Towards a Theory of Flexible Holons: Modelling Institutions for Making Multi-Agent Systems Robust

Michael Schillo[1], Ingo Zinnikus[2], Klaus Fischer[2]
Saarland University[1], DFKI GmbH[2]
Im Stadtwald, Building 43.8, 66123 Saarbrücken, Germany
schillo@virtosphere.de, {zinnikus, kuf}@dfki.de
Tel.: +49 681 302 4578, Fax: +49 681 302 2235

Abstract

Multi-agent systems have a reputation for bringing with them the property of robustness. However, multi-agent systems need to be specifically designed to display this property and until now it is unclear how this can be achieved. One way of approaching this, is to try to simulate social systems, as *social order* is in close analogy to robustness. We discuss this analogy, give an attempt to a definition of robustness, and a detailed analysis of *delegation* in multi-agent systems that we believe to help to achieve robustness. Delegation is an integral part of MAS and can be a source of flexibility. We discuss delegation on the spectrum between the absence of norms to the complete specification by norms. We argue that the concept of *flexible holons*, i.e. flexible grouping of agents due to task and social delegation, is a cornerstone to understanding the formation of institutions in multi-agent systems and to exploiting their potential contribution to robustness.

1. Motivation

Multi-agent systems are an integral part of current (D)AI research. Their success is in the majority of the literature ascribed to three major advantages: Firstly, many problems are inherently distributed and MAS reflect this distribution and provide off-the-shelf solutions. Secondly, the abstract concept of an „agent“ is of major importance to devise large systems and provide an abstraction to object-oriented programming comparable to that of object-oriented programming to its predecessors. And last but not least: Multi-agent systems are reliable, scalable, and robust:

- Distributed AI systems may be more *reliable* than are centralised systems because they provide redundancy, cross-checking and triangulation of results (Bond and Gasser 1988).
- A MAS has significant advantages over a single, mono-lithic, centralised problem solver: … more *flexibility* … and increased *reliability* (O‘Hare and Jennings, 1996).

- Agents can find ways to „work around“ unforeseen problems … an advantage of the agent-oriented approach is *scalability* (Bradshaw, 1997).
- As distributed systems they offer useful features such as parallelism, *robustness* and *scalability*, and therefore are applicable in many domains which cannot be handled by centralised systems (Weiß, 1999).

Although the literature treats robustness (or reliability, here almost used as a synonym) like an inherent feature of MAS (which it is not), there is hardly any discussion on what robustness actually means. A rare exception is the work of Klein and Dellarocas (1999, 2000) who define the normative behaviour of an agent, classes of exception and mechanisms for exception resolution to achieve. Another example is the work of Kaminka and Tambe (1998a, 1998b), who rely on ongoing agent monitoring and online behaviour adaptation. We can see that just like any other complex system, multi-agent systems need to be specifically designed to be robust. Compared to conventional computer science, the issue of robustness in MAS is different. Most computer science systems are *transformational systems*, which means they compute a function on some input. In this case, techniques for ensuring robustness do exist (cf. Ginsberg et al., 1998; Shlomi, 2000). However, the most interesting multi-agent systems are open systems (Hewitt, 1991), which do not explicitly compute a function (e.g. looking at the most predominant example of an open system, namely the Internet, as computing a function certainly misses the point).

In our research we are interested in making multi-agent systems robust. Our approach is to investigate analogies to human teams, institutions, and societies and try to transfer essential features to artificial teams, institutions, and maybe societies. For a discussion on the differences and similarities on the different levels of abstraction in human and artificial societies see (Schillo et al., 2001). On the search for a concept that helps to capture the aggregation of agents we belief that the concept of *holons*, i.e. agents that are made up by or represent groups of agents is most fruitful. Section 2 will deal with the description of the electronic market scenario we are developing our agents for. This description is followed by a discussion of the term social order (Section 3). We believe that beyond well-known definitions of agents from a technical point of view, some higher level properties are necessary ingredients for more socially competent agents to achieve self-organising capabilities and social order. We call the collection of these properties *bounded social rationality* (briefly summarised in Section 5). Building on these properties and the discussion of social order, we give a definition of robustness for multi-agent systems (Section 4). The central contribution of this paper is the analysis of delegation between these agents and its application to holonic systems in Section 6. Delegation will then be discussed in the context of institutions and norms in Section 7. We conclude with a discussion.

