On Cooperation in Multi-Agent Systems[a]

J. E. Doran[1], S. Franklin[2], N. R. Jennings[3] & T. J. Norman[3]

[1] Dept. of Computer Science, University of Essex.

[2] Dept. of Mathematical Sciences, University of Memphis.

[3] Dept. Electronic Engineering, Queen Mary & Westfield College, University of London.

Cooperation is often presented as one of the key concepts which differentiates multi-agent systems from other related disciplines such as distributed computing, object-oriented systems, and expert systems. However it is a concept whose precise usage in agent-based systems is at best unclear and at worst highly inconsistent. Given the centrality of the issue, and the different ideological viewpoints on the subject, this was a lively panel which dealt with the following main issues:

1) What is cooperation? What are its limits? How does it relate to concepts like communication, coordination and negotiation?

2) Can a generic typology of cooperation be identified? If so, what might such a typology look like? What are the key characteristics?

3) What sorts of cooperation are you likely to see in real multi-agent systems? How does the degree of autonomy, self interest, benevolence, affect cooperation strategies? What are implications of cooperation? Is it always beneficial, or are there costs associated with it that lead to adverse effects?

4) Must cooperation be motivated *a priori* or can it emerge or evolve through complex social relationships? Is it meaningful to talk about *reactive cooperation* and if so, how?

5) Is cooperation a mentalistic or a behavioural notion? Can cooperation be described solely by examining the mental state of the participants (e.g. their beliefs, desires, etc.) or can it only be described by examining their actions (i.e. irrespective of their internal state) or does it require a mixture of the two?

6) What are the key mechanisms and structures which give rise to cooperation and allow it to be sustained? To what extent are these mechanisms and structures necessary? Is it possible to identify a necessary set of conditions for cooperation?

The range of answers to these questions were many and varied, however the main cause of such variety was the difference in starting points (see questions 1 and 2)—Franklin offered a broad and all encompassing definition, Norman offered a precise and constrained definition, and Doran was somewhere between the two. The other main points of discussion related to questions 4 and 5.

So starting with Franklin. He offers a typology of cooperation[b] (see figure 1)—placing it in the

---

context of multi-agent systems and then defining specific types of cooperation (question 2). A multi-agent system is *independent* if each agent pursues its own agenda (Franklin and Graesser, 1997) independently of the others. A multi-agent system is *discrete* if it is independent, and if the agendas of the agents bear no relation to one another. For example, one agent could filter email while another gathered information on the web. Discrete systems involve no cooperation. However agents can cooperate with no intention of doing so. The puck gathering robots of Beckers et al. (1994) form an independent system, each following the agenda of moving in a straight line till an obstacle is encountered and then backing off and going in another direction. Puck gathering is an emergent behaviour of the system in that, from an observer's viewpoint, the agents appear to be working together, but from the agent's viewpoint they are not. They are simply carrying out their own individual behaviour.
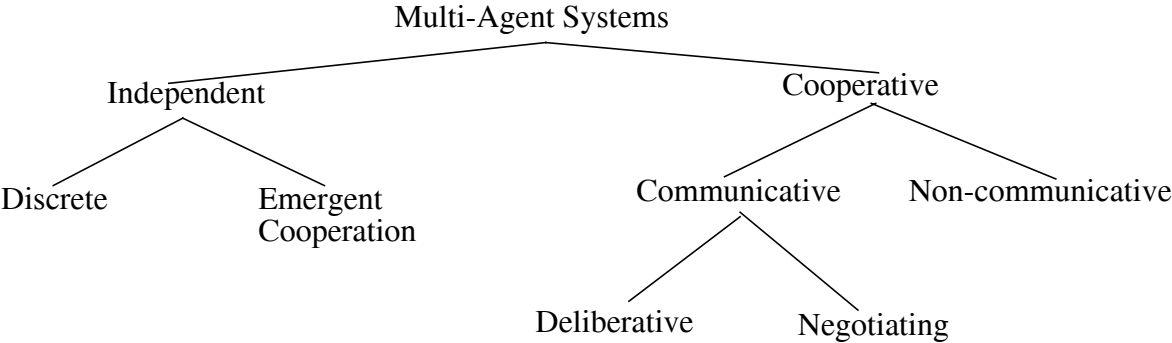

Figure 1: Cooperation Typology

The complement of independent systems are systems in which the agendas of the agents include cooperating with other agents in the system in some way (*cooperative* systems). Such cooperation can either be *communicative* in that the agents communicate (the intentional sending and receiving of signals) with each other in order to cooperate or it can be *non-communicative*. In the latter case, agents coordinate their cooperative activity by each observing and reacting to the behaviour of the other, as do lionesses on a hunt (Franklin, forthcoming). Intentional communication can take at least two forms—agents can deliberate or they can negotiate. In *deliberative* systems agents jointly plan their actions so as to cooperate with each other. Such cooperation may, or may not, entail coordination. *Negotiating* systems are like deliberative systems, except that they have an added dose of competition.

