

A Structural Functionalist Conception of Artificial Agent Societies

Sascha Ossowski

School of Engineering, Rey Juan Carlos University,
Camino de Humanes 63, 28936 Móstoles (Madrid), Spain
S.Ossowski@escet.urjc.es

Abstract. This paper reports a case of the successful application of sociologic theories in the area of Distributed Artificial Intelligence. We argue in favour of a rather deep conceptual transfer of sociologic models: it is shown how *analytical* theories from the school of structural functionalism, aimed at explaining observed phenomena in *human* societies, can be put “upside down”, so as to come up with a *constructive* co-ordination mechanism for *artificial* agent societies. Setting out from such an engineering perspective, we develop a formal model of this mechanism, sketch its operationalisation and illustrate its dynamics by an example.

1. Introduction

Ever since the emergence of the area of Distributed Artificial Intelligence (DAI) in the early eighties, reference of work in Sociology has been abundant. There are many different motivations for research at the edge between both disciplines. In this paper, we are approaching the field from an *engineering* perspective: our primary interest lies in using sociological theories as a source of models and metaphors that help us to *design* and implement *systems* which serve a purpose in the real world. Although computer simulation of social phenomena is a quite exiting area of research (Gilbert and Conte, 1995), it is of subordinate importance for us.

Our work sets out from perceived lack of “social plausibility” in the way that the problem of *co-ordination* among artificial agents has been treated within mainstream DAI so far. There are essentially two classes of approaches:

- In *centralised* co-ordination architectures, one individual agent orchestrates group behaviour respecting a certain set of goals within a group of agents (e.g. von Martial, 1992). It could be considered an *individual intelligence* that tries to make use of social resources.
- *Distributed* architectures comprise several agents that intend to shape group co-ordination. However, this is usually done from the viewpoint of a *replication* of an individual intelligence: agents develop joint goals and pursue joint intentions, which constitute (potentially different views of) the same multiagent plan *iterated* in the interdependent agents’ minds (e.g. Decker and Lesser, 1995). The problem lies in how to deal with uncertainty, i.e. when to exchange individual plans so as to merge them into one.

Both approaches rely on the “classical” AI stance on co-ordination. In the first, *one* sole agent co-ordinates the plans of others, while in the second *many* agents do the same. We propose to consider models of *emergent* co-ordination based on metaphors and mechanisms found in human societies, or *social* co-ordination for short. A truly social co-ordination mechanism needs to take into account that in social life, people are neither entirely manipulated nor fully conscious of the reasons and the effects of their actions, because concepts that do not get a hold of reality will have an undesirable impact on the design of artificial systems (Conte and Castelfranchi, 1995).

This paper presents a social co-ordination mechanism called *structural co-operation* as one step in this direction. It is organised as follows: first, we use findings in Social Science to motivate the development of structural co-operation; it is shown how observed social phenomena can be explained in terms of their social finality, what role social structure plays in such rationalisations, and how such a model can be useful from an engineering perspective. Subsequently, in section 3, an operational model of structural co-operation for a specific class of domains is presented, which is illustrated by an example in section 4. Concluding remarks can be found in section 5.

2. The metaphor: functionality of social structure

This section aims at motivating the conceptual framework within which our mechanism will be developed. It is subdivided in two parts. We will first make a short detour to the area of social science and examine the type of analysis that a particular school, structural functionalism, applies in order to explain observed characteristics of human societies. In the second part, we analyse as to how far structural functionalist explanations of social phenomena can provide us with a conceptual and logical framework for constructing social co-ordination mechanisms for multiagent systems.

2.1 Structural Functionalism

The conceptual basis of research in multiagent systems has been largely influenced by economy. Terms such as negotiation, contract, marginal costs etc. are in heavy use within the MAS community. All these metaphors are borrowed from a discipline that describes a developed society where the division of labour is predominant and where the exchange of “goods” is realised within the frame of some monetary system. As the sociologist Durkheim noticed around the turn of the century, the behaviour of the agents that comprise such complex social systems is mainly determined by individuality; the influence of socialisation on the self-interested behaviour of individuals decreases with the complexity. In contrast, within simpler societies the socially conditioned sphere of individual consciousness is predominant. In the sequel, we will examine as to how far this argument, and the study of “primitive” societies in general, can provide ideas for the design of co-ordination mechanisms within multiagent problem-solving systems.

Up to the early days of our century sociology, and social anthropology in particular, has been approached on the basis of the paradigm of evolutionism. Social phenomena were often analysed in rather simplistic schemes, as Darwin’s ideas respecting biological evolution were rigidly transferred to human societies. Habits and customs were deterrred from their cultural context to be compared indiscriminately, and often conjectures about the societies’ histories were presented in order to “explain” them by their supposed genesis.

