterances. Each party makes utterances that are meaningful only to the speaker; to the other, the utterance is meaningless. The absence of a common code or common language renders communication impossible.

Consider, finally, the case in which one party again speaks English ('correct' English according to its conventions) and the other speaks English with the exception that every noun is replaced by its antonym. Communication will initially be impossible between the interlocutors, but, once the replacement of noun by antonym becomes known to the standard English speaker, communication becomes possible, albeit at a reduced rate, and possibly with more comprehension errors. The reason that communication becomes possible is that the code being used by the second party becomes known to the first: a common code is established.

For sucessful communication, agents must employ a code that is meaningful both to the producer of the utterance and to the audience: a common code. The most common example of such a code is that of a single language (English, French, Bulgarian, Macedonian, etc.), but this is not necessarily the case as the third example shows[9]. The existence of a common code is necessary (but not sufficient) for mutual understanding: it is a code that enables expression in *mutually* meaningful ways. The code adopted for communication between the agents of a Multi-Agent System need not be as complex as a natural language. Even when the focus of Multi-AgentSystems research is people, it is still not necessary in every case to consider natural language in all its complexity. There are many examples of the use of shorthands or simpler codes for human communication. However, care must be exercised when adopting a code.

By way of an example, consider the following. The CASSANDRA-II [11] and BLONDIE-III [25] systems both used a similar approach to the common code for inter-agent communication. Neither of these systems, it must be admitted, was intended as a cognitive model, but the approach is intersting and salutory. Both systems used a blackboard-like approach to the construction of agents and, hence, maintained an internal database. The internal database of each system contained complex structures represented as attribute-value pairs. In both systems, these

[9]There is another example worth noting. Consider the case of communication in which one party speaks one language, and the other party speaks a different one, but both parties understand each other's language. Although each is unable to speak in the other's language, because they understand each other's language, they can engage in meaningful communication.

structures were communicated to other agents in an uncoded form. That is, a structure held in the internal memory could be directly transmitted to another agent without the need to translate the internal object (the structure) into some public code. This approach to communication has the advantage that it is simple and easy to implement. However, it suffers from a number of problems when considered from a cognitivist perspective.

The internal database of an agent contains items that are, at least notionally, private to the agent: it is intended to be the analogue of short-term memory in people (and I cannot *directly* access the contents of your short-term memory, and you cannot directly access the contents of mine). The way in which things are (precisely) encoded in my short-term memory is something that relates to me, and is probably something that will never be discovered: for example, the way in which I interpret the items in my short-term memory depend, *inter alia*, upon my previous experiences and the ways in which I interpreted them in the past and interpret them now. It is impossible for anyone else to have *exactly* the same experiences as me, so it is impossible for anyone else to interpret things in *exactly* the same way as I do. What is in my short-term memory is private to me in more than one sense, therefore: what is in your short-term memory is also private in the same ways. It is impossible for anyone to have *direct* access to the contents of my short-term memory, just as it is impossible for them to have access to the contents of anyone else's.

The short-term memory state of an agent contributes to the state of an agent—this is an important point and worth remembering. However, the approach adopted in CASSANDRA-II and BLONDIE-III requires the following assumptions to be made:

1. The representational vocabulary in all agents is the same. This is necessary because these two systems do not translate the internal representations into a public code. In essence, the internal code used to represent items in the local database *is* the public code used for external communication. Another way of stating this is that the encoding used by agents in representing their internal (and private) states is necessarily the same for all agents.

2. The internal state of one agent *must* be accessible to other agents. The reason for this is that when an agent in one of these systems receives a message from another agent, it often merely adds it to its own local database, integrating it with the structures that are already present as a result of purely local reasoning[10].

An assumption that ismade about autonomous agents (and people) is that their internal state is inaccessible to other agents. In short, there is no mechanism by which one agent can directly inspect the state of another. However, in the systems

[10] We are here referring to messages with *explicit* propositional content. Some messages—those without such content—caused other actions to be performed. In CASSANDRA-II, for example, a **stop** message caused agents immediately to abandon all processing and to terminate.

under consideration, the internal state (or part of it) is communicated from one agent to another. Thus, agents are given direct access to the internal state of others: this directly contradicts this assumption about agents. Not only are items taken directly from short-term memory, they are communicated in the form in which they are stored internally. This has the consequence that all agents *must* have the same internal representations as well as the consequennce that the privacy constraint is violated. Finally, it must be observed that this violation is completely implausible psychologically.

An issue related to communication will serve to connect the discussion of this section with that of the following one. This is about the dissemination of information within a collection of agents. This issue is raised because it introduces a number of arguments about agents.

A very simple method for communicating information between agents is the following. Agents sometimes have information that they consider should be communicated to other agents (how they decide that it should be communicated need not bother us now). The information might be the results of problem-solving activity or of some enquiry or deliberation. An agent might decide that the information is sufficiently interesting or useful that other agents might benefit by possessing it. The agent then communicates the information to other agents. But to which other agents? The simplest answer is for the agent to send the information to *every* other agent in the collection. This simple method ensures that every agent that can benefit from the information will do so. However, this method also has the disadvantage that the information will be communicated to agents that do not need it. This seems not much of a problem until it is considered that the bandwidth provided by any communications medium is finite: if enough information is transmitted, bandwidth is absorbed, and, in the limit, the medium becomes saturated. Saturation can cause messages to be lost or for information to be corrupted; it can also cause delays and is to be avoided where possible.

There is, though, another problem with this communication strategy (which is sometimes called "result sharing"). The second problem is that messages that are communicated in this fashion are not equally interesting to all agents. Some agents will not be interested in or concerned with the contents of any randomly taken message: agents have specific interests. This might not seem to be a problem, but it must be pointed out that the interpretation of a message consumes time and resources. While an agent is interpreting a newly received message, it diverts its attention from other tasks. By reading the message, an agent is diverted from what they were doing previously. In certain contexts, such a diversion can cause an agent to miss important events or to miss deadlines: agents are always finite and only have finite resources (as we have noted).

As a consequence of these arguments, the result sharing method of communication, although simple and easy to implement, can be seen to be a poor way of organising inter-agent communications. The strategy is too expensive in terms of communications and in terms of resource consumption to be generally applicable. When people receive messages that they consider irrelevant or uninteresting, they tend to dismiss the message, sometimes angrily.

4 State

At the end of the last section, we discussed an issue connected with the state of agents. In abstract terms, each agent has an internal (local) state that is disjoint from the internal states of all other agents. An agent's state is completely private in the sense that it cannot be inspected or updated by any agent other than the one whose state it is. The internal state of an agent can be *influenced* by the actions (including communications) of other agents, but other agents can, in no way, directly alter the agent's state.

A more general problem, or so it would seem, is that the global state of a collection of agents is inherently *distributed*. The absence of a global state becomes a problem in certain circumstances, in particular it can be construed as co-ordinated activities such as a problem for group problem solving.

In conventional problem-solving systems, there is a global state to which all operators are applied. In a paradigm like heuristic search, the heuristic evaluation function is applied to the global state to determine the next operator to apply and to determine how much progress is being made towards the solution state (the two need not coincide, note, although commonly they do). The system has available to it a global representation of the state: this allows it access to the *entire* solution process. All points of the state can, in principle, be accessed and measured by the heuristic evaluation function. In a system whose state is distributed, only local measures of progress can be made: when the evaluation function is applied in one region of the space, its result does not necessarily have any relation to the value of the function in another region. Global measurement of the space cannot be directly performed because the space has been distributed: instead, it must be constructed from the various local measurements, and this leads to the problem that the global function must be expressed in terms of some combination of local measurements.

