erschienen in: Klaus Fischer/Michael Florian (Hrsg.) (2005), Socionics: Its Contributions to the Scalability of Complex Social Systems, Berlin u.a.: Springer, S. 218-241.

From Conditional Commitments to Generalized Media: On Means of Coordination between Self-governed Entities[1]

Ingo Schulz-Schaeffer

Technische Universität Berlin, Institut für Soziologie, Franklinstr. 28/29, Sekr FR 2-5, 10587 Berlin, Germany
schulz-schaeffer@tu-berlin.de

Abstract. In the absence of pre-established coordination structures, what can a self-governed entity – i.e. an entity that chooses on its own between its possible actions and cannot be controlled externally – do to evoke another self-governed entity's cooperation? In this paper, the motivating conditional self-commitment is conceived to be the basic mechanism to solve coordination problems of this kind. It will be argued that such commitments have an inherent tendency to become more and more generalized and institutionalised. The sociological concept of generalized symbolic media is reinterpreted as a concept that focuses on this point. The conceptual framework resulting from the considerations is applicable to coordination problems between human actors as well as to coordination problems between artificial agents in open multi-agent systems. Thus, it may help to transfer solutions from one realm to the other.

1 Introduction

Coordination is the central theme for multi-agent research (cf. Van de Velde/Perram 1996: VIII). The key problem in this research area centres around ensuring coordination between agents (cf. Jennings 1996: 187). This problem is caused by the basic characteristics of agents: autonomy and pro-activeness. According to a well-known definition, agents are "hardware or (more usually) software-based computer system(s)" (ibid.) with at least these two properties: They operate autonomously "without the direct intervention of humans or others, and have some kind of control over their actions and internal state" (ibid.). And they function pro-actively, meaning that they "do not simply act in response to their environment", but "are able to exhibit goal-directed behaviour by *taking the initiative*" (ibid.). Humans in important respects can also be considered autonomous and pro-active entities and human societies have accumulated some thousand years of experience confronting coordination problems relating to autonomy and pro-activeness. Therefore, it is a promising idea to develop inter-agent coordination in analogy to forms of human interaction, which have proved to be successful.

[1] I am very grateful to Ying Zhu for the revision of my English.

However, to avoid that this analogy remains only metaphorical, we need to develop sufficiently precise concepts to define the common ground from which coordination in human as well as in agent interaction may evolve. In this paper, a general framework for dealing with coordination problems between self-governed entities will be provided. The basic suggestion is to conceive the initial situation of double contingency between self-governed entities, whose only means to affect others' behaviour is to make commitments, to be this common basis. This paper will show how conditional commitments can be employed by an entity to motivate another self-governed entity into responding cooperatively. Yet, this solution contains many restrictions. In adapting the concept of generalized symbolic media of interaction (or communication) to the problem of coordination, the emergence of generalized and institutionalised forms of commitments will be introduced as a means to overcome some of these restrictions. However, we will see that this general framework does not apply to closed multi-agent systems, since under the condition of closed systems less elaborate ways to deal with coordination problems are available and sufficient. On the other hand it is all the more important with respect to coordination problems in open multi-agent systems. Here, the analogy based on the common ground assumption works pretty well. In conclusion, I point out to the need for extending the analogy to hybrid systems.

2 Coordination in the Face of Double Contingency: Motivating Conditional Commitments

In the absence of given coordination rules or procedures, self-governed entities – i.e. entities that choose on their own between their possible actions – face a particular coordination problem when they aim to mutually adjusting their actions. This problem is defined by the initial situation of double contingency. Talcott Parsons and collaborators describe this situation with respect to two entities, ego and alter, as follows: „On the one hand ego's gratifications are contingent on his selection among available alternatives. But in turn, alter's reaction will be contingent on ego's selection and will result from a complementary selection on alter's part." (Parsons et al. 1951: 16). First of all, the behaviour of both parties involved is contingent on their own selections. Moreover, if alter is part of ego's relevant environment and vice versa – that is, part of the environment the respective entity takes into account when choosing its own behaviour – then one entity's selections are contingent on the selections of the other one. Thus, from the perspective of ego – the entity that wants to start a sequence of coordinated interaction – the situation of double contingency implies a double uncertainty: Ego does not know which behaviour to choose because it does not know which behaviour alter will choose in reaction to its action.

With respect to the goal of achieving coordination, the double uncertainty that accompanies situations of double contingency leads to a deadlock. If coordination is defined as establishing situations where two or more entities select their actions in a suitable way to commonly produce certain results, then the effect of this double

uncertainty is to prevent coordinated behaviour.[2] Hence, deadlock can only be broken through reducing uncertainty. At least the behaviour of one entity must become predictable to a certain degree so the other entity can rely on it while choosing its behaviour. However, one has to remember that we are dealing with self-governed entities. This means only the respective entity itself can make its own behaviour predictable. For a self-governed entity, the only means to do so is through self-commitment. For example, ego may announce:[3] "I commit myself to perform the action P every time I will be in the situation S." From alter's point of view such a self-commitment on the part of ego may or may not be useful: Alter will welcome ego's commitment if ego's action P contributes to what alter wants to achieve in the situation S. But as long as ego makes its commitment under the condition of uncertainty about alter's selections it is more probable that for alter this action P is without use.

From ego's point of view, an even more crucial problem is that this commitment does not – or only by chance – make alter cooperate with respect to ego's goals. Let us assume that ego is interested in the result produced by the combination of its action P and alter's action Q in the situation S. By fulfilling the self-commitment ego will do its part to bring about this result. But so far there is no reason why this commitment should enhance the probability of alter to react by performing the action Q. Consequently, ego runs the risk of constantly investing resources (by performing action P each time situation S comes around) without achieving its desired results. Upon reflection, ego may avoid this useless waste of resources by narrowing the self-commitment as follows: "I commit myself to perform action P in situation S on condition that you, alter, perform action Q.'[4] Does such a conditional commitment solve the coordination problem? The answer is yes, but only if alter as well as ego is interested in the results stemming from the combination of actions P and Q. In this case, if alter estimates ego to be trustworthy, it will perform action Q followed (if this estimation was appropriate) by ego's action P. But without such coinciding interests, again, alter is not inclined to cooperate.

As long as the entities involved cannot rely on given coincidences of interests, there is only one way to solve the remaining coordination problem: by producing such coincidences. But since a self-governed entity can affect the behaviour of other self-governed entities only by making commitments, producing coinciding interests has to be achieved by making commitments. Thus, the general strategy towards producing coinciding interests lies in ego to commit itself to act in the interests of alter (or to commit itself to refrain from acting against alter's interests) on condition that alter

---

[2] As I will argue below, the coordination problem resulting from the situation of double contingency is not only a problem of harmonizing actions with respect to common or complimentary goals but at first a problem of establishing common or complimentary goals. For this reason, I use the term coordination in a much broader sense than it is used for example by Esser (2000: 59-71).

[3] To simplify matters I will assume that the entities in question are able to use a common language. This is obviously a nontrivial assumption. But since the problem at hand is not how a common understanding between self-governed entities can occur but the problem of their coordination, this simplification seems to be justified.

