

# Sociological Agents for Effective Social Action

Mark d’Inverno  
Cavendish School of Computer Science  
University of Westminster  
London, W1M 8JS, United Kingdom  
M.dInverno@wmin.ac.uk

Michael Luck  
Department of Computer Science  
University of Warwick  
Coventry, CV4 7AL, United Kingdom  
mikeluck@dcs.warwick.ac.uk

## Abstract

*This paper is concerned with the problem of how effective social interaction arises from individual social action and mind. The need to study the individual social mind, suggests a move towards the notion of sociological agents who can model their social environment as opposed to acting socially within it. This does not constrain such social behaviour; on the contrary, we argue that it provides the requisite information and understanding for such behaviour to be effective. We argue that effective social agents must be sociological in modelling agents and agent relationships. In this paper, we show how an existing agent framework leads naturally to the enumeration of a map of inter-agent relationships that can be modelled and exploited by sociological agents to enable more effective operation.*

## 1 Introduction

Over a number of years, the work of Castelfranchi and Conte has focussed on issues relating to the *social foundations* of multi-agent systems. For example, they point out that the problem of how to allocate tasks and resources and how to coordinate actions is typically raised only after a collective or social problem or goal is assumed [1]. One of the questions they explicitly ask in consideration of this is *how society is implemented in the minds of social agents*. There are many definitions of what constitutes a social agent, but the term is more often associated with social *activity* such as involved in the process of interaction. The need to study the *individual social mind*, however, suggests a move towards the notion of *sociological* agents who can *model* their social environment as distinct from acting socially within it.

## 2 Inter-Agent Relationships

Earlier work on a formal agent framework [2] allows an explicit and precise analysis of multi-agent systems with a

very basic set of conceptual primitives. In particular, when an agent uses another non-autonomous entity, the entity adopts or satisfies the agent’s goals and creates a specific kind of *social* relationship, known as an *engagement*. If this occurs through *direct* intervention, we refer to a *direct engagement*, to distinguish engagements in which there are intermediary agents. Once autonomous agents have generated goals and engaged other non-autonomous agents, they may, in turn, engage other non-autonomous entities with the purpose of achieving or pursuing the original goal. These *engagement chains* provide more information with which to analyse multi-agent systems than using engagements alone, since the flow of goal adoption is explicitly represented. They represent the goal and all the agents involved in the sequence of direct engagements. Since goals are grounded by motivations, the agent at the head of the chain must be autonomous. A *cooperation* describes a goal, the autonomous agent that generated the goal, and those autonomous agents who have adopted that goal from the generating agent.

These fundamental relationships underlie all multi-agent systems and provide a means for analysing the interdependence of agents in terms of the goals some agents achieve for others. Moreover, from this basic set of constructions, we can derive a detailed map of the relationships between individual agents for a better understanding of current social interdependence. In particular, different situations of interdependence each suggest different possibilities for interaction. We enumerate the various possible relations that might result in this view.

The direct engagement relationship specifies the situation in which there is a direct engagement for which the first agent is the client and the second agent is the server. In general, however, any agent involved in an engagement chain engages all those agents that appear subsequently in the chain. To distinguish engagements involving an intermediate agent we introduce the indirect engagement relation *indengages*; an agent *indirectly* engages another if it engages it, but does not *directly* engage it. If many agents directly engage the same entity, then no single agent has

complete control over it. It is important to understand *when* the behaviour of an engaged entity can be modified without any deleterious effect (such as when no other agent uses the entity for a *different* purpose). In this case we say that the agent *owns* the entity. An agent, *c*, owns another agent, *s*, if, for every sequence of server-agents in an engagement chain, *ec*, in which *s* appears, *c* precedes it, or *c* is the autonomous client-agent that initiated the chain. An agent *directly owns* another if it owns it and directly engages it. We can further distinguish the *uniquely owns* relation, which holds when an agent *directly* and *solely* owns another, and *specifically owns*, which holds when it owns it, and has only one goal. Finally, an agent, *b*, *cooperates* with agent, *a*, if and only if both agents are autonomous, and there is some cooperation in which *a* is the generating agent, and *b* is in the set of cooperating agents.

### 3 Towards Sociological Agents

The framework described above, together with the relationships arising from it, are suitable for reasoning both *about* entities in the world, and *with* entities in the world. That is to say that in addition to providing us with a way of understanding and analysing agents, agents themselves can also use the entity hierarchy as a basis for reasoning about other agents within their environment. Our work enables an analysis that can reveal opportunities to exploit existing relationships and how to minimise effort and avoid conflict.

To illustrate this greater reasoning capacity of sociological agents, we describe below some examples of categories of goals, agents and plans (with respect to the models of the sociological agent), that may be relevant to an agent's understanding of its environment.

- A *self-sufficient plan* is any plan that involves only neutral-objects, server-agents the agent owns, and the agent itself. Self-sufficient plans can therefore be executed without regard to other agents, and exploit current agent relationships.
- A *self-sufficient goal* is any goal in the goal library that has an associated self-sufficient plan for achieving that goal. These goals can then, according to the agent's model, be achieved independently of the existing social configuration.
- A *reliant-goal* is any goal that has a non-empty set of associated plans for achieving that goal, each of which is not a self-sufficient plan.

For each plan that is not self-sufficient, a sociological agent can establish the autonomous agents that may be affected by its execution, which is an important criterion in selecting a plan from competing active plans. An autonomous agent *A* may be affected by a plan in one of two ways: either it is required to perform an action directly, or it is engaging a server-agent *S* required by the plan. In this latter case,

a sociological agent can reason about either persuading *A* to share or release *S*, taking *S* without permission, or finding an alternative server-agent or plan. To facilitate such an analysis, we consider the following categories of agents and plans.

- The *cooperating autonomous agents* of a plan are those autonomous agents, other than the agent itself, that are involved in performing actions of that plan. They will need to cooperate with the agent for the plan to be executed.
- The *affected autonomous agents* of a plan are those autonomous agents, other than the agent itself, that are engaging an entity required in the plan.
- The *least-direct-fuss plans* for any reliant-goal are those plans that require the fewest cooperating agents.
- The *least-fuss plans* are those requiring the fewest affected autonomous agents.

### 4 Conclusions

Castelfranchi's perspective of sociality arising not through collective action, but through *individual social action and mind*, lies at the heart of the challenge taken up in this paper. In it we try to show how effective social action arises naturally from the way in which a *sociological agent* can model its environment. Though such modelling is not new, the relationships that are subject to the modelling, and the way in which such modelling impacts on the world (and agents) around it, offers a way to understand and develop effective social behaviour.

The analysis is broadly applicable, but the particular value can be seen, for example, in applications that embody critical notions of dependence among agents, especially Social Power Theory and its computational counterpart, Social Dependence Networks(SDN). Our work provides two key benefits: first, it addresses some of the weaknesses in the SDN model in relation to the ambiguity of some constructs such as the nature of an *owned resource*, which we have clarified and tightened; second, it balances that earlier work which focussed on the problem *situations*, such as dependencies, by considering configurations of *solutions* to minimise effort and take advantage of opportunities in dealing with dependencies (through self-sufficient or least-fuss plans, for example).

### References

- [1] C. Castelfranchi and R. Conte. Distributed artificial intelligence and social science: Critical issues. In G. M. P. O'Hare and N. R. Jennings, editors, *Foundations of Distributed Artificial Intelligence*, pages 527–542. Wiley, 1996.
- [2] M. Luck and M. d'Inverno. A formal framework for agency and autonomy. In *Proceedings of ICMAS'95*, pages 254–260. AAAI Press / MIT Press, 1995.