In a distributed problem-solving system (a system whose purpose is to solve a problem in a distributed manner), the problem of distributed state is especially pressing. In such systems, each agent is an independent problem solver to which a part of the problem is assigned. Each agent is intended to solve its (local) part of the problem and pass its local (partial) solution to other agents who then integrate the local solutions into more global ones: this process continues until a global solution, a solution to the entire problem, has been constructed. Each agent works at a different rate (as in parallel and concurrent systems, there is no *a priori* way to determine how fast each agent will work, nor is there any effective way to control the work rate).

At various points in the problem-solving process, it might be necessary for an agent to have knowledge of what other agents are doing. There are numerous reasons for this, but we will give only some. One reason is that knowledge of the behaviour of other agents can be used to determine whether it is on the solution path. There is no way for an agent to make such a determination without information about other regions of the search space (this is a consequence of the need for integration of local information mentioned in the last paragraph). Agents may need to have non-local results in order to construct their local solution: such non-local results must be obtained from other agents who are working in different regions of the search space. Finally, a solution to a local problem may only be derived with the help of another agent for some other reason (e.g., resource constraints).

For example, it may be necessary for an agent to engage in synchronised activity with another agent. In order to synchronise, the agents in question must have some kind of knowledge about the state of the others. An example of human resource utilisation shows this clearly. Consider the case of a group of people engaged in translation from one language to another. They all need access to a bilingual dictionary, but only one person can have access to the dictionary at any one time. While one agent is using the dictionary, the others have to wait. While an agent is using the dictionary, the others are not permitted merely to grab the dictionary: social rules as well as knowledge that the dictionary might become damaged (and hence useless) help to prevent such behaviour. When the agent with the dictionary has finished with it, they pass the book to another agent according to some rule. However, the agent currently using the dictionary might be asked questions by the other agents as to how long it will take for them to release the dictionary. Answers can range from the relatively uncommittal ("In a minute", for example) to a detailed description of what must be done. The latter provides the other agents with information about the current task, and they are able to form an impression of the length of the task given their previous experience. The description of the task reveals information about the state of the speaker's activity.

There are limits on how much and what can be communicated. There are limits to how much information that can be communicated because there are bandwidth limitations on any communications medium. There is a limit to the number of bits that can be sent along a cable (copper or fibre-optic) or along a radio link. Available bandwidth tends to increase with time and improved technology, but there are physical limitations. There are also temporal limitations on physical media: because of propagation delays, there is a minimum time that it takes to send a message (this minimum is independent of the delays imposed by the electronic circuitry and by routing and other switching). In engineering terms, when we come to construct a Multi-Agent System on computers, there are physical limitations that limit the performance of the system.

When people are concerned, there are also limitations. There is a limit to the number of words we can utter in a second; there is a limit to the number of words we can write in a second, read in a second or hear in a second. Everything we do is limited by the maximum rate at which neurons can fire (of the order of once every 10 msec). Because our processing rates are limited, we humans can only process a certain amount of information per second. This limits the amount of information we can communicate.

There is a limit to what can be communicated, too. Pylyshyn [21] has argued that only certain aspects of the human mind are available to conscious introspection. This is called the *cognitive penetrability* hypothesis. For example, there are parts of the central nervous system that are actively engaged in sensation and perception, but to which we have no conscious access. Although this hypothesis is controversial, it is at least relatively easy to see what Pylyshyn means. For example, try to introspect on your sensing of the room in which you are reading this paper: try to get inside these processes. It will quickly become clear that such access is impossible. Another example is to try to introspect on what your visual cortex is doing at present.

A corollary of the cognitive penetrability hypothesis is that we cannot have control over every mental process. For example, try to stop yourself seeing the colour red—you will soon find that this cannot be done, try as you might. The consequence of this is that we cannot consciously access (observe and affect) many mental processes (we are using the term 'mental' in the widest possible sense). Because we cannot access them, we cannot talk about them to others. We are limited in what we can say in this case because we are unable to know of these processes and hence we are unable to form any propositions about them[11]. In terms of information about an agent's state, it is clear from the above that people do not have access to the details of their internal state. For example, when describing the content of the visual field, it is impossible for one to report on what a particular group of neurons is doing in Brodman's area 17 (primary visual cortex), nor can we know what is happening to neurons in Areas 22 and 23 (auditory cortex) when we are presented with an auditory stimulus. Complete descriptions of state are impossible because we do not have the necessary access to our neurophysiological processes, nor do we have access to some of our mental processes, particularly those closer to sensation and perception.

In the case of a distributed problem-solving system, it would appear ideal if the agents that comprise the system could produce what was, in effect, a core dump. Such a dump would contain all the information needed to reconstruct their state at any point during the problem-solving process. This is, of course, implausible if the agents of such a system are to be interpreted as models of human problem-solvers. Such an approach suffers from other problems as well.

The first problem is that such a dump would be extremely large, although it would be finite. It would require considerable amounts of communication resource in order to transmit the information to other agents: the resources consumed would

[11] The last proposition of Wittgenstein's *Tractatus* [26] is, on a superficial leve, close to our statement. We do not, however, intend our statements as an affirmation of Wittgenstein's earlier philosophy.

be time and bandwidth use (saturation of the communication medium might be a possibility). The next problem is that such a dump would require the interpreting agent(s) to have detailed knowledge of the code in which it was expressed: without such knowledge, the dump would be uninterpretable. Thirdly, the dump would require considerable resources in its interpretation: there would be a considerable body of data that would not relate to the state of the problem-solving process, but, instead, would relate to other aspects of the agent's functioning. This kind of interpretive process would take not only cognitive resources but would also require time to complete: once the analysis is complete, the world will have moved on and the dump will reflect only history (this is an extremely important point and we will return to it below).

These three arguments show that the "complete dump" approach is wholly unworkable. If we intend our agents to be models of people, we could not even have a complete dump. This is because cognitive penetrability only applies to some of our mental processes, and we cannot, *a fortiori*, communicate information about much that goes on in our nervous and mental systems. Secondly, there is the problem of converting all of the information into natural language: this is an encoding problem and it is a communication problem. There are many experiences that we have difficulty in expressing in words. If we had to express *everything* we would have to resort to vague feelings and our reports would become vague in some respects. Furthermore, a complete report on our internal state would take a considerable time to complete. There would also be the problem of determining what was relevant to the problem-solving process and what was not: for example, a vague itch above the left ear might be distracting, but it is, of itself, irrelevant to a description of how one is getting on with some problem—there is a problem of relevance. There is also the problem that much that goes on is concurrent in nature. It is notoriously hard to reproduce the behaviour of the simple systems we currently build—the interpretation of such a dump would also suffer from these problems, but in considerably more extreme ways.