[4] Or if action P has to be performed first: 'I commit myself to perform action P in situation S on condition that you, alter, commit yourself to perform action Q.'

does something ego is interested in (or refrains from acting against ego's interests). Let us assume that ego is interested in alter's action Q (which in combination with ego's continued actions will lead to a result it aims at) and that ego has reasons to believe that alter may be interested in its action P (which is, as we as godlike observers of the scene know, an appropriate assumption, since ego's action P in combination with alter's further actions will lead to a result alter aims at). By announcing: "I commit myself to perform action P on condition that you perform action Q." ego not only makes its own behaviour more predictable to alter but at the same time it tries to make alter's behaviour more predictable to itself. That is, the self-commitment now aims at reducing uncertainty from both sides of the double contingency problem. This is done by transforming the initial situation where only ego has an interest in alter to act in a specific way into a situation where alter has also an interest to act in this way, namely the interest in thereby bringing about ego's action P. And even though ego and alter's interests in alter's action Q are different, they are now coinciding interests in the sense that both parties are interested in alter performing action Q.

However, a number of preconditions have to be met so that such a self-commitment will work as described. First of all, ego requires certain resources at its disposal, resources allowing ego to change the situation in question to alter's advantage or disadvantage (and the same applies to alter with respect to the resources required to react as ego wants this entity to). Second, ego needs sufficiently reliable knowledge about alter's interests. Otherwise ego would not know how to use its resources to alter's advantage or disadvantage. Lastly, alter needs to trust ego's commitment towards the intended action. Alter will only be motivated to react accordingly when alter places confidence in ego's promised behaviour. None of this preconditions is trivial. Ego may or may not possess the resources necessary for motivating alter (and alter may or may not possess the resources required to adopt ego's proposal). And, in the positive case, ego may not know enough about alter's interests to employ its resources successfully. Even when ego holds the relevant resources at its disposal and knows how to motivate alter through the resources, alter could doubt ego's commitment, rendering ego's resources and knowledge useless. As we will see, all these preconditions push towards standardizing, generalizing, and institutionally framing such motivating conditional commitments to become more efficient means of coordination.

3 The General Framework as an Intermediary

When speaking about self-governed entities in the preceding section, I have avoided to specify the nature of the entities I have in mind. So far, I have only defined the entities as those able to choose between possible actions on their own. Speaking about possible actions implies that these entities have certain resources at their disposal that enable them to act in one or another way. The term "action" is only meant to designate a change in the entity's or its environment's state brought about by this entity. The preceding considerations also imply that these entities have interests in the sense that their self-governed behaviour is directed at the attainment or avoidance of

certain future states of affairs; that they are able to make plans that include actions on the part of other entities in order to realize or prevent these future states; that to this end they are able to reflect on their own interests as well as on the presumptive interests of other entities; that they are able to employ the concept of self-commitment; and that all interaction takes place under the rule of double contingency, meaning that no given structures guide the interaction.

Defining the initial situation in this way, the main intention was to choose a level of abstraction suitable to serve as a point of departure for dealing with both the problem of coordination between human actors (or corporative actors) as well as with the problem of coordination between autonomous software agents. Moreover, this general framework is intended to act as an intermediary between both realms of coordination problems. To this end it has been conceptualised so as to lie in between both of them. That is, some aspects apply more to human actors than to software agents. For example, pursuing interests and reflecting on presumptive interests of others are properties we usually assume an average competent human actor to possess. In contrast, software agents must first be programmed to possess any capabilities. But today certain agent architectures exist, especially the so called BDI agent architectures (cf. Shoham 1993; Haddadi/Sundermeyer 1996), that allow for implementing agents, which approximately show such properties. Hence, the assumptions in this respect are not altogether unrealistic.

On the other hand, some aspects of the general framework more appropriately describe the properties of software agents than those of human actors. This is the case with the absence of given structures. Since any structure guiding the agents' interaction must be pre-programmed by the designers of the respective multi-agent system, at first no structures exist that reduce the double contingency problem. Human actors, in contrast, grow up within given societies. From the individual's point of view the society is prior to him or her (cf. Mead 1967). Thus, there are always given social structures (cf. Durkheim 1982) reducing double contingency and possibly serving as coordination mechanisms. But if we are interested in understanding how such coordination mechanisms once came into existence and since we have all reasons to believe that they are constantly at risk to newly emerging uncertainties, our general framework seems to be a (to a certain degree counterfactual but nonetheless) useful point of departure for analysing the problem of coordination between human actors, too.

4 Motivating Conditional Commitments in Human Interaction

In the following section, I will apply the general framework to the problem of coordination between human actors. I shall counterfactually assume a situation of pure double contingency with respect to coordination issues within a certain population of human actors. In the absence of any pre-established coordination mechanisms, how may one actor (ego) gain another actor's (alter's) cooperation by means of motivating conditional commitments?

One way is to draw upon physical strength as a resource (or upon resources enhancing ego's capabilities to use physical force, such as weapons or other actors

ready to fight under ego's command) and to threaten to use it to alter's disadvantage. A respective self-commitment could run as follows: "I commit myself to harm you, unless you obey my orders." If, according to alter's estimation, ego possesses the resources required to make his threat come true (i.e. a physical strength or additional resources superior to his own), if alter believes ego will really use them accordingly, if alter possesses the resources required to obey, and if the threatened harm from the point of view of alter is more unfavourable than his obedience, then there is a good chance that alter will choose to obey, so that ego's self-commitment will result in alter's cooperation.

But perhaps ego wants to avoid the risk of alter's retaliation. He or she may be unsure of the own strength compared to alter's or may not see the coordination issue as worth the risk at all. Under these circumstances ego might prefer to motivate alter's cooperation by offering resources in return. Correspondingly, ego could announce: "I commit myself to place certain of my resources at your disposal on condition that you transfer certain of your resources to me." Again, if alter believes that ego really possesses the offered resources and trusts in his or her self-commitment, if alter has the resources at his disposal demanded by ego, and if from the point of view of alter getting those resources of ego serves his interests better than not giving away those resources of his own, then there is a good chance that alter will accept ego's proposal.

However, there are coordination problems of considerable relevance to human actors, which can not easily be solved by exchanging resources or by threat of force. This is the case when coordination requires participants to share common (or complementary) orientations. For example, if ego happens to be attracted to alter, much of their future interaction will depend on whether or not alter comes to feel the same. Or, if ego counts on alter to feel morally obliged to respond to him or her in a certain way, the participants' moral agreement is the basis for their coordination. Otherwise, the coordination rests on alter accepting as true, what ego holds to be true. Even though in these three cases the common (or complementary) orientations are different with respect to their content: truth, values or affective attitudes, the basic coordination problem is the same: To establish a certain kind of common or complementary orientations as a precondition of coordination.

For the time being, we want to do without referring to pre-established normative orientations – Parsons' solution to the problem of double contingency (cf. Luhmann 1984: 148-151). Moreover, we want to do without any given structure facilitating coordination (such as common knowledge or common affective attitudes). Thus, ego's request to adopt a certain aspect of his world view poses a problem to alter. Even if alter is sufficiently confident to gain his share of a successful cooperation based on his adoption of ego's orientation (this will be a necessary subject of ego's self-commitment), he can not know whether or not he is well advised at all to follow ego's suggestions. Therefore, ego has all reasons to try to convince alter of the truth of his assertions, the moral rightness of his convictions or the veracity of his feelings (you will have recognized the three Habermasian validity claims of communicative action, cf. Habermas 1987, Vol. I: 410-427).