Doran offers an alternative (and more tightly defined) viewpoint on cooperation. He regards it more as a property of the actions of the agents involved (question 5). Thus, given a multiple agent system in which the individuals and the various sub-groups therein may be assigned one or more (possibly implicit) goals; cooperation occurs when the actions of each agent satisfy either or both of the following conditions:

    i) the agents have a (possibly implicit) goal in common, (which no agent could achieve in isolation) and their actions tend to achieve that goal.

    ii) the agents perform actions which enable or achieve not only their own goals, but also the goals of agents other than themselves.

[b.] He makes no claim that the classification is complete nor that the boundaries are sharp.

Note that this definition focuses on the actions of the agents and their goals, irrespective of how they arise. Thus, like Franklin's definition of emergent cooperation, Doran does not require that the agents deliberate and in some sense "intend" to cooperate for cooperative problem solving to occur. Also the definition does not require that the goals be explicit within the agents. Consider, for example, two robots jointly carrying a large object. This would be cooperation according to definition variant (i)—assuming that we regard each agent as having the goal of moving the object. If two robots are building separate towers of differently coloured bricks, then for a robot to pass a brick of the appropriate colour to another robot is an example of cooperation variant (ii).

However this behaviouristic view of cooperation is not the whole story. For agent designers it is important to know what leads to the choice of actions which result in cooperation. Doran identifies two important influencing factors: (i) what types of intra-agent processing lead to cooperative actions; and (ii) what is the basic decision principle to which agents adhere when deciding which cooperative actions to follow. Each is now considered in turn.

Patterns of cooperative actions can arise from very different intra-agent processes. Here these processes will be called cooperation generation processes (CGPs). Hence cooperation may be classified (question 2) by the types of intra-agent CGP which give rise to it:

- Reflex or reactive agents simply act. Therefore when cooperation occurs it does so without reflection upon possible actions. There is no prediction or predictive planning and therefore no intention (Steels, 1990). Hence cooperation is "emergent".

- Deliberative agents reflect upon the combinations of actions they and others might perform. They are given some means to choose, in principle, between different combinations of action, possibly after some process of negotiation, which lead to a convergence of their behaviour. For example possible plans may be passed around, modified, and selected as in Partial Global Planning (Durfee and Lesser, 1988; Decker and Lesser, 1995; Doran and Palmer, 1995)

- Concept based agents are designed to maintain an explicit concept of cooperation (i.e. an intra-agent knowledge/belief structure) which the agent uses to help select its actions and plans and which itself may be modified in the light of experience and also may be passed between agents The implication is that the agent possesses concept utilisation and manipulation processes which *a fortiori* apply to the concept of cooperation and therefore engender actual cooperation. Such an agent may decide to cooperate prior to any particular set of actions being considered.

Where cooperation is by deliberative or concept-based agents, we can also classify (question 2) cooperation by reference to the degree of altruism implicit in the CGPs.

- Purely self-interested support for cooperation. The agent deliberates, and selects only those (cooperative) actions which it believes will further its own goals (Franklin's negotiating agents).

- Partly or wholly altruistic support for cooperation. The agent deliberates, and selects cooperative actions which it believes to further a group's interests (typically a group of which it is a member) irrespective of its own particular goals (Franklin's deliberating agents).

An orthogonal question to the nature of CGPs relates to where the CGPs come from. Here there are a number of distinct sources:

- by explicit design—the creator of the agents deliberately designs the agent's behaviours or rules so that various instances of cooperation occur.

- by adaptation—the creator of the agents arranges that within the "lifetime" of individual agents a tendency to cooperate is developed or augmented (i.e. the group learns to cooperate).

- by evolution—the designer of the population of agents has arranged that CGPs are selected for in an evolutionary process.

Given these statements about what is involved in cooperation, and from where CGPs originate the next question addressed by Doran is what types of cooperation are appropriate in what types of environment (question 3). He argues that what is needed to address this question is a theory that, given the characteristics of an agent population and its goals, and of the population's environment, identifies: the optimal patterns of cooperation and the type of cooperation engendering processes the agents should best incorporate to achieve it. Such a theory should also identify the types of failure which can occur in cooperation and should predict when such failures will happen. An initial set of failure types includes:

i) no cooperation is possible (i.e. there are no combinations of actions available to the agents which satisfy the definition of cooperation).

ii) cooperation is possible (i.e. suitable actions can be performed by the agents) but the agents cannot collect the information needed to decide when to perform them.

iii) cooperation and the requisite information collection is possible, but the CGPs available to the agents are inadequate for the agents to select and execute it.

iv) cooperation is possible and executed by the agents, but there are side-effects which are not considered by the agents and which render it ineffective.