Of course, people do not communicate in anything like the way suggested by the last couple of paragraphs. Instead of these enormously detailed reports, we rely on comparatively short utterances that must be interpreted by the audience. In response to a question about progress, we tend to report on those matters that are, or are believed to be, of relevance to the question of progress. A group of human problem-solvers, like any other group of individuals, is a distributed system in the sense that (what could be described as) state is distributed amongst the members of the group. At one level of description, there is no global state in a social group (at other levels, there is because one can describe the state or the intention (etc.) of a group, a crowd, a mob, or of, e.g., a football team). Very often when we (people) need a relatively detailed description of the global state of some group or social organisation, we resort to external aids. For example, we draw pictures and diagrams, or write reports, or use electronic aids to represent this information. Furthermore, we seem tacitly to accept that these representations of global information can be of comparatively limited utility, in particular that they can soon be outdated.

In a large software project for instance, the current state of the project may be represented by a collection of reports from those responsible for the main subsystems. These reports are written to a particular deadline and actually represent information that *predates* the deadline.

It is very important to remember two lessons from the above discussion of local and global state in relation to communication. The lessons are these:

- Information about state that is communicated from one agent to another (by any means) is *always out of date* when it is received (it is probably out of date when it is sent). This is because it takes time to formulate the message and it takes time for the message to be communicated. (It also takes time for a message to be interpreted, of course.) Television news, even 'live'reports, is always history (live reports can involve a few milliseconds' to a couple of seconds' delay, so they are only just history).

- Information about state that is communicated from one agent to another (by any means) is *always incomplete* (for reasons given above). Reasoning about such information must always involve an element of doubt (or an element of uncertainty if that term is preferred). This incompleteness results from limitations in bandwidth as well as the inability of agents to provide summaries that are precisely accurate (summaries necessarily cannot contain all the information of the original), in addition to the fact that agents, if they are people, cannot have access to everything that could yield state information.

Information about states, dispositions, goals and other intentions is often exchanged. As an experiment, the reader could try to listen to people talking in restaurants or other places and you will hear conversations on such subjects. Sometimes this information is about past or current activities, but it can also be about the future. We have ignored the future above, but it is important: we very often communicate our future intentions to others. By telling others of our future plans or our intentions, we can elicit supportive or cooperative behaviour: this is part of the intuition behind Durfee's concept of Partial Global Planning (PGP) [14], although PGP is unsupported by any cognitive theory.

5 Behaviour and Control

We have often mentioned behaviour above, but we have said relatively little about what kinds of behaviour we should expect from the agents in Multi-AgentSystems. In this section, we will discuss these issues in a little more detail. The reader should be warned, though, that the discussion is necessarily incomplete: we do not have space to do anything more than scratch the surface. For reasons of space, we will concentrate on the kinds of behaviour that have been historically associated with Multi-Agent Systems of a particular kind—those discussed in [4] (even though this collection of papers is now somewhat old), even though they are almost always have an orientation to distributed problem-solving. The kinds of behaviour we discuss in this section can also be referred to as forms of *controlled, collective behaviour*: the fact that control of a particular kind is involved is of considerable importance. It is conventional to describe agent behaviour as falling into one of the three following categories:

- Cooperation.
- Collaboration, and
- Competition.

(To these we could add negotiation, but we will defer consideration of this kind of behaviour.) These three kinds of behaviour are often considered as the most important kinds of behaviours that can be exhibited by collections of agents. A central issue is that the agents should behave in ways that are *coherent* in nature: we have already discussed coherence in terms of integration of local views into a global view of a problem-solving process.

We begin with the concept of competition. The meaning of the term 'competition' in Multi-Agent Systems is the same as in everyday use. Although it has been shown theoretically [17] that competition produces inferior results to cooperation, competition nevertheless has advantages. For example, during the 1970s, ARPA awarded contracts to a number of research groups to enable them to work on the problem of speech recognition: the groups were in competition to produce the best system. In Multi-Agent Systems, the idea behind competitive behaviour is often the same as that behind the ARPA project: to set different groups of agents off on some task with each group proceeding in a different fashion. The efforts of all the groups can then be compared and the best chosen (according to some criteria). An alternative, of course, is to let all the groups work on the problem and then accept the first solution that is produced. (Note that the concept of problem solving enters into the discussion, just as it will with the two other kinds of behaviour.)

The two other kinds of behaviour can be basically described as "mutually supportive" in nature. Some researchers have made distinctions between collaborative and co-operative working in Multi-Agent Systems. One such distinction is that co-operative working is less formal and involves less interaction than does collaboration. However, a moment's reflection on these terms suggests that such distinctions are incorrect. Before there can be any collaboration (literally, 'working together'), there must be an intention to co-operate. There is no sense to collaborative working in which the parties are not prepared to co-operate: collaboration cannot occur when the parties involved are unco-operative. The parties involved must have a commitment to act in ways that are mutually supportive as far as the common aim is concerned. When other matters are concerned, the behaviour of the parties may vary and be at variance with each other: when it is a matter of the aim that is to be fulfilled by their interaction, they must act towards that common aim in such a way as to promote it and to satisfy it. In other words, *co-operation precedes collaboration.*

We can then treat collaboration as an activity that is performed by agents that have already expressed a commitment and willingness to co-operate on the common project: in other words, we are arguing that collaboration is *precisely* joint working. It is necessary to point out that in systems where competition is the chosen startegy, there is still room for collaboration. Collaboration would occur within competing groups. There is no conflict inherent in this design decision.

We will conclude this discussion with a short examination of a method of control that seems plausible, but which is severely problematic. The method appears plausible in a purely engineering on traditional AI grounds; it even looks plausible on some organisational grounds. Unfortunately, it is a poor method. The method is easily expressed. A collection of agents is divided into a central planner and a collection of worker agents. There is only one planning agent, but there are many worker agents (in a collection of n agents, there will be $n-1$ workers). The central planner takes in a specification of the problem and outputs a complete plan of action for each of the workers (for a collection of n agents, the pallner produces $n-1$ plans, note). Each worker is sent its plan and, as soon as it is told to start, it begins work, slavishly following the prescriptions of the plan.

The kind of plan that we have in mind is a highly detailed one: we can consider plans of this detail because they allow us to have extremely simple worker agents. The workers have enough cognitive structure for them to execute plans. Worker agents may have some limited sensing capabilities, but they are not required to engage in sophisticated cognitive processing (such as planning).

The plans produced by the planning agent can be thought of as being akin to those produced by a classical, hierarchical, non-linear planner. This is not a necessary assumption for what follows, but is intended merely to show the level of detail that we are assuming for the plans. The classical planning model also implies the existence of mechanisms for plan modification and error correction.

This kind of Multi-Agent System would probably work as follows. The planner produces a detailed plan for each agent in response to the problem specification. The worker agents' plans are sent out and received by the workers. A start signal is emitted by the planner and the workers start operations. All goes well until one of the workers finds that its plan has failed. The reasons for failure can be numerous, but a common reason is that the world has changed, causing the operator to be executed to become inapplicable. When an operator is inapplicable, its effects cannot be produced, so subsequent parts of its plan cannot be executed (remember that we are assuming relatively stupid worker agents—they are not endowed with sophisticated cognitive apparatus). A plan failure has occurred, and the only agent that can repair the plan is the central planner (by assumption). The central planner has to be informed (by means of a message) of the fact that a failure has occurred, the nature of the failure and its location in the worker's plan.

When the planning agent receives a failure message, it can attempt to diagnose the problem and attempt to patch the plan. Failure diagnosis, as well as patch creation, requires knowledge of the way the world is near to the worker agent. There may be some considerable distance between the planning agent and the workers, so the world may look (and be) very different at the various sites. Information describing the worker's local environment must be communicated to the planning agent in order for this information to be employed in the repair process.