To achieve this, self-commitments unfortunately are of limited help, since ego's statements are only subjective whereas alter should prefer to obtain objective information. Nevertheless, there is one thing ego by means of self-commitments can do: Demonstrating alter his persuasion concerning the truth, rightness or veracity of

his suggestions by committing himself to bear the consequences following from them. With respect to claims of truth this means for ego to commit himself to assume responsibility for the reliability of his assertions, e.g. to commit himself to compensating alter for damages that may occur if his knowledge turns out to have been unreliable. In doing so, ego places a kind of a bet on the reliability of his assertions, thereby disclosing the degree to which he himself is convinced (a strategy already recommended by Kant to assess a persons subjective persuasion, cf. Kant 1956: B 849f.; Krohn 2003). Under the condition that there is no other way to verify ego's assertions, such a commitment may serve as an auxiliary proof and motivate alter to adopt them. With respect to moral convictions, ego can demonstrate to alter his own persuasion by committing himself to obey to the values he wants to establish and, additionally, by committing himself to treat alter as if he, too, was subject to them. If ego believably exemplifies his moral convictions through his own behaviour and if alter for whatever reasons has an interest in being deemed to be a respectable person according to ego's moral standards, self-commitments of this kind may enhance the chance of alter to adopt ego's convictions. And with respect to feelings of relatedness, affection, and solidarity all ego can do is to make self-commitments to the effect that he will act according to those feelings, hoping thus to motivate alter to reciprocate.

Obviously, the success of all these attempts to initiate coordinated interaction by using motivating conditional commitments depends on many "ifs", particularly when coordination requires commonly shared or complementary orientations. But attempts to coordinate actions by threat of force or by exchange of resources include considerable uncertainties and restrictions, too. Thus, it is reasonable to assume that the chances of success could be enhanced substantially if not all these prerequisites have to be established anew, each time such a self-commitment is made. This is where the concept of generalized symbolic media comes into play.

5 Generalized Symbolic Media of Coordination

Parsons introduces the concept of generalized symbolic media in order to describe a "family of mechanisms" (Parsons 1963a: 42), which have in common that they are "ways of getting results in interaction" (ibid.) by "fac(ing) the object with a decision, calling for a response" (ibid.). According to Parsons, within the social system, this family of mechanisms comprises of four generalized media: money, political power, influence, and value-commitments (cf. Parsons 1975: 94-95). These "mechanisms are ways of structuring *intentional* attempts to bring about results by eliciting the response of other actors to approaches, suggestions, etc. In the case of money, it is a matter of offers; in the case of power, of communicating decisions that activate obligations; in the case of influence, of giving reasons or 'justifications' for a suggested line of action." (Parsons 1963a: 42). Starting from Parsons' concept of generalized symbolic media of interaction, Luhmann has developed a concept of symbolically generalized media of communication, where he arrives at a somewhat different list: In his opinion, besides money and power, truth and love are the most elaborated generalized media in modern societies. Additionally he considers religious

belief, art and basic civil values to be rudimentary forms (cf. Luhmann 1975: 176-179; 1984: 222; 1997: 332-358). But with respect to the role they play as mechanisms to coordinate the actors' selections (cf. Luhmann 1997: 320), his view on the generalized media resembles Parsons' perspective: They allow to condition the respective selection "so that it works as a means of motivation, that is, so that it can adequately secure acceptance of the selection" (Luhmann 1984: 222).

Both Parsons' and Luhmann's description correspond to the way in which the motivating conditional commitment works as a means of coordination. Thus, the question arises what is the difference and what is the advantage if an actor draws upon one of the generalized symbolic media when trying to motivate other actors to adopt his or her selections. Indeed, there is a certain similarity to the use of self-commitments as described above. But this is not so amazing, since proposing a selection with reference to one of these media is nothing else but making a motivating conditional commitment. In this respect no difference exists between an actor offering money in exchange for certain goods or services and another actor trying to arrive at the same result through a barter exchange. In both cases the respective actor, ego, attempts to motivate alter to agree to a certain transaction by committing himself or herself to transfer to alter something of value on condition that alter responds according to this request.

In other respects, however, drawing upon generalized symbolic media makes a difference. By referring to them, the possible success of attempting coordinated interaction is considerably enhanced. This is due to two basic properties of these media: generalization and symbolization. As indicated above, the success of the motivating conditional commitment alone is always in danger from the problems and restrictions posed by the particular circumstances of the prospective participants' individual situation. Drawing upon one of the generalized symbolic media, in this respect has the effect of transforming the concrete situation into an instance of a much more general situation, thereby overcoming at least some of these problems and restrictions. The term 'symbolization' refers to the fact that by using one of these media, the means to elicit a certain response is not the relevant resource itself, but a symbolic representation thereof. Again, one effect is decontextualisation, and to the degree this is the case, the generalized symbolic media are symbolically generalized media. Additionally, but not less important, the emergence of symbolic representations of this kind comes along with the emergence of institutional arrangements, whose function is to make sure that these symbolic representations work as if the 'real' resources, they stand for, were present (cf. Parsons 1975: 96). This, in turn, has the effect of simplifying matters, since some of the prerequisites for successful coordination now no longer have to be brought about by the prospective participants themselves, but can be left to these institutional arrangements.

The paradigmatic case of a generalized symbolic medium is money (cf. Parsons 1975: 94). In order to illustrate how the media's properties of generalization and symbolization can contribute to overcome problems of coordination, I will look at the case of money first. One of the major problems of barter trade is that the actor, who offers a certain commodity in exchange for another commodity, not only must find someone, who is interested in the offered commodity, but someone, who additionally is capable and ready to provide the desired commodity for exchange (cf. Coleman 1990: 119; Esser 1993: 557-558). Obviously, this problem of "double coincidence of

wants" (Coleman 1990: 119) is the greater the more uncommon the commodities in question are. But basically, this is a problem inherent in each attempt at making a barter exchange since the occurrence of coinciding complementary offers and requests is more or less unlikely on condition that each party is characterized by its own individual constellation of disposable and desired resources.

Money deals with this problem, since money, as Coleman puts it, "enables two parties to break apart the two halves of the double coincidence of a barter transaction. For example, B can engage in one half of the transaction with A, by providing services to A (in return for money), and then engage in the second half with C, who provides B with services 'in return' for those B provides to A (concretely in return for money B earlier received from A). B need not discover a D, who both needs what he can provide and has what he needs." (Coleman 1990: 120) In this way, money helps overcome a major impediment to economic exchange: "the fact that at any given time and place only one party of a pair who might engage in a transaction has an interest in what the other party has" (ibid.: 121). When dividing the transaction into two halves, money transforms a situation which is relatively specific and therefore relatively unlikely to occur into a situation that is much more general and therefore much more likely to occur: The resource that is offered in return for the desired commodity is now a resource everyone has an interest in, at least everyone who has interests in any commodities different from the ones already at his or her disposal. This is because the resource offered, a certain amount of money, represents exchange value, that is, the generalized capacity to exchange it for a certain amount of any commodity offered for money. In addition to this property to mediate exchange by serving as a general equivalent form of value (cf. Marx 1971 <1890>: 83-85), money has at least two further properties which by generalization of the situation help to enhance the chances of exchanges to occur: The property to be used as a store of value, and its property to function as a measure of value. The property of money as a store of value allows temporally to separate the single exchanges within an overall transaction: Because of this property an actor may be ready to provide a certain commodity in return for money at a given time even if the exchange for which he wants to employ this money will take place only some time later. Thus, the use of money not only allows to break up the double coincidence of actors complementary offering and looking for certain commodities, but the temporal aspect of this coincidence, too. And last but not least, the property of money as a measure of value "makes goods and services …, which in other respects such as physical properties are incomparably heterogeneous, comparable" (Parsons 1975: 95) and therefore much more easy to exchange.