For example, consider the two tower building robots mentioned earlier. Each robot is trying to build a tower with blocks only of its "own" colour. Case (i) is illustrated by the situation where each robot cannot move the other's colour. Case (ii) occurs when the robots can move but not detect blocks of the other colour. Case (iii) occurs when the agents straightforwardly do not select actions which aid one another (for whatever reason). Case (iv) occurs when blocks are passed, but tower building ultimately fails because of a time overhead in "thinking about" helping one another.

Norman offers the most precise definition of the panellists. Many of the social phenomena labelled cooperation by Franklin and Doran would not be considered as such by Norman. Paraphrasing Longman's Dictionary, Norman argues that to cooperate is to act with another or others for a common purpose and for common benefit. But what does acting for a common purpose mean? An agent's purpose is what drives its behaviour, but there are two primary ways in which to give an agent a purpose. The agent may be provided with a set of behaviours that are designed in such a way that the agent pursues some purpose; in such a control system the purpose (or goal) is implicit in its design (a goal-oriented or purely behaviour-based control system). Alternatively, the agent may be motivated by explicit goals, possibly derived from

more basic motives (Norman, 1997). Such an agent may then employ planning and other decision-making processes to direct its action towards the achievement of these goals (a goal-directed control system[c]). Are both of these broad classes of agent capable of entering into a state of cooperation? Each agent must be individually motivated to act on the same purpose; this is possible with both goal oriented and goal-directed control systems. However, simply sharing the same purpose (implicit or explicit) is not sufficient for agents to cooperate (cf. (Conte and Castelfranchi, 1995)). Consider the example described by Bratman (1992) of two people painting a house. They share the same purpose of having the house painted, but if neither are aware of the other, or if they are aware but neither cares that the other is also painting the house, then they are not cooperating. Thus, there is more to cooperation than each participant having the same purpose, they must both intend to *act together* (a commitment to joint activity (Bratman, 1992; Jennings, 1993)). This is not possible without internal state, and so it is not possible to engineer a purely goal-oriented agent system in which the agents can participate in cooperative activity (as it is defined here)[d]. This contrasts with the position of both Franklin and Doran who do not require an intention to cooperate. Having offered this strong definition of cooperation, Norman considers a number of exemplar scenarios to illustrate what cooperation is and under what circumstances it may occur.

The airline example. Airline companies wish to maximise their profit through ticket sales and minimise the costs of ground services and flight windows, etc. They may increase ticket sales by offering flights to more destinations, but this requires investment in a presence at those new destinations. An alternative is to cooperate with another airline that provides services to these other destinations. Suppose that Air International provides services between London and other cities in Europe and North America, but not to other destinations within the UK, and GB Airlines provides services between London and other UK airports. These two companies may agree to cooperate in providing services from an airport that is served by only one of the companies to airports served by only the other. For example, between Paris and Edinburgh. The set of routes that these two cooperating companies can offer to customers is greater than the sum of their individual routes. With minimum investment, but through cooperation, both companies can benefit.

For these independent commercial organisations to cooperate, they must generate an explicit agreement to act for a common purpose and for common benefit. The motivation to cooperate is derived from their individual motivations to maximise profit (see Doran's basic decision criteria) while minimising their costs. Therefore, in certain cases cooperation is beneficial to both parties, but also detrimental to airlines that do not cooperate because of increased competition from other cooperating airlines. Suppose that Euro Jet is a company that offers flights from UK provincial airports to European cities including the route between Paris and Edinburgh. The cooperating Air International and GB Airlines are now in direct competition with Euro Jet, cooperation that will probably reduce the revenue that Euro Jet gains from flying this route. Successful cooperation can therefore be generated between autonomous agents that are not *a priori* cooperative through negotiating a mutually acceptable agreement to which they

---

[c.] Although a distinction has been made between behaviour-based and model-based control systems, a combination of the two may prove to be the most effective solution (Norman, 1994). Furthermore, the boundaries between the two may not be clear; behavioural competence may migrate between these layers as the agent adapts in its environment.

[d.] Norman argues that his position does not preclude the possibility of engineering a goal oriented agent system in which the agents are designed to coordinate their activities in the pursuit of implicit goals in an effective manner. All that is being said here is that these agents are not cooperating.

are both committed that describes how they are to act (i.e. a binding agreement of cooperative intent).