Structural Functionalism became popular especially in the United Kingdom as a response to conjectural reconstructions of evolutionists. Radcliffe-Brown, one of the major defenders of this school, conceived social anthropology as a natural science that applies inductive methods. Large amounts of data were gathered in field-works and common structural principles were abstracted that allowed for a classification of societies (Kuper 1973). The point lies in a shift from diachronic to synchronic explanations of phenomena; the reasons for certain observed patterns of behaviour are sought in the present, not in the past: instead of speculations about past origins which are necessarily opaque to the scientist it is asked for the “purpose” or the “meaning” of a specific custom in the contemporary situation. In structural functionalism this type of explanation is organised around three basic concepts: *process* and *structure*, whose relation is referred to as *function*.

The object of anthropological study is identified as the *process* of social life. This concept comprises the multitude of actions and interactions of human beings acting within a particular group and in a particular place. It constitutes the concrete, observable phenomenological reality.

Social structure is seen “an arrangement of persons in institutionally controlled or defined relationships” (Radcliffe-Brown 1952). Social relations arise either from person to person (father and son) or between social roles (king and subjects). It exists “between two or more individual organisms when there is some adjustment of their respective interests, by convergence of interest, or by limitation of conflict that might arise from divergence of interests” (p. 199). Any relationship “is one in which the conduct of persons in their interactions is controlled by norms, rules or patterns” (p. 10). The established moral codes, norms of conduct, the rights and the duties, pertaining to particular social structure are referred to as *institutions*.

Function is conceived as the “functioning of its structure”. The function of a particular set of norm is the contribution it makes to the continuity of society and the maintenance of its structure: normative standardised behaviour gives rise to recurrent social processes whose effects correspond to the conditions of persistence (or the “needs”) of society. The basic need of all societies has been termed *co-aptation* by structural functionalists, meaning the mutual adjustment of members of society: morals, etiquette, religion, government, and education are all parts of the complex mechanism by which a social structure exists and persists.

Within the above conceptual framework observed phenomena are explained through a *functional* analysis. It sets out from the assumption that all parts of a social structure work together with a sufficient degree of harmony or internal consistency, that is, without producing persistent conflicts: the social structure as a whole is considered *a priori* to be “functioning”, i.e. to be apt to meet the necessary conditions for the existence and the reproduction (or “survival”) of society. On this background a particular institution is analysed with respect to its instrumentality: the structural functionalist aims to show how it fits in the network of different interests and conflicts that exist within a society, so as to contribute to an *equilibrium* of force which leads to a stable social structure. Social processes, a religious ritual for instance, are not explained by their evolution, but by their “objectives” or functions, which, in the end, are always assumed to promote the maintenance of a social equilibrium.

Although the main interest lies in the description and comparison of social structure, it is acknowledged that norms and institutions “work or ‘function’ only through their effects in life, i.e. in the thoughts, sentiments and actions of individuals” (Radcliffe-Brown 1952, p. 185). Social institutions are characterised by being external to individuals and coercive in nature and are thus an instance of Durkheim’s notion of *social facts*: they exist before an individual’s birth and endure after his death; the individual cannot freely choose to adopt them but just has to live in a society where a set of norms is in force (i.e. being generally recognised as binding and widely observed). So, society and its structure are seen as influencing individual behaviour by creating external constraints: norms and values, rules of descent and residence are just social facts, that contribute to maintain the society in equilibrium, that is, to ensure its survival.

It is accepted that the explanations, that individuals in society provide for their own behaviour and for their abidance to social norms, are different from structural functionalist interpretations, as such “apparent” functions of social structure are believed to be analytically insignificant. The interest of the structural functionalist lies in the *latent* functions, which the individual ignores or is unaware of. These can only be analysed from the outside, taking a holistic stance: it is argued that the function of a norm can only be explained in its relation to the social structure and the external conditions of society, i.e. by taking the entire system into consideration.

Structural functionalism within social anthropology was strongly tied to the study of indigenous societies during the context of colonialism. The relative rigidity of these societies in colonial regions gave the (false) impression of harmonical, internally consistent social structures with stable social equilibria. So, it is not surprising that the decline of colonialism after the Second World War

accelerated the transition from structural functionalism to more diachronic, process-oriented theories which underline the dynamic aspects of human society.¹

2.2 A constructive perspective

At first glance, the above excursus to social anthropology, which aims to build an analytical model of phenomena observed in reality, appears to have pretty little relation to our attempt to engineer intelligent computer systems that efficiently solve problems. Still, the key problem of applying the paradigm of multiagent systems to problem-solving is to coerce autonomous agents to mutually adjust their behaviour so as to achieve a certain task. There is a growing awareness in the DAI community that “pure individualism”, as reflected by the principle of individual rationality, is not a good choice as a guideline for decision-making of autonomous agents that make up those systems (Jennings and Campos, 1997).