The plan repair process is not as simple as merely constructing a new plan (or patching the old plan) for the worker agent. A worker's plan may involve interaction with other workers. Interaction might take the form of synchronised activity rather than communication of information: both are possible, but the latter requires sophisticated cognitive apparatus, so we will concentrate on the former. The synchronisation of one worker's actions with those of another must also be taken into account in the new plan—this is a kind of interaction, note. The process of planning such interactions can, in the limit, involve planning interactions between *all* worker agents. If interaction between all workers needs to be replanned when the original plan has failed, *all* worker agents will have to be stopped when one of them detects a plan failure. This is not necessarily the worst case scenario because there will always (unless matters are strictly controlled) be the possibility that such interactions will occur: thus it is necessary always to stop all worker agents as soon as a plan failure is detected.

When the plan failure notification is sent to the planning agent, bandwidth is used this losing potentially valuable work. If there is more than one plan failure at any one time (and this is highly likely), there will be many notifications. These notifications will consume communications resources. A significant amount of information has to be transmitted with the failure notification in order for the planning agent to have sufficient to attempt to modify the plan. There is a significant risk of saturation of the communication medium. When the planning agent receives the failure notifications, it must decide on an order in which to attempt the replans, and, during replanning, it must determine which, if any, of the other plans are automatically repaired as a result of the fix it has just made. These processes are computationally expensive: in real-time contexts, very heavy demands are imposed upon the planning agent. Once a fixed plan has been created, it must be returned to its worker agent.

In a domain that exhibits significant dynamism (say, the real world), local (i.e., worker-specific) plan failure can be expected to be an extremely common event. At any one time, it is almost certainly the case that *at least* one agent will be reporting a plan failure. The planning agent will have considerable amounts of work to do and will require the exchange of significant amounts of information in order to effect a repair. Indeed, in the limit, the planning agent will be unable to effect a complete repair because of the unavailability of the right kind of information needed in order to effect the best possible repair: under these circumstances, plan failure can be expected to occur very soon. Furthermore, the resource demands imposed upon the planning agent can exceed its available resources: this will lead to degradation in performance and eventually to planning agent failure. In dynamic environments, the world might change so fast that no amount of catching up can take place: it is quite possible that the information that is required by the planner cannot be sent by the worker (for reasons connected to the worker's cognitively impoverished nature, for reasons connected with the range and limits of sensing, or for reasons connected with the nature and capacity of the communication medium).

Then there is the problem of planning–agent failure. If the planning agent fails, no plans are produced. When the problem specification is presented to the system, planning agent failure means that no workers will have any work to do. If failure occurs during a working session, no plans can be repaired, so the collection of agents cannot progress. We have argued that the most appropriate thing to do when an agent signals a plan failure is to stop all workers from doing any more, so the combination of plan and planning agent failure will cause the system to stop for an indefinite period of time. (In fact, this is equivalent to deadlock.)

The centralised, planning agent approach has been cited as a good engineering solution to the problem of behavioural control. The arguments above show that it is heavily flawed. Indeed, it is an approach to control that is doomed either to failure or to enormous inefficiency, even though it has been adopted in real societies, e.g., the Soviets, and, possibly by the British Government since 1979.

With the possible exception of the central pallner (whose power is only illusory) the kinds of collective behaviour that we have described are all long-term and somewhat rigid in nature. There must be stringent controls in order to ensure that all agents co-operate during the operation of the system. The concept of coherence implies that, as far as possible, all agents act in such a manner that promotes the collective attainment of some goal—it is no accident that these concepts arose in distributed problem solving. In order to achieve coherence, information must be exchanged between agents and agents' behaviour must be constrained so that they do not deviate from some ideal norm: such behaviour must be extremely rigid and highly controlled[12]. Such controls are, in fact, designed into systems of this kind.

When we consider 'natural' societies, on the other hand, the control mechanisms that are encountered there are of an entirely different kind from those we have considered in this section. Modern societies of all kinds are regulated by laws. Societies that we might prefer to regard as 'primitive' do not have the codified structure of a modern legal system, but, instead, rely upon the concept of a taboo. Just as in a legal system, the infringement of a taboo brings with it a penalty that

[12] This is almost an authoritarian concept: we would like to examine other kinds of social order.

must be paid. In the codified systems that we are used to in Europe, penalties are often expressed in terms of loss of liberty or of money: we are imprisoned or fined for infringing our laws.

The legal systems that we have in Europe tend to operate in (at least) two ways: prohibition and obligation. Certain acts are prohibited by laws (e.g., selling proscribed drugs, driving on the wrong side of the road, obtaining money by violent means), whereas other acts are mandated by law (there are many examples in social law). The system of regulation, works, on the one hand, by requiring us to act in certain ways, and, on the other, by attempting to make us to refrain from acting in other ways.

While we do not act in proscribed ways, we act legally. When we act in ways that contravene the law, we are expected to pay a penalty ("society exacts its penalty" is a common way of expressing this): the threat of punishment, perhaps together with less tangible, though still unpleasant consequences (e.g., ostracism) are the ways in which this kind of legal system operates.

There are two things worth noting here. The first is that the precise system of laws held by a society changes with time. Despite such things as the Divine Right of Kings, laws are made by special sections of society. In democratic countries, laws are made by an elected body of representatives who are *supposed* to act in the best interests of those who elected them[13]. Just as laws can be made, they can be unmade (repealed). When a law becomes outdated, it is repealed in favour of a more modern version. When a law is seen to be iniquitous, it should be replaced by one that is not so injurious[14]. Society at large, in a sense, allows legislation only by consent: where there is no concent, there can be no rule of law. The Poll Tax legislation of the early 1990s was seen as iniquitous by a significant proportion of the population. The law was so unpopular (even being the cause of riots in central London) that it could not remain on the statute books: it could not be enforced and the consent of the majority was not had by the legislators[15]. Control by legislation is a two-way process that requires, ideally, the co-operation of the legislature and of society at large.

The second point is that control by legislation is an indirect method of control. There are other mechanisms that can be employed to control a society: fear is one such. The imposition of martial law is one way in which to control the actions of an entire society: it is, of course, a radical method. Within smaller groupings, power structures develop: management hierarchies in companies and universities

[13] The current (October, 1994) corruption scandal amid the ranks of the UK's governing party strongly indicates, once again, that elected representatives often do not consider as binding or important their responsibilities towards the electorate and towards their duties as democratically elected representatives of others.

[14] Perhaps this is only a liberal dream, these days.

[15] It will be very interesting to see what happens with the new (1994) Criminal Justice act. It, too, is highly unpopular and is felt by some to be the beginning of the introduction of a police state.

are also examples. However, these also work, at least to some extent, on the basis of fear and of penalty and reward (reward is a concept absent from most legal systems). Reward in career is often expressed in terms of promotion, increased salary and stature, as well as greater power to determine the actions of others. Employees can also be penalised: for example by being passed over for promotion, being ignored when decisions are to be made, opinions solicitied, and so on. The fear component is clearly that of loss of job and attendant income, although it can (perhaps less commonly today than some years ago) be in terms of termination of one's involvement with a particular activity (e.g., a project).