In contrast to commodity money, such as gold, spices, or cigarettes that have been used to represent value, modern money no longer contains its value (in form of its value as a commodity), but merely symbolizes it. "It is symbolic in that, though measuring and thus 'standing for' economic value or utility, it does not itself possess utility in the primary consumption sense – it has no 'value in use' but only 'in exchange', i.e. for possession of things having utility." (Parsons 1963b: 236). Modern fiat money, such as the dollar, is "'valueless' money" (ibid.: 237), it "has no intrinsic utility, yet signifies commodities that do, in the special sense that it can in certain circumstances be substituted for them" (Parsons 1963a: 39). This feature of modern money – likewise to be abstracted from every commodity by only symbolizing value – "introduces new degrees of freedom" (ibid.: 40) in economic exchange, for example

because "money, unlike virtually all commodities, does not intrinsically deteriorate through time and has minimal, if any, costs of storage" (ibid.: 41). Hence, in the case of money, the positive effects of generalization are in part effects of symbolic generalization.

Another, but not less important consequence of the evolution of modern money is its dependence upon the co-evolution of certain institutional structures. This is the case, because the property of modern money to symbolize value is based on another symbolic property of money: Its property to serve as a symbolic representation of self-commitments of trusted third parties. This applies to fiat money as well as to its precursor, fiduciary money.[5] Fiduciary money represents a promise from its issuer (e.g. a bank or a trading house) to balance the debts it stands for, and fiat money is 'good' only as long as the government keeps the "promise to maintain a balance between growth in goods and services and growth in money supply" (Coleman 1990: 121). Thus, symbolic money requires in one or another way institutional arrangements securing the promise it embodies (i.e. its exchange value). Otherwise no one would accept intrinsically valueless money in exchange for intrinsically valuable commodities. But if trusted third parties of this kind do exist, a part of the commitments, which otherwise would have to be made by those engaged in a transaction can now be substituted by these trusted third parties' promises. Consequently, less trust must be invested in the respective other party involved in the exchange, serving as a further contribution of this generalized symbolic medium to make economic exchange more likely to occur.

If Parsons is right that money is only one, if perhaps the most prominent member of a "much more extensive family of media" (Parsons 1975: 94), and if it is appropriate to treat these media as media of coordination between actors,[6] then similar effects of generalization and symbolization should also be observable with respect to other media. I will address the issue of power only very briefly, since coordination by means of power is of little importance within multi-agent research, as I will argue below. Afterwards, I will discuss the case of influence in more detail.

Parsons describes power, in line with Weber, to be "the capacity of persons or collectives 'to get things done' effectively, in particular when their goals are obstructed by some kind of human resistance or opposition" (Parsons 1963b: 232). According to him, the difference between an attempt to obtain obedience by threat of force and the respective attempt by exercising power is comparable to the difference between a barter exchange and an exchange mediated by money: "Securing possession of an object of utility by bartering another object for it is not a monetary transaction. Similarly, … securing compliance with a wish … simply by threat of superior force, is not an exercise of power. … The capacity to secure compliance must, if it is to be called power in my sense, be generalized and not solely a function of one particular sanctioning act which the user is in a position to impose, and the medium used must be 'symbolic'." (Parsons 1963b: 237-238) Power, as well as

---

[5] With respect to the distinction between commodity money, fiduciary money and fiat money see Coleman 1990: 119-120.

[6] And not only as media of communication in the sense that they mediate the autopoietic emergence of specialized social systems (what is Luhmann's main focus, cf. Luhmann 1997: 359-371) or as media of interchange between the functional subsystems of the society (what for Parsons is of major interest, cf. Schimank 2000: 110-117).

money, has a 'real basis': "For the case of power, the basis of unit security corresponding to economic 'real asset' consists in possession of effective means of enforcing compliance … through implementing coercive threats or exerting compulsion." (Parsons 1963a: 47) Along with money, power is the generalized symbolic representation of this 'real basis'. It is this generalization and symbolization of physical force that makes binding obligations more likely to occur: "(J)ust as possession of stocks of monetary gold cannot create a highly productive economy, so command of physical force alone cannot guarantee the effective fulfillment of ramified systems of binding obligations." (ibid.)

Without any symbolic representation of physical force, the only way for alter to determine if ego maintains the capability to enforce his compliance is to test ego's physical force. But when ego was right in claiming superiority, the outcome of such a test is a disadvantage to both parties. Alter must face ego's sanctions as consequence to his disobedience, a situation alter would have avoided were he informed in advance. And ego is compelled to employ his force although he would prefer gaining alter's cooperation through deterrence. Power as a means to symbolize capabilities enforcing compliance in a generalized way, a way allowing a comparison between different amounts of such capabilities, helps to overcome such problems by enabling the participants to assess their relative capabilities in advance. However, in one respect, power as a coordination medium is substantially different from money. Power is not a medium in the sense that it intermediates between the parts of an overall transaction (cf. Esser 2000: 413-414). Normally, power is not a 'currency' that could be traded in exchange for compliance. Rather, it is a medium only in the sense that it makes it easier to grasp a special kind of relationship between actors: the power relation as the basis of coordination by dominance and submission. With respect to intermediating capacities, influence, the generalized medium I will turn to now, is much more similar to money than power can ever become.

As we have seen, Parsons and Luhmann agree that money and power are generalized symbolic media, but disagree about the remaining media. According to Parsons these are influence and value-commitment, whereas Luhmann holds that truth, love, and to a certain degree, religious belief, art, and basic civil values additionally play the part of generalized media. I will argue – partially in accordance with Parsons (1963a: 51-58) – that all such media should be viewed as representing different types of influence so that the medium influence constitutes a kind of a sub-family within the media family. Starting from a position that treats the generalized symbolic media as mechanisms to overcome coordination problems in situations where the motivating conditional commitment at first is the only means to initiate coordinated interaction, such an assumption makes some sense. As I have argued above, coordination on the basis of shared assertions of truth, of shared moral (or religious) beliefs, or of mutual feelings of affection (and, to include art: of shared aesthetic feelings) are similar, because in each case the establishment of one or another kind of a commonly shared (or complementary) perception of the particular situation at hand is the means to achieve coordination. In the absence of a pre-established common ground this leads to the question of why alter should be motivated to adopt ego's assertions, beliefs, etc. The admittedly unsatisfying answer I gave above refers to ego's degree of persuasiveness. As we will see now, influence of one or another kind generalizes and symbolizes persuasiveness, thereby helping to

overcome certain problems and restrictions that arise when alter's cooperation depends on his or her estimation that acting according to ego's definition of the situation lies in his or her own interest.