Structural functionalism provides a model of how society, or social structure, “moulds” individual behaviour so as to achieve globally functional behaviour. Still, in order to use this model as a metaphor in the design of multiagent systems, we need to turn structural functionalist analysis upside down: instead of explaining observed behaviour in terms of social structure, we would like to *design* a “social structure”, so as to achieve the “co-aptation” of the behaviour of autonomous agents towards a “social equilibrium” that leads to a satisfying achievement of a global task. There are two ideas in structural functionalism that are of special interest for this enterprise:

- the *biasing* effect of social structure
norms and social institutions impose external constraints on an agent’s freedom of choice. They are primitive notions, not reducible to any other, that actually *influence* the behaviour of a “socialised” autonomous agent. Still, they are not assumed to completely determine agent behaviour, but just bias self-interested action in a certain (“socially desired”) direction.
- the *teleological* character of social structure
the social function, that social structure provides, gives an external observer the impression that society is a subject that deliberately pursues some (collective) goal. This does not require that agents are “aware” of the functionality of this prescriptive apparatus, i.e. they need not reason about *why* some norm exists and prescribes a certain pattern of behaviour. Still, when abiding to norms they behave *as if* they deliberately intend to bring about a certain social function.

Seen from the outside, social structure induces the *co-ordination* of the behaviour of autonomous agents towards a global function of society. From an agent’s perspective social structure induces a specific form of *co-operation* in the sense of mutual adjustment of individual action.

We will term *structural co-operation* such a mechanism that has the co-ordination of autonomous agents emerge from social structure. Figure 1 graphically depicts the notion of the functionality of social structure, that structural co-operation is based on, from a constructive perspective. It sets out from autonomous agents that pursue their individual goals upon the background of a common social structure. The bold arrows indicate social interaction processes that the agents are involved in and that influence the characteristics and the achievement of their goals. Thin arrows denote functional effects. A chain of *functional effects* can be recognised: norms influence social interactions, which modify the set of individual goals of each agent so that its actions become instrumental with respect to the collective function. It is the task of the multiagent system designer to establish a social structure that induces the co-ordination of agents with respect to the problem that the system is to solve. He has to develop a “legislation” or a “code of moral behaviour” so as to make the collective function coincide with the desired functionality of the system.

¹ In fact, an essentially synchronic analysis of human societies in the tradition of structural functionalism is strongly discouraged. It must be underlined that we are concerned with *artificial* agent societies: we are seeking guidelines for constructing computational mechanisms to be applied within an *engineering* process.

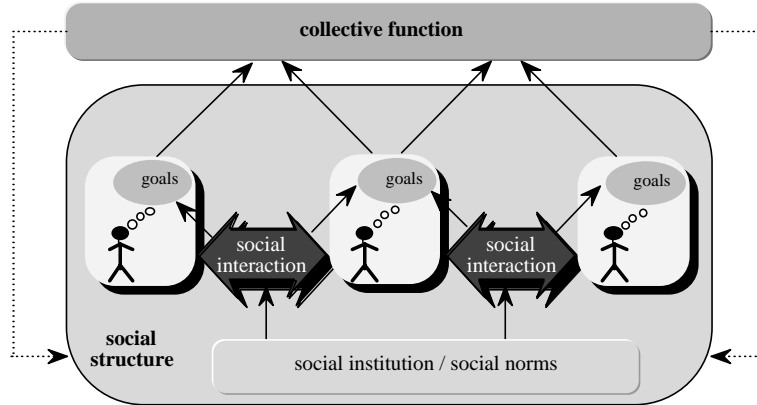


Figure 1. Emergent co-ordination through structural co-operation

3. The model: social structure in artificial agent societies

In order to make these quite abstract ideas more precise and to turn structural co-operation into an operational mechanism within artificial agent societies, two key questions must be tackled: we need to model self-interested action of autonomous agents in a multiagent world, and how social structure influences this process. In this section we first outline the class of domains that we are interested in: Distributed Problem-solving (DPS) societies of autonomous, cognitive agents that perform reactive planning. We then model structural co-operation within bargaining theory, and finally sketch a distributed algorithm that operationalises this mechanism.