2. Scenario

The domain, which we are working on, is the domain of an electronic market for transportation orders. However, there are many features in this domain that are also relevant to other domains, which is basically a *task-oriented domain* (Rosenschein and Zlotkin, 1995). This renders our work applicable to other contexts as well. We describe these common features in what we call our scenario of a dynamic electronic market (see Figure 1). We assume here an agent communication platform that spreads over enough machines, and thus agents do not have to take care about how to contact other agents (as long as they exist) and access to computational resources. The agents act semi-autonomously for an institution. They acquire orders for some abstract kind of service and make sure that this service is produced. How the latter is achieved is up to the agent. Either it uses the institution's own resources for production or it tries to resell parts of the order (or the whole order) to subcontractors (other institutions obliged to fulfil orders from the agent) or cooperation partners. Agents are able to acquire new resources for production to increase their capacities. The efficiency with which orders can be processed, interacts, i.e. several orders of the same kind can be processed more easily than a similar amount of orders of different types. Therefore, evaluation of orders is not trivial.

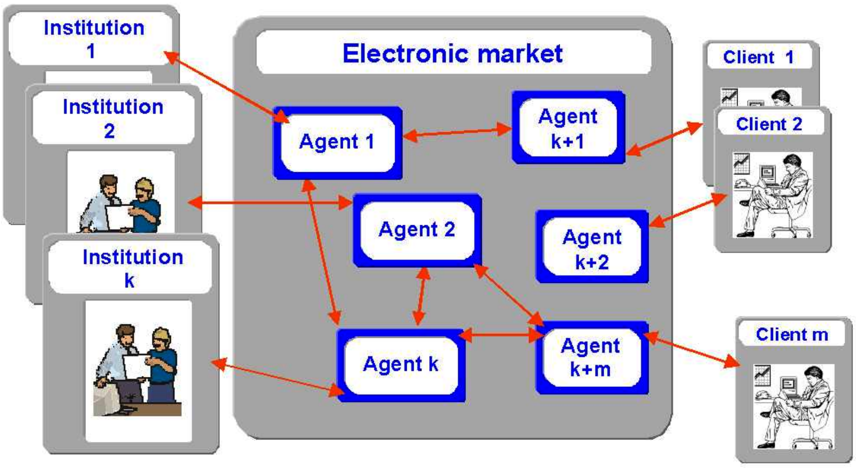

Figure 1 Agents and real world objects in our electronic market scenario.

The entire interaction necessary here is done by negotiation in auctions, specified by protocols. For a first step we use the *Contract Net Protocol* (Smith, 1980), but other auctions can be included at a later stage. In order to reflect practice in real markets we do not assume that price is the only criterion, which clients use for deciding for a product or a producer. Reliability, reputation, and past experience can (but not necessarily do) override simple price preferences. It is useful to assume that agents can observe auctions for orders they are involved in and can also observe whether any of their subcontractors or partners is also engaged in the same negotiation. In this case, agents should delegate bidding for this order among themselves to the one most likely to be accepted by the auctioneer.

In theory, this scenario allows for different roles of agents. Firstly, there is of course the agent who specialises on production by acquiring sufficient production resources. Secondly, there is the role of the "reseller" who merely seeks to live off the difference in price it obtains as middle agent when buying and reselling. Its advantage lies in having the contacts to bring together producers and clients and its reputation of being able to provide a broad range of products in a reliable way. Thirdly, there is the agent who does not need to produce everything by itself, but knows about the production resources of others and can competently offer subcontracts and -by repeatedly providing subcontracts- will be an attractive cooperation partner especially to agents of the first kind introduced above. One can easily see that given a set of concrete tasks and a continuous overlap of these roles inside one agent gives way to a magnitude of possible configurations for such agents to make up an institution.