Human societies, whether 'developed' or 'primitive', do not rely exclusively on their systems of laws or taboos (and note that there can still be taboos even in societies that have extensive legal systems). There are also systems of *moral codes* that apply to the members of a culture. Moral codes may, of course, relate to taboos as well as to laws. For example, the Christian religion has an injunction against killing. One of the Ten Commandments—the central moral tenets of the religion—explicitly forbids killing. In a similar fashion, theft and adultery are also forbidden to followers of this religion. Other religions have parallel codes of conduct.

What is interesting about these systems of religion-based moral codes is that their authority is claimed to be above the social level: authority is invariably invested in a supernatural deity. This has the consequence that mere mortals cannot question the moral code accompanying the religion, for to do so would be to commit an act of heresy. Furthermore, because the moral code comes, ultimately, from a higher authority, it is not society's position to alter the codes. The code can be changed only with divine intervention.

Here again, we can see the operation of a reward/punishment system. The rewards, at least in Christianity, are a pleasant afterlife. The penalties are possible excommunication from the Church or ostracism from society while one is alive, or eternal damnation when one is dead. Other religions have different approaches to both rewards and punishments (even Buddhism which aims at enlightenment—a punishment for failure to follow the codes of this atheistic religion is to fail to find enlightenment). The two concepts however appear to be of considerable some importance to these systems of belief.

Moral systems, such as the Christian one we have discussed, serve to limit the potential behaviours of members of society of believers. The injunctions against killing, adultery, theft (and theft might be regarded as encompassing adultery under some older or more reactionary interpretations), and so on, delimit the bounds of conduct. Action that falls within the moral code is permitted; action that contravenes the code is sinful and is to be punished. Christianity, at least in some of its forms, has added to this the concept that even thinking about a sin is equivalent to committing the sinful act: thoughts count as deeds.

As we noted earlier, some moral codes, and Christianity is a prime example, serve as the basis for legal systems. This confers upon the legal system an ultimate authority that must be respected by all members of the society (at least, as long as they all profess the same religion: there clear is opportunity for dissent in a pluralist society). Amongst the differences between moral and legal codes is the fact that the concepts of punishment and reward are considerably less clear-cut for moral than for legal systems.

Finally, let us consider the issue of 'thought control'. Above, we mentioned advertising and propaganda as methods of thought control: here we eill discuss these in a little more detail. In addition, both can be employed in controlling the behaviour of a society or of social groups. In Nazi Germany, propaganda was used to instill anti-semantism in the population. In a similar fashion, during the Cold War both sides were subjected to propaganda about the 'evils' of the other (the reader might recall Ronald Reagan's description of the USSR as the "evil empire"). On both sides of the Iron Curtain, propaganda was applied, during that period, to shape opinion: suspicion and even hatred of the other side was encouraged, while each side portrayed itself as being righteous, humane and liberal.

Advertising is perhaps a less obvious form of behavioural control, but, as is relatively well known, it has effects similar to those of propaganda. Unlike the other forms, advertising is typically commercial in aim (there are exceptions, for example public information films that are intended to inform the audience about things such as road safety or the availability of welfare benefits). The purpose of advertising is to persuade the audience to buy a product. As part of this process, it might be desirable to change the audience's attitudes, for example to persuade them that the kind of 'lifestyle' associated with the product being advertised is one that will benefit the members of the audience. Alternatively, the advertiser might attempt to form an association between the product and a stereotype. There are many examples of both cases.

As examples of the association between stereotype and product we can cite: the Marlboro cowboy and the UK beer TV commercials that portray drinkers of a particular beer as being tall, thin, good-looking men who are successful with women. The cowboy symbolises a rugged, outdoor existence based upon strength and the ability to stand up to the vicissitudes of nature. The beer drinkers always attract the sexiest women, and so are, according to one version of the western male, successful men *precisely* because they drink this beer. Equally, there are many UK TV commercials for detergents that show them being used either by highly competent housewives, or else by independent women who are in complete control of their lives and their children (the detergents are shown as a liberating force in these women's lives, the one product upon which they can rely for such a lifestyle). For a long time, Coca-Cola and Kelloggs have used advertising that has appealed to lifestyle. Cornflakes advertising has suggested that one will live a carefree, healthy, sun-filled life if one eats this product for breakfast. During the period of détente, Coca-Cola TV commercials emphasised peace between nations: a considerable change in lifestyle from the then predominant antagonism between

East and West.

Lifestyle is a characteristic message of the other kind of advertising. Suntans were popularised by Coco Chanel, but their massive popularity has, arguably, come about from the printed media and later from film and television. There has been an association forged during the last sixty years between the tanned body and a healthy lifestyle: the aim of becoming tanned was seen until recently as being one which promotes health. (In recent years, the dangers associated with over-exposure to ultra-violet radiation has begun to change attitudes towards the suntan.) Any product associated with being suntanned or with obtaining a suntan was promoted as something that would lead to a healthier life: very often these products also had a glamour value associated with them,

A second example, a more recent one, is that of environmentalism. Many products have appeared in the last five years that have been claimed to be environmentally friendly in the sense that they are less polluting, require fewer environmentally damaging chemicals in their production or growth, or in consuming fewer natural resources. The perception of the environmenally conscious member of society has changed since the mid– to late–1980s. The image of conservation first attracted public awareness through the activities of such movements as the German Green Party. Thereafter, issues such as global warming became a subject that people worried about, as was the quality of the air we breathe[16] and the quality of the water we drink; the consequences of ozone layer depletion have also come to be seen as a significant issue that will affect us all. In addition to this increased awareness (to a significant extent, brought about by the media), advertising has turned what was once a fringe concern into a central focus for consumers' attention. Advertising and the media have played a large part in altering the habits of a number of western societies with Germany being arguably the most changed.

Advertising has had significant a effect in changing behaviour and, more importantly, in moulding opinions. There are areas other than environmentalism and suntans where advertising has altered behaviour. For reasons of space, let us mention just one more: the youth culture. Advertisers have promoted the concept that youth is all-powerful and that those who are no longer young are, in some sense, useless. It is interesting to read a newspaper, magazine or to watch TV commercials because one comes away with an over-riding impression that youth is valued highest, at least in the popular imagination. There are products that make one look or feel 'younger'; there are products specifically aimed at the younger consumer; many products are advertised by young and attractive (often very thin) actors or models. In conjunction with this, there is the push from business to concentrate on the young (universities are also engage in this): there is a widespread perception that only the young are capable of performing while the old and middle-aged are

[16] Air quality is a matter for severe concern in the UK. As this paper is being written, a UK Royal Commission has today (26 October 1994) recommended that motor fuel prices should double in the next five years as a measure to prevent the increased use of cars.

only fit for the scrap heap. One reason for this is, of course, that young employees do not call on large salaries. A second reason is that they can be worked hard before being thrown away (there is always a supply of young workers and, in some sectors, there is always a demand for employment): if conditions are right, young workers do as they are told because they fear dismissal (which can amount to a loss of face)[17]. Furthermore, young workers can be moulded into the company's image: if they do not accept the mould, they can be thrown aways.