The crucial problem of coordination through adopting assertions, beliefs, or feelings lies in bringing about "a decision on alter's part to act in a certain way because it is felt to be a 'good thing' *for him*, … for positive reasons, not because of obligations he would violate through noncompliance" (Parsons 1963a: 48). For example, if it is a matter of adopting a certain information, "there must be some basis on which alter considers ego to be a trustworthy source of information and 'believes' him even though he is not in a position to verify the information independently – or does not want to take the trouble" (ibid.). As I have argued above, ego's commitment to bear the consequences following alter's adoption of his suggestions may provide such a basis to a certain degree. But these self-commitments' capacities to work in this way are limited: Alter is neither sure not to become subject to ego's fraud, nor can he or she rule out that ego is mistaken even if he himself truly believes what he says. For both reasons, it would make alter's decision easier if he or she could obtain more general knowledge about ego's performance in comparable situations. Thus, it would help alter to assess ego's trustworthiness in both respects, if he or she could relate the actual situation to prior experiences with ego in similar situations, or if he or she could find out to which degree other people feel positive about having adopted ego's suggestions in similar situations. And if such comparisons turn out to confirm ego's trustworthiness, ego will welcome them, since they enhance the persuasiveness of his suggestions. Influence of one or another kind can be understood to be the generalized symbolic medium that represents the accumulated perceptions of certain actors with respect to their trust in another actor's capability and willingness to make suggestions that will improve their situation, if adopted. In this sense, "(i)nfluence is a means of *persuasion*" (ibid.).

Since space is short, I will illustrate only one type of influence, namely scientific reputation. According to Merton, "graded rewards in the realm of science are distributed principally in the coin of recognition accorded research by fellow-scientists. This recognition is stratified for varying grades of scientific accomplishment, as judged by the scientist's peers."(Merton 1968: 56) In characterizing recognition by fellow-scientists, that is, reputation within a scientific community, as "the coin of the scientific realm" (Merton 1957: 644), Merton implies it to bear analogy to money. The basis of this analogy is the observation that in science reputation has become "symbol and reward for having done one's job well" (ibid.: 640). This leads to some questions: In which way is reputation a generalized symbolic media in the sense of an intrinsic valueless representation of something else of value? In which way does this medium help to overcome problems of coordination? And what does this currency buy?

In the case of scientific reputation, to have intrinsic value would mean to contain scientific truth, what reputation certainly does not. Rather, reputation contains information about the capability of a scientist to produce information of scientific value, as judged by fellow-scientists. Compared with what the scientist in question concretely has contributed to science, this is a rather general information, since it represents the accumulated recognition of several of this scientist's contributions by several of his fellow-scientists. And reputation is a symbolic representation, because it

does not somehow recapitulate or summarize the content of this scientist's accomplishments, but merely symbolizes his peers' estimation of how serious his contributions at large should be taken.

According to Luhmann, the "plausibility of reputation" (Luhmann 1990: 246) depends on the assumption that reputation will be attributed according to scientific accomplishment. But as we have seen, reputation does not directly follow from a researcher's contribution to science, but results from his fellow-scientists' perceptions thereof. Hence, the question arises as to how to make sure that reputation sufficiently corresponds with actual accomplishment. One part of the answer lies in the self-adjusting properties of reputation. If fellow-scientists involved in a particular research problem recommend referring to a certain scientist's contributions and other scientists, after following this recommendation came to the conclusion doing so was of no help, this will (at least in the long run) not only affect the recommended scientist's reputation, but those fellow-scientists' reputation, too. The fellow-scientists would appear to have given bad advice. In order not to compromise their reputation, scientists have a certain interest not to claim much more than their research really can contribute, and more general: an interest to live up to the expectations raised by the reputation they have already obtained.[7] The same applies to recognition by fellow-scientists. They, too, must be cautious not to misjudge other scientists' accomplishments, in order to save their own reputation.[8] In addition to this informal institutionalisation of the reputation mechanism, the referee system as a more formal mechanism to attribute scientific reputation has been established (cf. Zuckerman/Merton 1971). Like the institutions which are backing money, those informal or formal institutions' function is to make sure that reputation becomes and remains a sufficiently adequate symbol to represent scientific accomplishment.

In which way does reputation as a generalized symbolic medium help to overcome problems of coordination? As we have seen, the answer with respect to influence in general is that this medium communicates information about the presumptive quality of an actor's suggestions, what is useful in situations where on the part of the addressee of such a suggestion it is either impossible or too costly to verify its quality independently. This applies to scientific reputation too: "Studies of the communication behavior of scientists have shown that, confronted with the growing task of identifying significant work published in their field, scientists search for cues to what they should attend to. One such cue is the professional reputation of the authors. The problem of locating the pertinent research literature and the problem of authors' wanting their work to be noticed and used are symmetrical" (Merton 1968: 59): Since the readers' "behaviors in selecting articles" are, to a considerable degree, "based on the identity of the authors" (ibid.), their reputations, scientists must acquire

---

[7] As Merton (1968: 57) observes with respect to Nobel laureates, "the reward system based on recognition for work accomplished tends to induce continued effort, which serves both to validate the judgment that the scientist has unusual capacities and to testify that these capacities have continual potential. … It is not necessarily the fact that their own Faustian aspirations are ever escalating that keeps eminent scientists at work. More and more is expected of them, and this creates its own measure of motivation and stress. Less often than might be imaged is there repose at the top in science."

[8] This is a major reason of why a considerable part of the citations to be found in scientific texts refer to authors, whose scientific accomplishments are beyond doubt.

a reputation as a means of producing interest in their research results. To the extent to which reputation is a reliable indicator of scientific accomplishment, referring to it simplifies (cf. Luhmann 1990: 249) the scientists' search for those contributions of other scientists that prospectively are most important to their own work.

Thus, reputation helps to overcome coordination problems by supporting the allocation of scientific findings to those who will need them in their own research. Like money as a medium of economic exchange, reputation, too, enables two parties to break apart the two halves of a double coincidence: the double coincidence that the scientist looking for scientific findings useful for his work finds someone who is capable and ready to provide him with such findings. One half of the overall transaction consists of offering scientific findings to everyone who might be interested, i.e. of publishing them, in order to be rewarded by the fellow-scientists' recognition. The other half of the transaction exploits reputation, which those who offer their findings already have accrued, to single out what appears to be most promising contributions and in turn to pay recognition to its' authors if their findings actually turn out to be useful. Thus, from the perspective of those who pay recognition, reputation is a means to get authoritative advice with respect to their own scientific work. From the viewpoint of those attaining recognition, it is a means to strengthen their position "within the opportunity structure of science" (Merton 1968: 57), that is, their chances of "access to the means of scientific production" (ibid.). In this sense, "status, or recognition from others … has a characteristic that makes it somewhat like money: The value of a particular act of deference from a person is proportional to his own status. It is as if he has a particular quantity of status and pays out a certain fraction of it through the act of showing deference to another." (Coleman 1990: 130-131)

As we have seen, referring to generalized media has the effect of transforming particular coordination problems into instances of much more general ones, thereby reducing the need to meet the specific preconditions of the particular situation. The parties mutually have to know much less about their individual interests, strategies, capabilities, and trustworthiness, since part of what otherwise had to be negotiated between them now can be left to the respective generalized medium. Consequently, the generalized media are means of "disembedding", that is, of "the 'lifting out' of social relations from local contexts of interaction and their restructuring across indefinite spans of time-space" (Giddens 1990: 21). By helping to overcome restrictions imposed by local social contexts and the limitations of co-presence, they are powerful means of coordination making even cooperation between complete strangers probable to occur.