We mentioned earlier that our agents are semi-autonomous. Autonomy of agents is a very complex phenomenon (Castelfranchi, 2000a). An agent can be autonomous (independent) or dependent on others concerning information, the interpretation of information, planning, its motivations and goals, resources, and authority ("being allowed to do *X*", deontic autonomy). What we mean by autonomy of an agent here is that agents act on behalf of the user (in this case the institution) without asking the user for confirmation for each interaction. Especially in the case of automated negotiation, interaction is massive and asking for confirmation here is not feasible (but of course the user can be asked for confirmation once an agreement between agents is found and before any commitment is made). A further aspect of autonomy is that the agent's knowledge and reasoning is not observable to others, i.e. the only way to find out about it is observing its behaviour and status in the agent society. The fact that autonomy and agents are linked so closely, makes multi-agent-based social simulation for this kind of research most attractive.

3. Robustness in Social Science: Social order

As Castelfranchi (2000b) states, social order is due to two different mechanisms. Firstly, there is the approach that focuses on providing incentives for certain behaviours and penalties for others (which we may call the extrinsic approach). And secondly, there is the approach of relying on the goals towards social order inherent in individuals (which we may call the intrinsic approach). Although the extrinsic approach is well investigated in multi-agent system research (in applications of game theory, mechanism design etc.), the research on intrinsic creation of social order is less investigated. This is due to the fact that interaction of intrinsic goals (or values) is much more difficult to investigate in theory and from our point of view requires more empirical research. It appears that social systems gain much robustness from the fact that (human) agents do not only follow economic rationality, but also have different, sometimes conflicting goals. In the empirical research of our project, there is proof that even in trading, the field where economic rationality almost by definition seems the ultimate criterion for decision making, many different utility expectations interact, e.g. reputation, ethics, past experiences, reciprocal relationships building on past experiences of cooperation, or hope on such cooperation in the future. Managers admit that informal cooperation between competitors exists for specified tasks with little profit and that subcontractors are deliberately given jobs in times of low income to let them survive economically to the next period of many customer requests, where the subcontracting will again be beneficial. And indeed, as Castelfranchi and Conte (1998) state, there is a fundamental problem underlying the notorious notion of rationality as simply economic rationality, namely the merely implicit notion of goals[1]. Not only is it unrealistic to assume that rationality can be reduced to simply maximising a single utility (be it money, level of satisfaction etc.). Looking at the practice shows that a whole range of criteria influence the decisions made, not all of which can be reduced to a definite monetary worth.

4. Bounded Social Rationality: A Collection of Basic Properties for Social Agents

Thus, we are confronted on the hand with a call for broader concepts of rationality for action and on the other hand with the need to cover all the different criteria that are taken into account by human decision-makers. The necessary set of properties we proposed to term *bounded social rationality* (Schillo et al., 2000b). This term covers the concept of decision making based on multiple criteria, which can refer to different goals and takes into account forms of rationality that are shaped by the individual's history as well as the history of the social field in which the agents are operating. It also encompasses that agents do actively evaluate their position in their social fields and can formulate and reason about these positions to pursue their goals. The idea for this concept arose from three orthogonal concepts:

- *Subjective rationality* as perceived by Castelfranchi and Conte (1998), which consists of two notions: a) an agent is "*deliberative*, such that it's action is the output of the individual agent's evaluation of existing alternatives as means to reach its goals" and b) the agent is "*self-interested* or *self-motivated*, acting only in service of its own goals, and choosing the action which allows the achievement of the highest number of most important goals at the lowest cost". This is a notion of rationality concerned with goals of an agent.
- *Social intelligence*, as the idea that the intelligence of the individual is shaped by the history of its social situatedness, a concept which also appears in Shoham and Tennenholtz's concept of *co-learning*. These concepts are based on the idea of co-development by being socially situated and the learning of agents instead of static definition of knowledge and abilities. A more operational definition is given by Conte for social reasoning, defined as the process of producing a social intelligent action: "A social intelligent action is an action which either takes among its inputs a mental state of another agent (...), in order to adapt to them, or is aimed to give as an output a change in the mental state of any given agent (…)" (Conte, 1999).