3.1 The problem: reactive social co-ordination

Many real-world domains are highly dynamic: perceptions are error-prone, actions fail, contingencies occur. A common way to deal with this problem is to build systems that only plan their actions for a short-time horizon, in order to assess the effects of their interventions as early as possible, and to adapt future behaviour accordingly (Cuenca and Ossowski, 1998). When such systems are modelled on the basis of a multiagent architecture, two essential constraints have to be taken into account: first, agents need to cope with the fact that their plans and actions interfere because they share an environment with only limited resources; second, agents should be prepared to consider actions that attain their goals only *partially* due to resource limitation and environmental contingencies. In the sequel we formalise essential features of this type of problems.

Let S be a set of *world states* and Π a finite set of *plans*. The execution of a plan π changes the state of the world which is modelled as a partially defined mapping

$$res: \Pi \times S \rightarrow S.$$

A plan is *executable* in s , if only if res is defined for a certain world state s , fact which we express formally by the predicate $exec(\pi, s)$. At least one *empty plan* π_e is required to be included in the set of plans Π ; it is modelled as identity.

There is a set of agents A , each of which can act in the world thereby modifying its state. An agent $\alpha \in A$ is characterised by the following notions:

- a predicate $can(\alpha, \pi)$, determining the *individual plans* $\pi \in \Pi$ that α is able to execute. An agent α is always capable of executing the empty plan π_e ;

- a predicate $ideal(\alpha, s)$, expressing the states $s \in S$ that the agent $\alpha \in A$ would ideally like to bring about;
- a metric function d_α , which maps two states to a real number, representing agent α 's estimation of "how far" one state is away from another. It usually models the notion of (relative) "difficulty" to bring about changes between world states.

In the scenarios that we are interested in, an agent usually cannot fully reach an ideal state. So, we will use the notion of ideal states together with the distance measure d_α to describe an agent's preferences respecting world states. Note that the agents in A may have different (partially conflicting) ideal states and may even measure the distance between states in different scales.

We now introduce a notion of interdependent action. The set M of *multiplans* comprises all multisets over the individual plans Π , i.e. $M = bagof(\Pi)$. A multiplan $\mu \in M$ models the simultaneous execution of all its component plans, i.e. of the individual plans $\pi \in \mu$ that are contained in the multiset μ . The commutative operator \circ denotes multiset union and hence states that its operands are executed together². By identifying an individual plan with a multiplan that contains it as its only element, the partial function res is extended to multiplans:

$$res : M \times S \rightarrow S .$$

The function res is undefined for a multiplan μ and a state s , if some of the individual plans that it contains are *incompatible*, i.e. in case that in a state s of a modelled domain it is impossible to execute them simultaneously. Otherwise, μ is said to be *executable* in s (formally: $exec(\mu, s)$). The empty plan π_ϵ is compatible with every multiplan and does not affect its outcome.

The notion of capability for executing a multiplan is also a natural extension of the single agent case. We define the set of groups Γ as the powerset of the set of agents, i.e. $\Gamma = \wp(A)$. A group $\gamma \in \Gamma$ is capable of executing a multiplan μ , if there is an assignment such that every agent is to execute exactly one individual plan and this agent is capable of doing so, i.e. there is a bijective mapping ψ from individual plans to agents, such that

$$can(\gamma, \mu) \equiv \forall \pi \in \mu. can(\psi(\pi), \pi) .$$

Definition 1. A co-ordination setting D is defined by the sets of individuals S , Π , A , M and Γ , the functions res , d_α and \circ as well as the predicates $exec$, can and $ideal$.

3.2 The mechanism: structural co-operation

The outcome of a co-ordination process within a co-ordination setting D can be conceived as a multiplan. When it represents the result of social co-ordination, the multiplan reflects the self-interested choice of all agents, i.e. agents do *not* refer to a notion of joint utility and do *not* care for whether the multiplan contributes to the overall functionality of the DPS system or not. Structural co-operation provides a model of such self-interested action, and introduces a mechanism to bias the outcome of co-ordination among social DPS agents³. The designer is supposed to use that mechanism so as to make the outcome of social co-ordination instrumental with respect to the desired functionality of the DPS system (Ossowski et al., 1996).

In the sequel, we will sketch both, a model of multiagent rational action as well as a biasing mechanism. Subsequently, we model these ideas within bargaining theory. Finally, we show how the outcome of social co-ordination in a co-ordination setting D can be determined and computed.

² In the sequel we will use a set of cardinality one and its only element indiscriminately. So, for instance, we write $\mu = \pi \circ \pi' = \pi' \circ \pi = \{\pi, \pi'\}$ and $\mu \circ \pi'' = \pi'' \circ \mu = \{\pi, \pi', \pi''\}$

³ As will be outlined below, we require a "social" agent to be norm-abiding.