We have now seen some of the ways in which a social system controls itself or is controlled or influenced by power groups. There are, naturally, many other kinds of control. We have only had space to review a highly restricted set of controls, and this has been done in a highly cursory fashion. What is clear, though, is that the kinds of control we have discussed here have not been considered in Multi-Agent Systems. There has been work on organisational structure (Fox [15] was amongst the first to examine this avenue), but this has been viewed in a purely organisational, and not an organisational *and* control, fashion. The kinds of control structure typically employed in Multi-Agent Systems are static and highly rigid: the control structures we have just considered are flexible and dynamic (and, in some cases, open to question and negotiation)[18]. In addition, the kinds of control exhibited by human societies are subject to change from internal pressures: this is not the case for those examined in the Multi-Agent Systems literature. Finally, control in human societies develops over time (e.g., the development of a legal system), whereas the control aspect of a Multi-Agent System is designed into the system at the start and can never change: all behaviour is strictly controlled by the design, and, hence, is inflexible and is not of the kind that affords insights into natural phenomena.

6 Cognition and Multi-Agent Systems

So far, we have not made explicit reference to cognitive processes, even though we said at the outset that part of our motivation for studying Multi-Agent Systems is the examination of the relationship between cognition and social behaviour. In this section, we will be concerned with exactly this topic. We will consider some of the ways in which the complex phenomena we discussed in previous sections relate to cognition.

Before moving on, we need to address an issue that might be raised by a reader well-versed in social theory. A distinction is often drawn between macro and microscopic social phenomena. Macroscopic phenomena relate to entire societies (e.g., a national society), while the microscopic refers to small collections of people (eventually referring to single individuals). The emphasis in this section will, quite em-

[17]In addition, there is the fact that the young may have fewer expectations of work and have less experience, both of which imply that they will accept orders without question.

[18]Gasser [16] has considered control and organisation from the viewpoint of commitments, an approach that is also flexible, evolutionary and dynamic.

phatically, be on the microscopic side: we will be concerned only with the processes active within an individual. We will, therefore, ignore the ways in which the behaviour of individuals contributes to the maintenance of a social structure: these important issues are left for another time.

6.1 Action

Above, we have frequently appealed to the concepts of action and behaviour in social settings. We have, for example, discussed the concept of a *legal, permitted*, or *licenced* action. Everyday conduct involves much action and action of different kinds. Speech is a kind of action; it can be argued that thought is another kind (but one with somewhat different properties from all others). The production of behaviour is clearly an important aspect of Multi-Agent Systems, just as it is important for Cognitive Science in general. Let us consider some aspects of action and behaviour and their production.

There are two kinds of behaviour. The first is unmotivated by any rational cause, and cannot be explained in terms of rational cause. This class is exemplified by reflex actions. The second kind is relatable to rational causes[19] and, therefore, acts of this kind have their origins, at least in part, in the beliefs and intentions of the agent: these acts are intentional and are performed for reasons that could, in principle, be explained to other agents if the circumstances were right. Part of their rational basis can be described as the property that they are goal-directed or done for a reason: equally important[20] is the fact that others can ascribe a rational basis, can ascribe reasons, for actions of this kind—the observer can make *meaningful* inferences about the action and the actor's relationship to it—the reasons that are *ascribed* are, together with knowledge of the action itself, what we normally take as being constitutive of rational action.

The view that we have just advanced is concerned as much with the observer's role in action as with that of the actor. However, in more traditional cognitivist terms, terms that do not normally deal with the observer's position, we need to consider the relationship between intentions, beliefs and actions: the account that is given within traditional cognitive science is couched in terms that relate entirely to what goes on inside an agent's head. The traditional account runs (roughly) as follows.

Agents have beliefs and intentions. Their beliefs are about themselves and about other agents, as well as about other objects (typically inanimate) in the external world. Intentions include such things as desires, wants and goals: we will concentrate on goals because they are more familiar. An explanation is that actions are performed in order to satisfy goals: there is an association between actions and goals

[19] We should be somewhat more careful here. Some authors, [24] for example, would disagree with many of our remarks about what we term "rational action" and question the explicit nature of the origins of such acts.

[20] Although we should probably say *more importantly*.

such that the actions associated with a goal are those most likely to satisfy the goal. Action is seen as secondary to an agent's goals: goals determine behaviour in the sense that actions are performed in order to satisfy goals.

It is usually claimed that not all goals are such that they can be directly translated into action (consider the goal of redecorating one's sitting room): the translation is mediated by a plan. Plans are executed to produce behaviour. When actions are performed, agents can observe them and form new beliefs (as well as modifying old ones). There is, therefore, a circularity: beliefs lead to plans, plans lead to actions and actions lead to new beliefs.

The picture presented above cannot be the whole story. What is the relationship between motives and action? Where do goals and other intentions enter the picture? Although the concept of motive is considered important by some workers, we will defer any discussion: the symbolic interactionist framework that we have adopted for much of our more recent work on Multi-Agent Systems leads us to call into question the concept of motive (see, for example, [6] for a short discussion). Furthermore, various conceptions of situated action, including that of Suchman [24], also suggest that this concept should be viewed with some suspicion. Goals have a relatively more stable position. Goals can be described as states that an agent would like to obtain. Agents form goals as a consequence of holding beliefs about the world as it might be and about how a different future of the state of the world might be of benefit to the agent. Thus, an agent holds beliefs about what would be desirable or beneficial to it: the desire to bring about this state can be described as a goal.

In much of the literature, the relationships between goals and beliefs are ignored: the planning literature, for example, treats goals as given and never questions their origins, let alone their genesis. There are only a few discussions of ths relationship that we are aware of: Cohen and Perrault [8] and Appelt [1] explicitly consider the relationship between goals and actions (in both cases, the actions are linguistic, but they also consider *cognitive* consequences of linguistic action).

The question as to how does the cognitivist view of action impact upon Multi-Agent Systems must now be raised. Agents must act in the world in various ways. In particular, agents must act in such a way that their aims and purposes are served. Random behaviour is inadequate for this: what is needed are *purposeful* acts that meet the performer's goals. This is a severe constraint upon an agent's behaviour—the conventional mechanism of the plan is often suggested as the mediating device: plans are conceived as mechanisms for translating goals into actions[21]. Plans and goals are all very well, but they have to serve appropriate purposes: this is where beliefs enter. Agents must be able flexibly to act in the world: they must respond to the actions of others as well as to events with other causes. The beliefs possessed by an agent will change with time. Given the concept of goal presented above, this implies that at least some of an agent's goals will change as the world changes. Some

[21] Perhaps we ought not to consider the concept of a 'plan', but, instead, consider the *activity* of 'plan formation' or planning alongside the reconceptualisation of a plan as a form of strategy.

goals will remain, of course, perhaps because they have not been achieved.

In traditional cognitive science, and almost always in AI, agents are considered to have beliefs about various things. In many contexts, agents are construed as holding beliefs about the tasks they perform and this is *all* they believe. If we consider people and what their beliefs are about, we find something different. People hold beliefs about a variety of things: their beliefs are not confined to the tasks that they perform. People hold beliefs about other people, likely or past behaviour (their own as well as of others, and about the social settings in which they find, or are likely to find, themselves. Many of the beliefs we have are about social matters: for example, political beliefs often relate to social matters. As beliefs change about people and other social matters, our goals and actions change in a fashion similar to that discussed above. Many of the acts we perform are social. Communication is fundamentally a *social* act. Above, we discussed Multi-Agent Systems as social systems: it can be seen that they also relate to action and the cognitive processes traditionally associated with them.