However, it can be expected that modelling the environment and its inhabitants as well as doing complex computations for action selection the agent runs the risk that by the time it has selected an action, the basis on which it did so may no longer be valid. As has been treated by a number of (possibly layered) agent architectures, the

[1] There are of course other points of criticism: the omniscience assumption, unbounded computation resource assumption etc. which, however, are commonly agreed unrealistic assumptions in game and decision theory.

agent behaviour needs some kind of anytime algorithm behaviour. This means, that its reasoning process must be interruptible, the performance of this process must be of non-decreasing performance over time, and the resources needed for reasoning must provide the feature of dynamic reallocation. This idea is expressed by the work of Russell and is the third concept we build our work on:

- *Bounded optimality*, as stated by Russell (1997) is defined as "…the capacity to generate maximally successful behaviour given the available information and computational resources." According to Russell, this capacity is necessary for the agent because of the imposition of constraints on actions and computations that the agent designer does not directly control.

5. Towards a Definition of Robustness in MAS

Robustness in multi-agent systems made of bounded social rational agents is more than introducing redundancy. Redundancy will not solve problems such as malicious agents in an open system, information overload and intractable increase of communication in large systems. Robustness requires re-organisation of a multi-agent system and thus, a broader conception of robustness is required. From our point of view, *robustness* can only be defined in relation to some definition of *performance measure*. Robustness is the ability of a system to maintain "safety-responsibilities" (Wooldridge et al., 1999) even though events happen that are able to disturb the system. So these safety-responsibilities must be defined, as well as it must be defined when the system performs well and how this can be quantified. When looking at an electronic market, we can for example identify the following three performance criteria. Firstly, how fast can a customer find someone who provides a desired service (speed of matchmaking) and secondly, do both parties, customers and producers, meet their needs, i.e. can producers earn enough to maintain their service and can customers find the service they need. Thirdly, the quality of service provided is of importance (product according to specification, on-time delivery of service, drop-outs, etc.). We define robustness *quantitatively* by the expected drop of the performance measure in four perturbation scenarios (i) **increase of population size**, (ii) **change of task profile** over time, (iii) malicious agent **intrusion**, and (iv) **drop-outs** of agents. Robustness can then be defined and measured by how much decrease of the performance measure will be the result of doubling the population size, five percent random drop-outs, etc. If the limit for calling a system robust is defined to be five percent, then there is a clear-cut *qualitative* definition of whether a system is robust or not. Mastering the perturbation scenarios corresponds to providing the following four properties:

Scalability: No decrease of performance if the size of the agent society is increased by a certain percentage. Specifically, this requires that patterns of interaction can react to the increased size of the community. Possible strategies are choosing different protocols, employing matchmakers, or organisation of participating parties to bigger entities (cf. the section on *holons*).

Flexibility: If any change in the environment occurs and the safety responsibilities cannot be maintained at the moment, is the system able to recover? How fast can the system recover to such disturbances and how fast can agents adopt their models of others (some kind of modelling is present in almost any multi-agent system, if not it is interesting to investigate how it deals with lack of knowledge about others *in combination* with scalability).

Resistance: Mechanisms to avoid a damaging effect of malicious behaviour of agents (namely lying in communication about facts in order to manipulate knowledge of agents, abusing protocols) on the community and the performance measure. Work on trust and betrayal in agent societies can be found in (Schillo et al. 2000a), which deals with the issue of betrayal in communication about others. In this work we showed that it is possible to increase the robustness of systems towards malicious agents by using trust as a complex mechanism for evaluating agents and isolating malicious agents from interaction in the population.

Drop-out safety: The reasons for agents halting execution are manifold, handling these situations is difficult. Approaches in this context are "shadow agents" i.e. agents that monitor other agents and replace them (Zinnikus and Funk, to appear) or adapt their plans accordingly (Kaminka and Tambe, 1998a; 1998b) and market-based approaches where requests are announced in contract-net protocol fashion and drop-outs are compensated by other agents getting their bids accepted. This requires a theory of delegation.