6 Bridging the Gap: Generalized Media as Emergent Effects of Conditional Commitments

If we want to draw upon generalized media as a way to overcome problems of coordination between self-governed entities as characterized by the general framework description given above, one major problem still remains: the problem of how generalized media come into existence, starting from an initial situation where all

an entity can do to initiate coordination is to make motivating conditional commitments. It must be shown that from this initial situation at least some development towards generalization (and symbolization) of conditional commitments may occur. Otherwise it would be much less useful as a point of departure for considering solutions to problems of coordination between self-governed entities. Additionally, the general framework would lose much of its relevance as an intermediary, that is, as a basic concept that helps to transfer means of coordination between human actors to the realm of software agents.

Luhmann (1975: 174; 1997: 316-317), in particular, emphasizes that generalized media result from self-reinforcing processes and counts on the possibility that such processes can be initiated by nothing more than one first suggestion of an actor to adopt his definition of the situation. This may happen as follows: "In a still uncertain situation alter decides tentatively to act in a certain way, as a first step. He starts with a friendly glance, a gesture, a gift – and then awaits ego's reaction to the definition of the situation thus proposed. In the light of this first step each following action has the determinative effect of reducing contingency – whether it be in a positive or in a negative way." (Luhmann 1984: 150) In the long run, such attempts at testing other actors' reactions to suggestions of one or another kind leads to more reliable expectations of how they typically will react. The generalized media can be understood to represent reliable expectations regarding the prospective reactions of other actors to certain types of suggestions. Thus, the next step towards generalized media is disembedding expectations from their local contexts of origin, that is, their transformation into general patterns of orientation with respect to certain types of coordination problems.

Even this next step can be deduced at least to a certain degree from the initial situation as characterized by the general framework. According to Esser (1993: 560-561), successful solutions to problems are almost automatically adopted by other actors, since they see that they will be better off in doing so. Consequently, a successful selection changes the situation by affecting other selections, thus setting off a process by which general solutions to typical problems are developed and become institutionalised. Carl Menger's outline of an 'organic' development of commodity money, to which Esser refers in this context, follows this pattern of explanation: according to Menger, it takes only a simple observation to solve the problem of double coincidence in barter exchanges, the problem that "not only does A have something that B wants, but it is also true that B has something that A wants, and both want what the other has more than they want what they themselves have, which they are willing to give up in exchange" (Coleman 1990: 119): Each individual can easily discover that some commodities, in comparison to others, are on greater demand. So, when looking for a certain commodity, it is more likely to find someone, who offers it, among the many, who themselves look for a marketable commodity than among the few, who look for a less marketable item. Thus, it is an obvious idea for anyone who offers a less marketable commodity to exchange it not only for the commodity he looks for, but – if this is not possible – also to exchange it for other commodities which he does not need but for which there is a greater demand than for his own commodity. In this way he gets nearer to his aim, since now he can look among the many, who are in demand for this commodity, for someone who offers what he wants. And if he is successful, a transaction mediated by this demanded

commodity has been performed. Based upon this solution of the double coincidence problem, everyone will prefer to possess among the highly demanded commodities those with the highest marketability, the least deterioration, and the best divisibility, thus paving the way for the development of commodity money whose primary function is to be a medium of exchange (cf. Menger 1969 <1883>: 174-176).

In the same way, influence can be understood to result from aggregated and thereby generalized experiences with cooperation on the basis of conditional commitments. At least with respect to influence, as far as it is an informally institutionalised medium of coordination, this can be easily shown. An everyday experience serves as an appropriate example: When you have moved to a new place, how do you find a dentist, in whose skills you hope you can trust? Probably, you will ask your new neighbours or colleagues and follow their recommendations. The rationale behind this strategy is as follows: from the point of view of the neighbours or colleagues, the recommended dentist so far kept his promise to treat their dental problems successfully, otherwise they would not have recommended him. So, it seems likely that this dentist will be capable of dealing with other persons' dental problems as well. Following the neighbours' or colleagues' recommendations means adopting a solution to a problem which has proved to be successful. At the same time, it implies a certain degree of generalization of this solution as a precondition of its transferability. The person who follows such recommendations, concludes from other persons' individual experiences with this dentist's capabilities to his competency in general. Or to put it another way: He refers to this dentist's professional reputation as an indicator of his problem-solving capacity.

To a certain degree even power can be conceived to emerge from the initial situation in which the motivating conditional commitment is the only means of coordination. This can be observed in situations where the physical means to enforce compliance are distributed relatively equally among the parties involved, but their readiness to actually employ them is not. In such situations those more willing to use force will often gain a capacity to secure compliance far beyond their relative physical strength, thus accumulating power as distinct from force. Nevertheless, the emergent qualities of power relations are limited, since the basis of this medium, that is, the "possession of effective means of enforcing compliance" (Parsons 1963a: 47), is less affected by exchanges mediated by power than it is in the case of money or influence, because power usually is not a circulating medium[9] in the sense of transferring what it represents. Rather than to be an emergent effect of prior exchanges, the possession of such effective means primarily results from decisions of system-builders, for instance of those who found a company and delegate rights and resources to the different positions within this company. Thus, "designed institutions" rather than "emergent institutions" (cf. Conte 2001) are the basis of coordination mediated by power.

[9] In this respect, Parsons (1963b: 245-246) argues to the contrary.

7 Generalized Media of Coordination in Closed Multi-agent Systems?

In early stages of research in multi-agent systems one can already observe the implementation of coordination mechanisms, which at first glance resemble coordination by means of generalized media: The contract net protocol (Davis/Smith 1983: 77) emulates a simple mechanism of economic exchange. It implements a coordination structure, where the agents involved possess the capabilities to announce tasks they want to be carried out by other agents, to "evaluate their own level of interest" (ibid.) in performing tasks announced by other agents, to submit bids according to these evaluations, to evaluate received bids and to select between them. Often a certain quality of the task itself – for example the amount of time the bidding agents state they will need to perform it – is used to evaluate them (cf. Schulz-Schaeffer 2000: 30-32), but sometimes a more general measure of value, some kind of money, is used (cf. Wellman 1992). In both cases the respective coordination mechanism goes beyond barter exchange in that it employs a more or less generalized medium to make different offers (or different announcements) comparable.

Coordination by identifying the agent that is best suited for performing the respective task in a population of agents with different expertise is another strategy often employed in multi-agent research. An early example of coordination structured by the agents' particular skills is the distributed vehicle monitoring testbed (cf. Lesser/Corkhill 1983). In this case, the "spatially-distributed nodes detect the sound of vehicles, and each applies knowledge of vehicle sounds and movements to track a vehicle through its spatial area. Nodes then exchange information about vehicles they have tracked to build up a map of vehicle movements through the entire area." (Durfee et al. 1989: 74) Another way to use an agent's particular ability as basis of their coordination is not to refer to skills but to willingness. Much of the early multi-agent research starts from the so called benevolent agent assumption: "Agents are assumed to be friendly agents, who wish to do what they are asked to do." (Martial 1992: 41) In this case coordination is brought about by the agents quality to be "perfectly willing to accommodate one another" (cf. Davis 1980: 42).