3.2.1 Social Agents

In the above co-ordination setting, the need for co-ordination is expressed by the fact that, when an individual plan π is executed in conjunction with a multiplan μ , the resulting world state changes (i.e. $res(\pi, s) \neq res(\pi \circ \mu, s)$). When an agent α is able to execute π (i.e. $can(\alpha, \pi)$) and a group γ is able to execute μ (i.e. $can(\gamma, \mu)$), then the group γ has the power to influence α : the agent *depends* on γ with respect to π . There are different types and degrees of dependence. The strongest degree is *feasibility* dependence, in the presence of which a group γ can turn down or enable an agent's *possibility* to execute its individual plan π (in the frame of a multiplan μ) (Ossowski and García-Serrano, 1997).

We model self-interested choice within a co-ordination setting on the basis of this notion of dependence: the less an agent *depends* on the choices of others with respect to the outcome of its plans, the better is its position in society. And the better an agent's position in the agent society, the more weight will have its preferences in the outcome of social co-ordination; if an agreement is reached, it will be biased towards that agent.

In consequence, the outcome of social co-ordination is determined by the network of social dependence relations. We now introduce the notion of *normative prescription*: if in a situation s it is forbidden for a group of agents γ to enact a multiplan μ we write

$$forbidden_s(\gamma, \mu)$$

Our *social* agents are norm-abiding: they do not even consider executing plans that are forbidden for them. So, we introduce the notion of *preparedness* as capability plus the absence of such prohibitions

$$prep_s(\gamma, \mu) \Leftrightarrow can(\gamma, \mu) \wedge \neg forbidden_s(\gamma, \mu)$$

Reconsidering the above model of self-interested action on the basis of the notion of preparedness instead of capability, it becomes clear that by issuing prohibitions, a designer can modify the social dependence structure and, in consequence, also the outcome of social co-ordination. These ideas will be formalised in the sequel, taking into account that prescriptions are conceived to worsen the position of the involved agents, potentially improving the position of the remaining agents.

3.2.2 Social Co-ordination as a Bargaining Scenario

In a first step, a quantitative notion of preference over *agreements* is introduced. Agent α 's preference for a world state s is expressed by its distance to some ideal state, which can be written as

$$|s|_\alpha = \min\{d_\alpha(s, \bar{s}) \mid ideal(\alpha, \bar{s})\}.$$

Furthermore, we introduce the set X of *legally enactable* multiplans in a situation s . It comprises all plans that are executable in s and for which there is a group of agents prepared to do so:

$$X = \{\mu \in M \mid exec(\mu, s) \wedge \exists \gamma \in \Gamma. prep_s(\gamma, \mu)\}$$

On this basis we can define a quantitative preference over multiplans.

Definition 2. The *utility* for an agent α of a legally enactable multiplan $\mu \in X$ is given by

$$U_i(\mu) = |s|_\alpha - |res(\mu, s)|_\alpha.$$

The utilities that each agent obtains from a multiplan can be comprised in a vector. The set of utility vectors that are realisable over X is denoted by $U(X)$.

When agents have different points of view respecting which multiplan to agree upon, they may “flip a coin” in order to choose between alternative agreements. A probability distribution over the set of legally enactable multiplans is called a *mixed multiplan*. Let m be the cardinality of X , then a mixed multiplan is a m -dimensional vector

$$\sigma = (p_1, \dots, p_m), 0 \leq p_i \leq 1, \sum_{i=1}^m p_i = 1 .$$

The set of mixed multiplans is denoted by Σ . The *expected* utility of a mixed multiplan is given by the sum of each legally enactable multiplan’s utility weighed by its probability:

Definition 3. The utility for an agent α_i of a mixed multiplan $\sigma \in \Sigma$ is given by

$$U_i(\sigma) = \sum_{k=1}^m p_k U_i(\mu_k) .$$

The set of expected utility vectors that are realisable over Σ is denoted by $U(\Sigma)$.

When agents co-ordinate their strategies and agree on some mixed multiplan, the corresponding vector of utilities is what each agent expects to obtain. Still, agents are autonomous and not forced to co-operate. So, it remains to model what happens in case of conflict.

In a conflict situation we define the *response* of the set of agents γ to a single agent α ’s plan $\pi \circ \mu$, i.e.

$$response_s(\pi, \alpha_i, \mu, \gamma) \Leftrightarrow \mu = \min_{U_i(\pi \circ \mu')} \{ \mu' \in X \mid prep_s(\mu', \gamma) \} .$$

This models that in case of disagreement an agent must account for the unpleasant situation that all its acquaintances jointly try to harm it. As the possibility of reaching an incompatible multiplan has to be excluded, α can only choose from the set $FEAS_s(\alpha)$ of plans that are feasible regardless what others do. The empty plan π_ϵ is contained in $FEAS_s(\alpha)$ by definition (Ossowski, 1997). Agent α will choose the plan π out of $FEAS_s(\alpha)$, that maximises its individual utility value when combined with the response of its acquaintances.