6. Holons as a Way of Modelling Institutions in MAS

6.1. Definition of Holonic Systems

Many application domains of multi-agent systems can be decomposed into particular subtasks performed by several agents, and often a domain allows hierarchical decomposing of tasks. That means, analyzing a domain shows that a task requires combining the activities of some subagents. To model these combined activities the concept *holonic agent* or *holon* has been introduced (Gerber, Siekmann, Vierke, 1999). The concept is inspired by the idea of recursive or self-similar structures in biological systems. A holonic agent consists of parts, which in turn are agents (and maybe holonic agents themselves). The holonic agent itself is part of a whole and contributes to achieve the goals of this superior whole. The holonic agent may have capabilities that emerge from the composition of subagents and it may have actions at its disposal that none of its subagents could perform alone. The subagents can give up parts of their autonomy to the holon. To the outside, a holon is represented by a *head (agent)* which moderates the activities of the subagents and represents the holon to the agent society.

Three forms of association are possible for a holon: firstly, subagents can build a loose federation sharing a common goal for some time before separating to regulate their own objectives. Secondly, subagents can give up their autonomy and merge into a new agent. Thirdly, any mix between the first and second scenario is possible, considered that agents can give up autonomy on certain aspect while retaining it for others. In this case of flexible holons, the responsibility for certain tasks and the degree of autonomy that is given up is subject to negotiation between the agents participating in the holon, not a matter of pre-definition by the designer. Furthermore, the role of the head can be distributed according to different tasks to different agents. Just like *fat trees* (multiple bypasses to communication channels) in massive parallel computing, the distribution of the task of communicating to the outside is able to resolve bottlenecks in communication. This makes delegation a principle action in the context of flexible holons and provides the basic functionality for self-organization and decentralized control. This means that holons have the increased ability to deal with inevitable change and therefore provide their own potential to achieving robustness in multi-agent systems.

6.2. Types of Operation: Social Delegation vs. Task Delegation

In the following we want to extend a theory of delegation that will help provide the four properties of Section 5 in multi-agent systems. This extension is inspired by the Habitus-Field theory of sociologist Pierre Bourdieu (1992), which is of importance to DAI in that it helps to extend psychological (individual) theories to social (distributed) theories. From Bourdieu's theory, we can derive four mechanisms of interaction of natural distributed systems that explain the stabilisation as well as the flexibility of a group or team. The interplay of theses mechanisms contributes enormously to the robustness of social systems.

Delegating tasks to other agents is not new to MAS research. Research on traditional benevolent task-oriented domains has for a long time been involved in how to distribute the right task to the right agent. But this model of delegation is restricted to two kinds of settings: settings where agents are all designed to share common goals or settings where agents simulate authority relationships. Neither of these settings apply in an electronic market. Here, a further mode is present, namely the negotiation of delegation. This means that agents decide on a case by case basis whether they delegate a task and to whom involving trust and risk. Recent work on delegation (cf. Castelfranchi and Falcone, 1998 for an extensive treatment), has shown that delegation is a complex concept and at least in semi-open systems to be very relevant to multi-agent systems research. This point is stressed also by Bourdieu for human societies, when attributing a central role to delegation. The mechanism of delegation makes it possible to pass on tasks to other individuals and furthermore, allows specialisation of these individuals for certain tasks (functional differentiation and role performance). Thus, we need to differentiate two *types of operation: task delegation*, which is the delegation of a sequence of goals to be achieved and *social delegation*, which does not consist of creating a solution or a product but in representing, for example a group or institution in a negotiation. For holonic agents social delegation is the finding of the *head agent* (at least for a specified subset of communication subjects) and task delegation the distribution of tasks among the member agents of the holon. Both types of operation are also essential for institutions in the real world, as they organise who is doing which job and whose job is communicating with other groups or teams in the organisation. This does not so much depend on a particular individual, but rather on somebody being able to perform this particular role.