Coordination by implementing means to secure compliance seems to play a part in every multi-agent system, but mostly in an implicit way. An explicit suggestion is Shohams and Tennenholtz' idea of imposing social laws on agents. Taking the domain of mobile robots as their example, they argue: "Suppose robots navigate along marked paths, much like cars do along streets. Why not adopt a convention, or, as we'd like to think of it, a social law, according to which each robot keeps to the right of the path? If each robot obeys the convention, we will have avoided all head-on collisions without any need for either a central arbiter or negotiation." (Shoham/Tennenholtz 1992: 277) But most multi-agent researchers do not like this idea very much. While they concede that it is indeed an effective way to overcome coordination problems, they fear that pre-programmed conventions or laws of this kind will reduce the autonomy and pro-activeness of the agents (cf. Schulz-Schaeffer 2000: 45-48) and thereby affect what is held to be the distinctive feature of this strand of research: coordination as an emergent effect of interaction between agents without a central authority.

Though these solutions at first glance seem to resemble coordination by means of generalized media, in fact they bear only a poor resemblance, because – as Castelfranchi puts it – the respective approaches in multi-agent research "still remain in a world of 'pre-established harmonies'" (Castelfranchi 1990: 50; cf. Conte/Castelfranchi 1994: 268). At least within closed multi-agent systems, coordination is largely a result of predefined patterns of behaviour: The agents who are subject to rules, conventions, or social laws cannot act otherwise than to comply, the benevolent agents do not possess the option to act malevolently, the spatially or functionally distributed agents interact on the basis of given knowledge about their respective skills, and the agents announcing and bidding for tasks do not possess interests that could interfere with the performance of the overall exchange system.

Closed multi-agent systems are characterized by the fact that the development and implementation of all agents involved, as well as of the system's architecture (e.g. inter-agent relations, interaction protocols) are completely in the hands of one designer or designer team (cf. Schulz-Schaeffer 2000: 14). In this case, there is little reason why designers should provide agents with properties which pose impediments to coordination, then being forced to implement coordination mechanisms in order to overcome the coordination problems resulting from these properties. Why should they develop agents that are able to violate obligations, thereby raising the need to develop means to secure compliance? Why should they employ complicated reputation mechanisms to enable agents to evaluate their respective skills, if they possess complete knowledge about their agents' capabilities because they have programmed them and could easily distribute this knowledge among the agents? The answer is, they will not, and they actually do not, when the aim is to develop agents, which efficiently coordinate their actions within closed multi-agent systems. Rather, multi-agent researchers developing closed systems restrict their agents' conduct in a functional way so that collaborative problem-solving necessary results from their pre-programmed patterns of behaviour. Thus, coordination between agents in closed multi-agent systems differs from our general framework in that it heavily relies on pre-established structures. This does not mean that closed multi-agent systems cannot be modelled on human social systems. Since a large part of social interaction between human actors is very successfully governed by given social structures, quite the contrary is true. But it does mean that in closed multi-agent systems there is little need to refer to conditional commitments and to generalized forms of making commitments (i.e. generalized media) in order to ensure coordination.

7 From Conditional Commitments to Generalized Media in Open Multi-agent Systems: The Paradigmatic Case of Reputation

The situation completely changes when we move on from closed to open multi-agent systems. Open multi-agent systems can be pragmatically defined as systems where the behaviour of the agents involved is not developed and is not completely controlled by one designer or one designer team (cf. Schulz-Schaeffer 2002: 246). More precisely, open multi-agent systems are characterized by one or both of the following attributes: They are systems with open membership, in which every designer or user

who wants his agent to become a participant, in principal may do so (cf. Schulz-Schaeffer 2000: 17-21). And/or the (or some of the) agents involved may be subject to emergent properties. In other words, in the course of their 'life' agents are able autonomously to change their patterns of behaviour, for instance by 'learning' from prior experiences. The growing interest in research in open multi-agent systems mainly results from the consideration that there will be a lot of promising applications in the domain of agent-based web services, which presuppose open systems at least in the sense that all agents, which are authorized by their users to engage in certain transactions may possibly become cooperation partners.

With respect to the question of inter-agent coordination, the most important consequence of the two aspects related to openness is that agents now are – as Hewitt (1986: 322; 1991: 81-82) calls it – "at an *arm's length relationship*". This means that the "internal operation, organization, and state of one computational agent may be unknown and unavailable to another agent" (Hewitt 1986: 322) so that the agents know about one another only what they communicate to others. Since these agents act only according to their respective designers' or users' specifications (or in the case of emergent features: according to their own advancement of such specifications), there is no way to ensure collaboration by means of pre-established structures. Rather, nothing else but the negotiations between the agents account for success or failure of an agent's attempt to initiate coordinative interaction.[10] Thus, agents in open multi-agent systems are confronted with the coordination problem as characterized by our general framework.

I have argued that in the absence of pre-established coordination structures the only means a self-governed entity has to evoke another self-governed entity's cooperation is to motivate this entity to act in a certain way by making conditional commitments. Additionally, I have tried to show that this solution to coordination problems has an inherent tendency to become more and more generalized and institutionalised, thereby removing some of the restrictions of the initial situation, in which alter when deciding whether to follow a suggestion he is asked to adopt can consider nothing more but ego's conditional commitments. If this is true, similar ideas and efforts should be observed in research on open multi-agent systems.

Indeed, in the last decade we have witnessed a lot of pioneering work in establishing commitment as a basic concept of coordination between agents (see for example Bond 1990; Cohen/Levesque 1990; Jennings 1993; Castelfranchi 1995). In this period, some researchers have even gone so far as to claim that "(a)ll coordination mechanisms can ultimately be reduced to (joint) commitments and their associated (social) conventions" (Jennings 1993: 234). Within a few years the concept of coordination through commitments made by agents, according to Castelfranchi and Conte, has become the dominant view: "No preexisting relationships, no objective bases, no specific motivations for cooperation are supposed in the agents, no obligations and constraints, except their free commitments, are thought to be put on them." (Castelfranchi/Conte 1996: 537) Its wide acceptance can be underlined by the

[10] Again, the only given structure that has to be presupposed is the existence of a common language, that is, of a communication protocol such as KQML (cf. Finin et al. 1993; Labrou/Finin 1997) or the agent communication language of the FIPA (cf. http.//www.fipa.org), which is being used by all agents, participating in communication, to make sure that suggestions and responses are properly understood.

somewhat exaggerated assertion that "(t)ypically, multi-agent systems … use centrally the concept of commitment" (Aubé/Senteni 1996: 13).

In accordance with the considerations above regarding the motivating conditional commitment as the first step towards coordination, it has been emphasized that an agent's commitment should be thought of as a social commitment in the sense of a promise to act in a way another agent is interested in (or is interested in to prevent), that is, as a "(c)ommitment of one agent to another" (Castelfranchi 1995: 41, cf. ibid.: 42-45), but that it is based on an "*internal* commitment" (Cohen/Levesque 1990: 257): that the agent commits itself to act in this way (cf. Jennings 1993: 236). Likewise, it is seen that "(w)ith respect to coordinating the behavior of multiple agents, the most important feature of commitments is that they enable individuals to make assumptions about the actions of other community members. They provide a degree of predictability to counteract the uncertainty caused by distributed control" (ibid.: 240), that is, caused by the fact that the agents involved are self-governed entities.