Definition 4. The *conflict utility* of the agent α is

$$U_i^d = \max \{ U_i(\pi \circ \mu) \in \mathbb{R} \mid \pi \in FEAS_s(\alpha_i) \wedge response_s(\pi, \alpha_i, \mu, \gamma) \} .$$

We now outline how a bargaining scenario can be defined on the basis of the above notions. For this purpose, we define the overall conflict utility as

$$\vec{d} = (U_1^d, \dots, U_n^d) ,$$

and treat the conflict utility vector as an effectively reachable agreement, defining a set S to be the convex and closed hull (*cch*) of the legally enactable multiplans plus the conflict utility vector

$$S = cch(U(X) \cup \{ \vec{d} \}) .$$

The set S usually equals $U(\Sigma)$, but may also be a (convex) superset of the latter.

Definition 5. The *bargaining scenario* B associated with a social co-ordination problem is a pair

$$B = (S, \vec{d})$$

S is called the *bargaining set* and \vec{d} the *disagreement point*. B complies with the formal properties of bargaining models, so the whole mathematical apparatus of bargaining theory becomes applicable (Thompson, 1994).

3.2.3 The Outcome of Social Co-ordination

In this section we rely on Bargaining Theory to find a solution to the associated bargaining scenario (S, \vec{d}) : a vector $\vec{\varphi} \in S$ needs to be singled out upon which a bargaining process – and the social co-ordination that it models – is supposed to converge. Strategic bargaining theory relies on

a sequential setting where agents alternate in making offers to each other in a pre-specified order and eventually converge on an agreement. By contrast, the axiomatic models of bargaining that we will use in the sequel first postulate desirable properties of a bargaining solution, and then seek the solution concept that satisfies them.

The five requirements of *individual rationality*, *Pareto-optimality*, *symmetry*, *scale invariance* and *contraction independence* that Nash bargaining models state for “fair” solutions to a bargaining scenario (Thompson, 1994), provide an adequate model for our purposes.

Theorem 1 (due to Nash). A utility vector $\bar{\varphi}$, that complies with the above axioms, maximises the function

$$N(\bar{x}) = \prod_{i=1}^n (x_i - d_i)$$

So, a solution $\bar{\varphi}$ maximises the *product* of gains from the disagreement point. It always exists and is unique (Thompson, 1994).

3.3 The Implementation: Computing the Outcome

We are now endowed with a characterisation of the *outcome* of social co-ordination. So, there is no need to explicitly “simulate” the co-ordination process among norm-abiding autonomous agents. Instead, we use a distributed multistage algorithm that *directly* computes the solution (Ossowski 1997):

- stage 1 performs an asynchronous distributed search for Pareto-optimal multiplans;
- stage 2 determines the *mixed* multiplan that constitutes the solution to the associated bargaining scenario;
- finally, in stage 3 a specific multiplan is chosen by means of a lottery and the corresponding individual plans are assigned to the agents.

Note that although this algorithm requires rather benevolent agent behaviour, we can assure that its *outcome* corresponds to the result of social co-ordination among autonomous agents.

4. An example

We will illustrate the above model by a scenario within the *synchronous blocks domain* (Ossowski 1997), which is an extension of the well known blocks world. There is a table of unlimited size and a fixed amount of numbered blocks. Blocks can be placed either directly on the table or on top of another block. The only operation that can be performed is to place a block x on top of some block y (formally: $move(x,y)$), which requires x and y to be clear. There is a clock that marks each instant of time by a tic. A *plan* of length k is a sequence of k operations performed successively at tics. Instead of an operation, a plan may contain a NOP, indicating that nothing is done at a certain tic.

We define the result of a multiplan μ to be the “sum” of the effects of its component plans. Still, if the component plans “interact” the following rules apply:

- a multiplan μ is *not* executable if some component plans *access* the same block at one tic in different ways, or if they *obstruct* a block that another component plan uses at a later tic;
- a multiplan μ is executable despite of a non-executable component plan, if other component plans “complement” it, e.g. by providing a missing action;

Suppose a scenario as depicted in Figure 2: it shows the configuration of the four blocks of our example in the initial situation. There are two agents α_1 and α_2 , each of them capable of executing plans of length 2. Still, while the former can move all blocks but block 4, the latter is capable of

manipulating all blocks but block 3. The agents' ideal states corresponds to the configurations of blocks shown in Figure 2. Both of them measure the distance between two states s_1 and s_2 by the length of the shortest plan that transforms s_1 into s_2 .