6.2 Problem Solving and Reasoning

Problem solving, and, more generally, reasoning, has been a much-studied human competence. Research in the general area of reasoning has been conducted since the earliest days of cognitive psychology and much research has been conducted within the framework of cognitive science. There are topics within this area that are still controversial: the debate about how people do deductive reasoning still rages (a recent contribution being [19]), as does the debate over whether the everyday reasoning employed by people is deductive[22]. The research conducted into reasoning processes have concentrated on the behaviour of individuals when, typically, they solve puzzles of various types.

For Multi-Agent Systems, it would appear clear that reasoning plays a central role. The reason for this is that Multi-Agent Systems are often thought of as doing some useful work. This is, however, not the only focus of attention: we are interested in how Multi-Agent Systems illuminate cognition, but distributed expert systems does not even begin to achieve this aim. In a similar fashion, group problem-solving activity (a relatively little-studied activity in cognitive science) seems an obvious case in which Multi-Agent Systems have something to say to cognitive science. Although, this is undeniably the case, we have another set of issues in mind: these other issues, and we can argue that they are much more central, are, we believe, a central set of issues that bridge individual and collective cognition. We will, therefore, ignore the classical problem of how people (and models) perform reasoning tasks for the purposes of working for a living and for the purposes of psychological experiment. The instances of reasoning we have in mind are more "natural" in the sense that

[22]To our mind, this debate has run its course. The evidence against the primacy of deductive reasoning is overwhelming.

they occur in natural settings and take place without necessarily always incurring a significant deliberative overhead.

When discussing behaviour and action, we mentioned the fact that agents must reason about their actions. In particular, they have to decide whether an action is the most appropriate or the most effective. They also have to interpret the actions of others. Reasoning about action, both one's own and those of others is an important aspect of human behaviour. Furthermore, explanation is often required, both of one's own actions and in the interpretation of the actions of others: unless we can find a purpose or reason for an act, it remains an isolated gesture that has no significance. Part of this process of interpretation is finding rationales for actions just so that we can find *meaning* in the action. In order to find meaning, it is necessary also to relate the action to the previous behaviour of the actor or of other actors (including oneself): this, in turn, involves relating the action to previous meanings.

Reasoning about action is important. Such reasoning assists us in attempting to decide how others are likely to act in the future. If we understand, or think we understand, the actions that another has taken in the past, we will be better placed to predict their likely actions in the future. We need to make predictions in order, *inter alia*, to avoid unpleasant consequences of actions, to make the best use of actions, to avoid duplicating actions (think of buying birthday presents for a sibling knowing that your parents are also buying presents), and to perform complementary actions. In cases such as armed conflict, sporting activities, or commercial competition, knowledge of how the likely competing agency is to act brings an important advantage: even a good intuition of how they will act can often reap benefits.

When we perform a task or take some course of action, we do not do so in a vacuum. We always act in a context. We also act in specific situations. It can be argued that every action is performed in a specific and unique situation Situations are connected by a number of relationships: temporal relationships such as contiguity, part-of relationships between a situation and a wider, more encompassing one, or similarity relationships. Situations can be characterised in a number of different ways, and every situation brings with it information that can be used in the interpretation and generation of behaviour.

Situations can be characterised in various ways: a number of theories, for example Schank's scripts [22] and Symbolic Interactionism [6], exploit the regularities that can be detected in everyday situations. It is important to reason about the situation one finds, or expects to find, oneself in so that we can act appropriately. Most theories of situations of this kind (Barwise and Perry's theory [2] is not of this kind) require the identification of role within a situation, as well as the identification of other aspects of the situation in order for the actors to perform their assigned roles. Such identification comes after one has recognised (or engineered) the situation: it is necessary to recognise the sitation is before anything else can be determined. A fundamental assumption of these theories is that situations determine behaviour (or, at the very least, set the limits of appropriate or licensed behaviour). When one has identified the situation, one can identify the roles of the actors. Once roles have been identified, it becomes possible to reason about the actions that are made possible (or permitted) by the situation: this permits reasoning about action, and, in particular, it assists in the prediction of the future behaviour of the other participants *as well as* of oneself. A situation provides part of the context within which to reason about action (more of the context is filled in by previous experience, knowledge of one's self, and knowledge of the other agents involved in the situation).

Situations can also be seen in a wider context. A situation, or collection of situations, can have consequences for an agent. The situations in which an agent participates may have meanings for that agent: agents might see significance in certain kinds of situation (this idea forms part of the basis of psychoanalysis). Situations can be avoided or they can be encouraged: we know the kinds of situation which please us, those which displease us or cause us pain, and those to which we are neutral. We can understand what happens in situations in order to determine, in part, how we are going to behave at a larger scale. In order to do this, we need to be able to reason not just about actions but also the contexts within which they occur. We reason about situations, in other words. We can also infer the consequences of allowing ourselves to participate in various situations: we can reason at a level above that of the individual action. This, once again, enables us to navigate through the social world in which we live. Situations are about our relationships with others and about how we act within the constrained circumstances defined by situations.

Problem-solving behaviour and reasoning in general, then, can be seen to be of significance in our social behaviour. Without reasoning about action, intention, and about the situations in which we find ourselves, we could not act in coherent and purposeful ways. Furthermore, and this is something we have not considered in detail, there is the reasoning we engage in when considering other agents: we wonder about their likes and dislikes, their preferences, their goals and so on. When reasoning about other people in order to figure out how they will behave in various contexts: for example, we often wonder how someone will react to something we do, and we often wonder how people will receive information or requests from us. We reason about other agents as well as about action.

We reason about ourselves in ways similar similar to these. We want to know how we will act and react, how we might view something, and so on. Just as reasoning about others takes place within situations and about situations, so, too, does this kind of introspection. Introspective behaviour of this kind is more general than that usually considered in the literature and more work is needed on it (but see [13] for a more detailed treatment).

6.3 Communication

In previous sections, we have discussed inter-agent communication at some considerable length (although we have not devoted enough space to intra-agent communication, but see [13] for a discussion). For people, communication can be verbal or non-verbal with verbal communication being the more flexible and subtle. Verbal communication is language use: consequently, there is a considerable role for cognitive science in the examination of communication. However, what has already been done in cognitive science and computational linguistics does not automatically cross over into Multi-Agent Systems: there are issues here that are not always given enough emphasis by the older disciplines.

The problems of production and understanding are important to computational linguistics, but language production has always been the cinderella. One reason for this is that linguistics, at least since Chomsky's *Syntactic Structures* [7] has been dominated by issues that relate to a speaker's competence: recognition—what counts as a grammatical utterance—has formed the core of linguistic theory. Problems with pragmatics have also impaired progress in the area of production or generation, as well as understanding. One significant aspect of the more computational modelling of linguistic processes is that they have been conducted in relatively poor environments: that is, the discourse that is to be understood or is to be generated relates to a relatively sparse environment with little of the richness of the natural world. With Multi-Agent Systems, such rich environments are naturally part of the experimental setting, and research into communication between agents should make use of this.

We see communication as relating to many of the other aspects we discuss in this section: indeed, it might be argued that it relates to *everything* else. We have devoted a relatively large space to action for a number of reasons, one being that communication is a form of action: acts of communication are performed, often, for specific reasons. However, the particular problems of communication considered as a *linguistic* act are numerous and difficult. By placing communication at the centre of a linguistic study that is embedded in the context of Multi-Agent Systems, issues such as syntax and semantics can be considered in a more naturalistic setting and the *uses* to which language is put—uses that relate to setting and context—can be examined in a rich context.