6.3. Mechanisms for Delegation: How to Determine a Delegate?

It is important to note that finding the right individual for delegating a task to is not a trivial task. We observe four different mechanisms to determine a delegate: (i) a well-known mechanism is *voting*, whereby a group of equals determines by some voting mechanism (majority, two thirds, etc.) one of them to be the delegate. (ii) *Authority* is another well-known mechanisms, it represents the method of organisation used in distributed problem solving. (iii) *economic exchange* is a standard mode in markets. A good is exchanged for money, while the involved parties assume that the value of both is of appropriate similarity, i.e. the delegate is being paid for doing the delegated task or representation (iv) *social exchange:* economic exchange is not the only kind of exchange in social settings. Rather, gifts are being given (Think of a gift in an abstract way: This can for example mean that someone accepts dissimilarity in an economic exchange.) and favours are being done, all in expectation of either reciprocation or refusal of reciprocation. Both are indications to the involved parties about the state of their relationship. This kind of exchange entails risk, trust, and the possibility of conflicts (continually no reciprocation) and the need for an explicit management of relationships. The aim of this is to accumulate "social capital" that may pay-off in the future, much like a storage that can be used up in times of scarcity. Obviously, this kind of "storage" is also a source of robustness. Note that these four mechanisms work for both types of operation: economic exchange can be used for social delegation and voting for task delegation as well as vice versa. Their application corresponds to the explicitness of the norms specifying the roles, which are interacting in a given situation. If the norms completely determine the delegation, then the mechanism of choice is delegation by authority. The mechanism dealing with the other end of the spectrum, namely the lack of norms that reduce the complexity of the interaction, is the mechanism of the social exchange. Voting is also quite independent from norms, but at least requires the acceptance of the procedure of voting. Finally, the economic exchange is appropriate in situations, where no authority relationship is specified but relies on contracts and commitments both parties adhere to.

6.4. Delegation and Robustness

We believe that the two types of operation combined with the four mechanisms provide the basis to achieve the four properties necessary to master the perturbation scenarios. Social delegation supports scalability of multi-agent systems in that it structures groups of agents and reduces communication. It is important to note that similar findings show that in networks there are depending on the given situation, intervals between completely connected graphs and minimal spanning graphs, which exhibit the best of both worlds: reaching other nodes with minimal number of jumps and minimal number of edges in the graph (Watts and Strogatz, 1998). Holons are a measure for representing this in multi-agent systems, flexible representation is the mechanism to calibrate the network structure to the situation.

With the aid of task delegation we believe that multi-agent systems can achieve the flexibility to react to changing task profiles. Social exchange and the entailed concepts of trust and risk deal with norm-breaking agents (Schillo et al., 2000a). The thoroughly applied concept of delegation can provide the mechanism necessary to deal with dropouts. Possibly, the described types of operation and mechanisms are not complete, but we believe they are necessary ingredients for robust multi-agent systems.

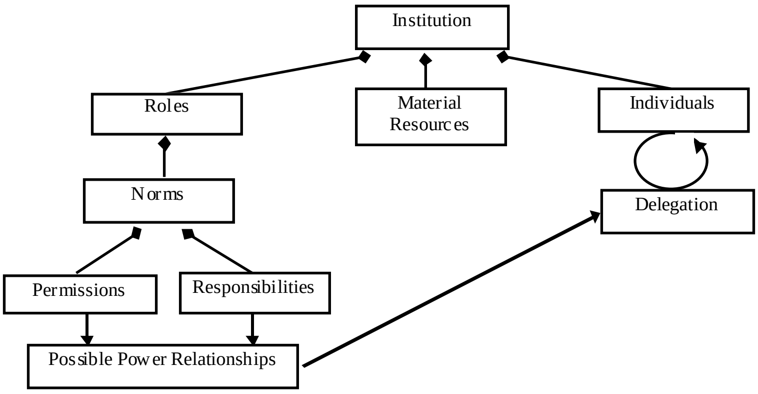

Figure 2 A sketch of the relationship between institutions, norms, individuals, and delegation. The diamond depicts (as in UML) the “consists of” relation.