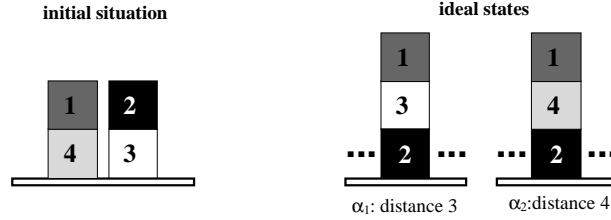


Figure 2. A scenario in the synchronous blocks domain

Table 1 shows some plans in this scenario, the capability of the agents to enact them and their executability in the initial situation s .

Plan	Operations	$can_i(\alpha, \pi)$	$exec(\pi, s)$
π_1	$[move(2, table), move(3, 2)]$	α_1	true
π_3	$[move(2, table), NOP]$	α_1, α_2	true
π_4	$[move(1, table), NOP]$	α_1, α_2	true
π_9	$[move(2, table), move(4, 2)]$	α_2	false
π_{10}	$[move(1, 2), move(4, 1)]$	α_2	true
π_{11}	$[move(2, 1), move(3, 2)]$	α_1	true
π_ϵ	$[NOP, NOP]$	α_1, α_2	true

Table 1. Some individual plans

Stage 1 of the distributed asynchronous algorithm sketched previously determines that there are three Pareto-optimal (pure) multiplans (π_1, π_4) , (π_4, π_9) and (π_3, π_4) . Their execution leads to the world states depicted in Figure 3, which comes down to utility vectors of (2,1), (0,3) and (1,2) respectively.

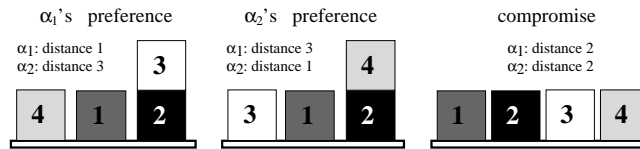


Figure 3. Efficient outcomes of the example scenario

Stage 2 now calculates the solution in Pareto-optimal mixed multiplans, which are obtained by randomising among (π_1, π_4) , (π_4, π_9) and (π_3, π_4) . First, the disagreement point is to be determined. The only plan that α_1 can execute and which is guaranteed not to become incompatible is π_ϵ , which α_2 counters by π_{10} , resulting in a conflict utility of -2 for α_1 . Agent α_2 also needs to choose π_ϵ in case of disagreement, to which α_1 's most malicious response is to enact π_{11} , giving rise to a conflict utility of -1 for α_2 , i.e.

$$\vec{d} = (-2, -1).$$

In accordance with Theorem 1 this leads to a solution vector of

$$\vec{\phi} = (1, 2).$$

Consequently, the outcome of stage two is to choose the "compromise" multiplan (π_3, π_4) with probability 1, so that in stage three just π_3 is assigned to α_1 and π_4 is assigned to α_2 .

Suppose now a normative structure in which it is *forbidden* for all agents to put a block 1 on block 2. As a result, agent α_2 is no longer *prepared* to execute plan π_{10} and it is *permitted* for α_1 to

enact π_4 . So, the worst situation that α_2 can bring about in the eyes of α_1 is to put the light grey block 4 on top of block 2. The conflict utility of α_2 remains unchanged so that the disagreement point changes in favour of α_1 to

$$\vec{d} = (-1, -1).$$

Stage two of the algorithm keeps adhering to Theorem 1, and because of the change in \vec{d} now produces

$$\vec{\varphi} = (1\frac{1}{2}, 1\frac{1}{2}),$$

which is reached by randomising equally between “compromise” (π_3, π_4) and α_1 ’s favourite (π_1, π_4) . Stage 3 tosses an equally weighed coin, selects one of the above multiplans accordingly, and assigns the corresponding component plans to α_1 and α_2 .

Imagine now another normative structure where it is *permitted* for agent α_2 to unstack block 1 at the first tic (e.g. by executing plan π_2). So, all plans that manipulate the left stack in the first step are *forbidden* for α_1 . Consequently, the worst thing that α_1 can still do is to obstruct block 2 for α_2 by putting block 4 on it in the second step. Now the disagreement point moves towards α_2 :

$$\vec{d} = (-2, 0).$$

The solution utility vector computed in stage two becomes

$$\vec{\varphi} = (\frac{1}{2}, 2\frac{1}{2}).$$

This is reached by selecting α_2 ’s favourite (π_4, π_6) and compromise (π_3, π_4) both with probability $p = \frac{1}{2}$. Again, in stage 3 a coin is tossed and the winning multiplan enacted.