6.4 Memory and Learning

Finally, there are the issues of memory and learning. Much of the discussion above has made an appeal to memory. For example, there was mention of remembering previous actions and of recalling previous events: these are impossible without memory. Without learning processes, it is impossible for agents to have such memories, so we consider learning, together with memory, to be of central importance in any full theory of Multi-Agent Systems. For reasons of space, we will not consider these issues in any depth here, but will refer the reader to [12] and [13].

7 Conclusions

This paper has been about the use of CBR in the development of a holiday planning program to be called HOLS. Although holiday planning does not seem a particularly spectacular problem in terms of difficulty, it does require the application of different kinds of knowledge and is also memory-based in the sense that previous experience is used in deciding what to do next—there is more to the problem than one would initially believe. Holiday planning can be considered to be based as much on constraint satisfaction as planning proper: this increases the interestingness of the problem from a technical stance. The knowledge employed in planning a holiday is of an everyday, non-technical kind: this has the pleasant consequence that there are no problems in acquiring the right knowledge—we are all potentially expert in the domain. Furthermore, holidays form an important part of many people's lives, and planning a successful holiday is clearly important to them.

We have outlined some of the important factors in planning a holiday and have also outlined some of the background knowledge that is needed to perform the task. We noted that holiday planning can be a domain that is relatively sparse in the sense that previous experience may be lacking. Furthermore, the indexing problem in this domain is acute: this is because there are many different ways to index a holiday in memory. We then gave the broad outline of a program to plan holidays. We suggested that the program could be configured in different ways, and that the problem has the pleasant property that the amount of knowledge and range of experience available to the program could be varied quite considerably.

We next considered holiday planning as a group activity. We noted that holidays are often planned by more than one person, and that the needs or wants of others in a group played an important role in determining which plans are proposed. We argued that the extension to group problem-solving brought benefits in terms of the expanded range of experience that can be drawn upon, but that this expansion comes at the cost of more work on the part of the participating agents. We also argued that concepts from social psychology play an essential role in a multi-agent planning system.

Finally, we sketched ways in which a community of holiday-planning agents could be arranged. We briefly considered various experimental arrangements, each of which depends upon a different distribution or provision of knowledge and experience.

We wanted to consider holiday planning as a group activity for reasons other than realism:

1. Group processes and social interaction are, we believe, important in understanding cognition. The holiday planning task seems ideally suited to this, as should be clear.

2. The multi-agent HOLS program is an initial exercise in the revision of our CASSANDRA architecture [11, ?].

In [?], we argued that the original definition of the CASSANDRA architecture was severely limited, and that additional facilities were necessary to make it as flexible as we orginally intended. In that paper, we cited extensions such as introspection, knowledge of organization and the inclusion of a declarative database in the system. The proposals were based on concepts in rule-based systems.

More recently, we have come to see that the needs for communication between agents are more extensive than the simple inclusion of a communications interface. In [11], we proposed using speech acts [?] as a basis for communication: many of the proposals in [?] derived from a more detailed examination of what was needed to support communications based on the speech act theory. At the same time, we have come to believe that effective acts of communication can only be performed when agents have knowledge of the agents with which they communicate. Paradoxically, such a requirement entails that they have knowledge of themselves. As part of this knowledge, previous experience is important. The HOLS programs are an initial attempt to build systems with this knowledge and which uses previous experience as an integral part of their operation.

References

[1] Appelt, D., *Planning English Sentences*, CUP, 1985.

[2] Barwise, D. and Perry, J., *Situations and Attitudes*, MIT Press, Cambridge, MA, 1983.

[3] Barwise, D., On the Model Theory of Common Knowledge, in Barwise, J., em The Situation in Logic, CSLI Lecture Notes No. 17, Center for the Study of Language and Information, Stanford, CA, 1989.

[4] Bond, Alan H. and Gasser, L., *Readings in Distributed Artificial Intelligence*, Morgan Kaufmann, San Mateo, CA, 1988.

[5] Burr, Vivien and Butt, Trevor, *Invitation to Personal Construct Psychology*, Whurr Publishers, London, 1992.

[6] Charon, Joel M., *Symbolic Interactionism*, Fourth Edition, Prentice Hall, Englewood Cliffs, NJ, 1992.

[7] Chomsky, N., *Syntactic Structures*, Mouton, The Hague, 1957.

[8] Cohen, P. and Perrault, R., Elements of a Plan-Based Theory of Speech Acts, *Cognitive Science*, Vol. 3, pp. 177-212, 1979.

[9] Craig, I. D., *Extending CASSANDRA*, Research Report No. 183, Department of Computer Science, University of Warwick, 1991.

[10] Craig, I. D., *Meanings and Messages*, Research Report No. 187, Department of Computer Science, University of Warwick, 1991.

[11] Craig, I. D., *The CASSANDRA Architecture*, Ellis Horwood, Chichester, England, 1989.

[12] Craig, I. D., *Agents that Model Themselves*, Research Report No. 266, Department of Computer Science, University of Warwick, 1994.

[13] Craig,I. D., Agents That Talk To Themselves, *in prep.*

[14] Durfee, E., *A Unified Approach to Dynamic Coordination: Planning Actions and Interactions in a Distributed Problem Solving Network*, COINS Technical Report 87-84, Department of Computer and Information Science, University of Massachusetts at Amherst, 1987.

[15] Fox, M., An Organizational View of Distributed Systems, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 11, pp. 70-80, 1981.

[16] Gasser, L., Social Conceptions of Knowledge and Action: DAI Foundations and Open Systems, *Artificial Intelligence*, Vol. 47, pp. 107-138, 1991.

[17] Genesereth, M. R., Ginsberg, M. L., and Rosenschein, J. S., Cooperation without Communication, *Proc. National Conference on Artificial Intelligence*, Philadelphia, PA, pp. 51-57, 1986.

[18] Giddens, A., *Sociology: A Brief but Critical Introduction*, Second Edition, Macmillan, Houndmills, Hants, 1986.

[19] Johnson-Laird, P. N. and Byrne, R. M. J., *Deduction*, Erlbaum, Hove, England, 1991.

[20] Orwell, George, *Nineteen Eighty-Four*, Secker and Warburg, London, 1949.

[21] Pylyshyn, Z., *Computation and Cognition*, MIT Press, A Bradford Book, Cambridge, MA, 1984.

[22] Schank, R. C. and Abelson, R., *Scripts, Plans, Goals and Understanding*, Erlbaum, Hillsdale, NJ, 1977.

[23] Sperber, D. and Wilson, D., *Relevance*, Blackwell, Oxford, 1986.

[24] Suchman, L. A., *Plans and Situated Actions*, CUP, 1987.

[25] Velthuijsen, H., *The Nature and Applicability of the Blackboard Architecture*, Ph. D. thesis, Department of Computer Science, University of Limburg, The Netherlands, 1992.

[26] Wittgenstein, L., *Tractatus Logico-Philosophicus*, Trans. Pears, D. F. and McGuinness, B. F., Routledge and Kegan Paul, London, 1961.

[27] Wittig, T., (ed.), *The ARCHON System*, Ellis Horwood, Hemel Hempstead, Herts, UK, 1992.