The driving example. Consider a lorry in the middle lane of a motorway moving to the inside lane as a car approaches from behind. To the driver of the car, the driver of the lorry is being cooperative. Although this is a possibility, the lorry driver may not be aware of the car approaching from behind; the lorry driver may simply be moving to the inside lane because his exit is approaching. The car driver interprets the lorry driver's behaviour as "being cooperative", but there is no true "cooperation" involved. Thus, to simply observe the external behaviour of a number of agents is not sufficient to determine whether or not those agents are cooperating (cf. Doran's position).

The contracting example. The third example concerns the generation of contractual agreements between agents for the provision of services (an example from the ADEPT project (Jennings et al., 1996)). Consider an agent that has agreed to design a network, but for this to be achieved the site at which the network is to be installed must be surveyed and the agent providing this service is not capable of performing the survey (Doran's first type of cooperation). It must therefore gain the agreement of an agent that does have this capability. Suppose that the service provider is aware of the consumer's goal; i.e. to have provided a network design. The agent providing this survey service is, in providing the service, acting in such a way that the consumer of the service is able to achieve its goal. However, the goal to have provided a network design is the goal of the consumer of the survey, not the provider of this service. Thus, these agents do not share the same purpose. They both benefit from the interaction (the consumer in having the survey, and the provider in having payment for this service), but they are not in a state of cooperative activity according to the position presented here.

In conclusion, given these three differing positions it is clear that definitions of basic social phenomena are important for multi-agent system research. Since all the panellists start from different suppositions, their answers to the remaining questions are radically different. While it is unlikely that universally acceptable definitions will emerge in the future, this panel highlights the need for precise formulations (MIKE: put links to other panel here) and consistent use of basic terms. The panellists feel that what may emerge in this field are definitions of a number of different types of cooperation, rather than one all embracing definition. In this case, only when all the definitions are taken together can the full range of phenomena which come under the umbrella of cooperative problem solving be defined. Once reasonable definitions and perspectives are in place, the next stage of the debate is to identify the types of mechanisms which need to be present within the agents to deliver desirable cooperation. Such mechanisms need to indicate when cooperation is appropriate, what forms of cooperation are appropriate in what circumstances, and how the agents should act in the given social context to benefit most from the potential of cooperative problem solving.

REFERENCES

Beckers, R., Holland, O. E. and Deneubourg J_L (1994). "From Local Actions to Global Tasks: Stigmergy in Collective Robotics," in R. Brooks and P. Maes eds. Artificial Life IV, Cambridge, Mass.: MIT Press.

M. E. Bratman. "Shared cooperative activity". The Philosophical Review, 101(2):327-341, 1992.

R. Conte and C. Castelfranchi (1995) "Cognitive and Social Action" UCL Press.

Decker K S and Lesser V R (1995) "Designing a Family of Coordination Algorithms" Proc. First International Conference on Multi-Agent Systems June 12-14th 1995, San Francisco, California Pps 73-80.

Doran J and Palmer M (1995) "The EOS Project: Integrating Two Models of Palaeolithic Social Change". In "Artificial Societies: the Computer Simulation of Social Life" (eds. N Gilbert and R Conte) London: UCL Press Pps 103-125.

Durfee E H and Lesser V R (1988) "Using Partial Global Plans to Coordinate Distributed Problem Solvers" In "Readings in Distributed Artificial Intelligence" (eds. A H Bond and L Gasser) Morgan Kaufmann Publishers: San Mateo California Pps. 285-293

Franklin, Stan and Graesser, Art (1997). "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents." Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, Springer-Verlag 1997.

S. Franklin (forthcoming) "Coordination without Communication" http://www.msci.memphis.edu/~franklin/coord.html.

N. R. Jennings. "Commitments and conventions: The foundation of coordination in multi-agent systems". The Knowledge Engineering Review, 8(3):223-250, 1993.

N. R. Jennings, P. Faratin, M. J. Johnson, T. J. Norman, P. O'Brien, and M. E. Wiegand. "Agent-based business process management". International Journal of Cooperative Information Systems, 5(2&3):105-130, 1996.

Norman, T. J. (1994) "Motivated Goal and Action Selection" In Working Notes of the AISB workshop: Models or Behaviours-Which Way Forward for Robotics? [University College London, Department of Computer Science Research Note 94/18].

T. J. Norman. "Motivation-based direction of planning attention in agents with goal autonomy". PhD thesis, Department of Computer Science, University College London, 1997.

Steels L. (1990) "Cooperation Between Distributed Agents through Self-Organisation" In "Decentralized AI" (eds Y Demazeau and J-P Müller) Elsevier Science Publishers B.V. (North-Holland) Pps 175-196.