5. Discussion

In this paper we have shown how conceptual and logical aspects of sociologic models can be applied to the area of Distributed Artificial Intelligence. We have put “upside down” *analytical* theories from the school of structural functionalism, aimed at explaining observed phenomena in *human* societies, so as to come up with the *constructive* mechanism of structural co-operation for *artificial* agent societies. Setting out from such an engineering perspective, structural co-operation has been formalised, and a distributed algorithm to compute the outcome of the co-ordination process has been sketched.

In a complementary line of research, we have developed an agent architecture, that provides operational support for Societies of Autonomous Problem-solving Agents. On this basis, the mechanism of structural co-operation has been applied to real world domains (Ossowski and García Serrano, 1998). The lessons learnt from this enterprise make us conclude that the transfer of sociologic theories is not just legitimate but also beneficial for both, theoretical and practical research in Distributed Artificial Intelligence.

References

- Castelfranchi, C. (1994): "Guarantees for Autonomy in Cognitive Agent Architectures". *Intelligent Agents* (Wooldridge & Jennings, editors). LNAI 890, Springer, pp. 56-70
- Conte, R.; Castelfranchi, C. (1995): *Cognitive and Social Action*. UCL Press
- Cuena J., Ossowski S. (1998): "Distributed Models for Decision Support". To appear in: *Introduction to Distributed Artificial Intelligence* (Weiß & Sen, editors). AAAI/MIT Press
- Decker, K.; Lesser, L. (1995): "Designing a Family of Coordination Algorithms". *Proc. 1st International Conference on Multiagent Systems (ICMAS-95)*, pp. 73-80
- Durfee, E.; Rosenschein, J. (1994): "Distributed Problem-solving and Multi-agent Systems: Comparisons and Examples". *Proc. 13. International Workshop On Distributed AI*, pp. 94-104
- Gasser, L. (1991): Social conceptions of knowledge and action: DAI foundations and open systems semantics. *Artificial Intelligence 47*
- Gergen, K.; Greenberg, M.; Willis, R. (1980): *Social Exchange: Advances in Theory and Research*. Plenum Press
- Gibbs, J. (1981): *Norms, Deviance and Social Control*. Elsevier
- Gilbert, N.; Conte, R. (1995): *Artificial Societies: The Computer Simulation of Social Life*. UCL Press
- Jennings, N. (1996): "Coordination Techniques for Distributed Artificial Intelligence". In: *Foundations of Distributed Artificial Intelligence* (O'Hare & Jennings, editors), Wiley & Sons, pp. 187-210
- Jennings, N.; Campos, J. (1997): "Towards a Social Level Characterisation of Socially Responsible Agents". *IEE Proc. on Software Engineering*, 144(1)
- Krogh, C.: (1996): *Normative Structures in Natural and Artificial Systems*. Ph.D. Thesis, Univ. of Oslo, Dpt. of Philosophy
- Kuper, A.: *Anthropologists and Anthropology – The British School 1922-72* (1972). Penguin Books
- Malone, T.; Crowston, K. (1994): "The Interdisciplinary Study of Coordination". *ACM Computing Surveys Vol. 26 No. 1*
- Ossowski, S. (1997): *On the Functionality of Social Structure in Artificial Agent Societies – Emergent Coordination of Autonomous Problem-solving Agents*. Ph.D. Thesis, Technical Univ. of Madrid, Dpt. of Artificial Intelligence
- Ossowski, S.; García Serrano, A. (1997): "Social Co-ordination Among Autonomous Problem-solving Agents". To appear in: *Proc. 3rd Australian Workshop on Distributed Artificial Intelligence (DAI-97)*, LNAI, Springer
- Ossowski, S.; García Serrano, A. (1998): "From Theory to Practice in Multiagent System Design: The Case of Structural Co-operation". To appear in: *Proc. German Annual Conference on Artificial Intelligence (KI-98)*, Springer
- Ossowski, S.; García Serrano, A.; Cuena, J. (1996): "Emergent Co-ordination of Flow Control Actions Through Functional Co-operation of Social Agents". *Proc. 12. European Conference on Artificial Intelligence (ECAI-96)*, Wiley & Sons, pp. 539-543
- Radcliffe-Brown, A.: *Structure and Function in Primitive Society*. Cohen & West Ltd., 1952
- Thomson, W.: "Cooperative Models of Bargaining". *Handbook of Game Theory* (Auman & Hart, eds.), 1994, pp. 1238-1284