7. Institutions, Norms, Delegation and Holons

As we view *institutions*, they are made up by their access to *material resources* (means of production, infrastructure etc.) and the *individuals* that belong to the institution (see Figure 2). In addition, the institution is also shaped by the *roles* the individuals have. These roles are defined by *norms*, which specify the role-specific *permissions* and *responsibilities* towards other roles/individuals and the material resources. This view is closely related to that of Carley and Gasser (1999). Delegation is a central concept here, as it describes what is happening between the individuals. Individuals can order someone else to do a certain task (if the roles are of appropriate constellation), or need to specify a delegate to represent their group in a committee. The possible power relations between individuals that are defined by normed permissions and responsibilities describe the enormous number of possible shapes of the social structure inside an institution. Many of them will never be instantiated. But depending on the individual needs in response to the actual situation, one or the other of the possible shapes of delegation between the individuals will actually happen. This means that the delegation that occurs is not arbitrary. It conforms to the permissions and responsibilities that come with the roles of the individuals. If we represent individuals by agents in this model, we can also represent that by social delegation the agents select a head agent. What would be a group or team in the real world, is here the creation of a holon. By continued social delegation we create holons that enclose more and more agents, which represents the instantiation of institutions. Re-organisation is then described by delegation of tasks and possibly by the change of social delegation during the life-cycle.

We need to note that norms do not mean they are necessarily obeyed, indeed there is evidence that some institutions only work, *because* norms are not fulfilled (a topic we could not cover here). The strength of institutions in the real world and potential for robustness in multi-agent systems is that to some extend, they do not rely on a particular individual, but on somebody able to perform this role.

8. Summary and Discussion

In this paper we reviewed the concept of "social order" in the social sciences and defined the basic properties for bounded social rationality beyond well-known agent definitions to describe what agents must possess to model them in complex social settings. We provided an attempt to a definition of robustness in multi-agent systems, which goes back to a performance measure and a set of perturbation scenarios, where the performance measure must maintain a certain level to call a system "robust". This can be the basis for flexible holonic systems i.e. the self-similar organisation of multi-agent systems. Their flexibility can be described and achieved by extensive use of the concept of delegation. We distinguish social delegation and task delegation. Both rely on four basic mechanisms by which the delegate can be chosen: economic and social exchange, voting and authority. We showed the principal role that delegation plays in (artificial) institutions and how the concepts role and norm relate to it.

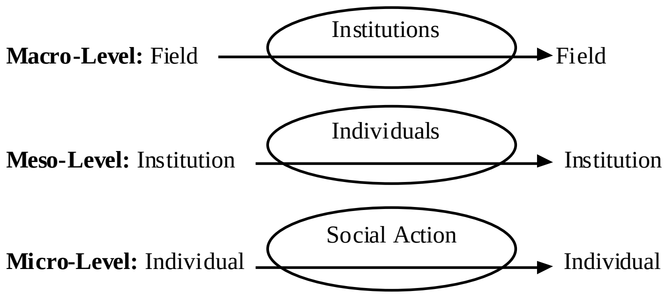

Figure 3 Levels of social aggregation and corresponding means of interaction.

In sociology the notion of scaling up a society is different to DAI (Schillo et al., 2001). While in DAI scaling up means increasing size of population, here scaling up means the increasing self-organisation on different levels of sociality. Note that the sociological scaling is according to different "levels" of sociality: the micro, meso and macro level (see Figure 3). While the micro level is defined by interaction between individuals through social action, the interplay of institutions as the meso level can happen only through individuals. The macro level is defined as the level where social fields (e.g. politics, religion, economy etc.) influence each other. We believe that this can only occur through institutions (which again interact only through individuals and so on). As the scaling in society provides much of the flexibility and robustness that human sociality possesses, institutions play a central role in reducing complexity in the real world and we believe that modelling institutions will also provide their potential to robustness to multi-agent systems.

References

Bond, A. H. and Gasser, L. (1988). *Readings in Distributed Artificial Intelligence*, Morgan Kaufmann.

Bourdieu, P., Wacquant, L. (1992). *An Invitation to Reflexive Sociology*, Chicago, Polity Press.

Carley, K. and Gasser, L. (1999). Computational Organizational Theory. In Gerhard Weiß, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence.* MIT Press.

Castelfranchi, C. (2000a). Founding Agent's "Autonomy" on Dependence Theory. In *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000), pp.353-357.*

Castelfranchi, C. (2000b). Engineering Social Order. In Omicini, A., Tolksdorf, R. and Zambonelli, F. (eds*.): Engineering Societies in the Agents' World (ESAW'00) at ECAI 2000.*

Castelfranchi, C. and Conte, R. (1998). Limits of economic and strategic rationality for agents and ma systems. *Robotics and Autonomous Systems, Special Issue on Multi-Agent Rationality*, vol. 24:127-139.

Castelfranchi, C. and Falcone, R. (1998). Towards a theory of delegation for agent-based systems*. Robotics and Autonomous Systems*, vol. 24, pp. 141-157.

Conte, R. (1999). Social Intelligence Among Autonomous Agents. *Computational & Mathematical Organization Theory,* **5:3**, pp. 203-228.

Dohlev, S. (2000). *Self-Stabilization.* MIT Press, Cambridge, Massachusetts.

Gerber, C., Siekmann, J., and Vierke, G. (1999). Flexible Autonomy in Holonic Agent Systems. In *Proc. of AAAI Spring Symposium 1999.*

Gilbert, N. and Conte, R. (1995). *Artificial Societies*. UCL Press.

Ginsberg, M., Parkes, A., and Roy, A. (1998). Supermodels and Robustness. Proceedings of the Fifteenth National Conference on Artificial Intelligence AAAI 98, pp. 334-339.

Hewitt, C. (1991). Open Information Systems Semantics for Distributed Artificial Intelligence. *Artificial Intelligence,* vol.*.47, pp. 79-116.*

Kaminka, G. and Tambe, M. (1998a). What is wrong with us? Improving robustness through social diagnosis. In *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).*

Kaminka, G. and Tambe, M. (1998b). The Role of Agent-Modeling in Agent Robustness. *Proceedings of the conference on AI meets the real-world.*

Klein, M. und Dellarocas, C. (1999). Exception Handling in Agent Systems. Proceedings of the Third International Conference on Autonomous Agents (Agents '99), Seattle, Washington, May 1-5, 1999, pages 62-68.

Klein, M. und Dellarocas, C. (2000). Towards a systematic repository of knowledge about managing multi-agent system exceptions. Proceedings of the internation Conference on Multi-Agent Systems, ICMAS 2000.

O'Hare, G. M. P. and Jennings, N. R. (1996). *Foundations of Distributed Artificial Intelligence*. John Wiley & Sons.

Rosenschein, J., and Zlotkin, G. (1994). *Rules of Encounter*. MIT Press.

Russell, S. (1997) *Rationality and Intelligence.* Artificial Intelligence, 94, p. 57-77.

Schillo, M., Rovatsos, M. and Funk, P. (2000a). Using trust for Detecting Deceitful Agents in Artificial Societies. In Castelfranchi, C., Tan, Y., Falcone, R., and Firozabadi, B. S. (eds.): *Applied Artificial Intelligence Journal*, Special Issue on "Deception, Fraud and Trust in Agent Societies.

Schillo, M., Fischer, K., Hillebrandt, F., Florian, M. and Dederichs, A. M. (2000b). Bounded Social Rationality: Modelling self-organization and adaptation using habitus-field theory. In Jonker, C., Lindemann, G., Letia, I., and Uthmann, T. (eds.) *Proceedings of the Modelling Artificial Societies and Hybrid Organizations (MASHO) of ECAI 2000,* pp. 112-122.

Schillo, M., Fischer, K., and Klein, C. (2001). The Micro-Macro Link in DAI and Sociology. In Moss, S. and Davidsson, P. (eds.) *Multi-Agent Based Simulation: Second International Workshop on Multi-Agent Based Simulation Boston MA, USA, July, 2000.* LNAI 1979, Springer.

Smith, R. G. (1980). The Contract-Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. *IEEE Trans. Computers*, S. 1104-1113, 1980.

Watts, D. and Strogatz, S. (1998). Collective Dynamics of "small world" networks. In *Nature*, vol. 393.

Weiß, G. (1996) Adaptation and Learning in Multi-agent systems: Some Remarks and a bibliography. In Weiß, G. and Sen, S. *Adaption and Learning in Multi-Agent Systems*. LNAI 1042, Springer.

Wooldridge, M., Jennings, N., and Kinny, D. (1999). A Methodology for Agent-Oriented Analysis and Design. In *Proceedings of the 3rd International Conference on Autonomous Agents, AA99,* pp. 69-76.

Zinnikus, I. and Funk, P. (To appear). Towards Self-stabilizing Agent-based Information Sources. Technical Memo, DFKI, Saarbrücken.