However, most of the early approaches assume 'good faith', postulating that "agents commit only to what they believe themselves capable of, and only if they really mean it" (Shoham 1993: 64; cf. Jennings 1996: 195; Castelfranchi 1995: 45), allowing obligations to be revoked only after "explicit release of the agent by the party to which it is obliged", or when it turns out that the agent "is no longer able to fulfill the obligation" (Shoham 1993: 65; cf. Cohen/Levesque 1990: 254-256) In the meantime it has become widely recognized that in open systems this good faith assumption is as unrealistic as the benevolent agent assumption of the early days was (cf. Rosenschein/Genesereth 1988: 227), since it does not take into account the possibility of incompetence or fraud. Consequently, the question of how to enable agents to assess other agents' trustworthiness in order "to make our agents less vulnerable to others' incompetent or malevolent behavior" (Marsh 1994: 97), has become a major topic in multi-agent research (see for example Castelfranchi/Falcone 1998). In particular, much research has been done in recent years on reputation mechanisms (cf. Conte/Paolucci 2002).

This interest in reputation has much to do with its emergent properties. For researchers who fear that pre-designed coordination structures might reduce the autonomous problem-solving capacities of interacting agents, but who nevertheless acknowledge that there is a need for means to reduce coordination problems, it is an intriguing idea that such a means of coordination may "emerge from a spontaneous process" (Conte 2001), that is, from the accumulated past experiences one agent has made with another agent, or additionally, from the accumulated recommendations of other agents reflecting their experiences with the performance of this agent. Thus, it is the property of being "an intrinsic enforcing mechanism" that does not need to be "controlled by a given external entity", but is controlled "by the whole group" (ibid.), that makes reputation being viewed as a promising means of coordination between autonomous agents. In accordance with the considerations regarding generalized media of cooperation, multi-agent simulations have shown that the effectiveness of the reputation mechanism grows in line with its generalization. For example, Castelfranchi, Conte, and Paolucci (1998) have compared two experimental settings with respect to the ability of 'respectful' agents (i.e. those who follow a certain norm) to identify 'cheating' agents (i.e. those who do not). In the first setting the agents

learn about other agents' behaviour only from their own experiences, in the second setting they exchange their experiences. The result of the simulation is not surprising (in the second setting the agents are much better in identifying cheaters), but clearly shows the use of accumulating experiences, and that means: the use of transforming individual evaluations into general indicators of agents' performance.

8 Closing Remark: The Need to Hybridise Open Multi-agent Systems

In the previous section I have dealt with the general framework of analysing coordination problems between self-governed entities, starting from commitments as the only means to motivate another self-governed entity's cooperation. I have tried to show how it may enhance our understanding of coordination problems between agents in open multi-agent systems and how it may help to identify impediments to coordination in open systems as well as the respective means of coordination that have emerged in human societies to deal with them. Obviously, in addition to the problem of whether to trust in an agent's commitments, there are a lot of further impediments to coordination posed by the initial situation of double contingency. Thus, identifying processes where more general means of coordination emerge from conditional commitments may in this respect prove to be of help to overcome problems of coordination in open multi-agent systems.

However, modelling coordination mechanisms between agents on conditional commitments and their emergent generalizations raises a problem that should not be ignored. I have argued that an important aspect of the generalized media's capabilities to facilitate cooperation is that they allow to substitute reliance on individual actors' intentions or resources by reliance on institutions. The more these means of coordination are symbolically generalized, the more important (and the more efficient) this institutional background becomes: Fiat money ultimately relies on the capability of the society's central bank to prevent inflation or deflation; political power in the end relies on the capability of the state to hold the legitimate monopoly of force. This leads to the question of how ultimately to ensure the reliability and trustworthiness of the institutions supporting coordination in open multi-agent systems.

So far, this question has been answered with reference to the self-adjusting properties and to the intrinsic enforcing mechanisms of emergent institutions. For example, the agent who does not stick to his commitments in the long run will gain a bad reputation, and, consequently, will be avoided by other agents. But these intrinsic properties fail to prevent certain malpractices: a user might employ the strategy to always kill his or her agent after having gained a bad reputation and create a new one. Or he or she might choose to create additional agents who deceptively recommend this agent's trustworthiness so that it will gain a good reputation (cf. Spiegel Online 2003). It should be obvious that malpractices of this kind cannot be avoided by means of coordination mechanisms, which only affect the behaviour of the agents. Rather, institutional arrangements are required to make sure that what affects an agent affects its user as well. To this end, access rules to open systems and rules regarding the

users' responsibility for their agents behaviour have to be established. If only for this reason (in fact, there other good reasons, too, cf. Schulz-Schaeffer 2001), research on open multi-agent systems necessarily leads to research on hybrid systems, that is, on systems of interaction among and between computational agents and human actors.

References

Aubé, Michel/Alain Senteni (1996): Emotions as Commitments Operators: A Foundation for Control Structure in Multi-Agent Systems, in: John P. Perram (Hrsg.), Agents Breaking Away. 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW '96, Proceedings, Berlin u.a.: Springer, S. 13-25.

Bond, Alan H. (1990): A Computational Model for Organization of Cooperating Intelligent Agents, in: (Hrsg.), Proceedings of the Conference on Office Information Systems, Cambridge, Mass., S. 21-30.

Castelfranchi, Cristiano (1990): Social Power. A Point Missed in Multi-Agent, DAI and HCI, in: Jean-Pierre Müller (Hrsg.), Dezentralized AI, Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Amsterdam u.a.: Elsevier Science Publishers (North-Holland), S. 49-63.

Castelfranchi, Cristiano (1995): Commitments: From Individual Intentions to Groups and Organizations, in: Victor Lesser (Hrsg.), ICMAS-95. Proceedings of the First International Conference on Multi-Agent Systems, June 12 -14, 1995 in San Francisco, California, Menlo Park u.a.: AAAI Press/ The MIT Press, S. 41-48.

Castelfranchi, Cristiano/Rosaria Conte (1996): Distributed Artificial Intelligence and Social Science: Critical Issues, in: N. R. Jennings (Hrsg.), Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons, S. 527-542.

Castelfranchi, Cristiano/Rosario Conte/Mario Paolucci (1998): Normative Reputation and the Costs of Compliance, in: Journal od Artificial Societies and Social Simulation 1(3), http://www.soc.surrey.ac.uk/JASSS/1/3/3.html.

Castelfranchi, Cristiano/Rino Falcone (1998): Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification, in: (Hrsg.), ICMAS'98. Proceedings of the International Conference on Multi-Agent Systems, Los Alamitos, Cal.: IEEE Computer Society, S. 72-79.

Cohen, Philip R./Hector J. Levesque (1990): Intention Is Choice with Commitment, in: Artifical Intelligence 42, S. 213-261.

Coleman, James S. (1990): Foundations of Social Theory, Cambridge. Mass. u.a.: The Belknap Press of Harvard University Press.

Conte, Rosaria (2001): Emergent (Info)Institutions, in: Cognitive Systems Research 2, S. 97-110.

Conte, Rosaria/Cristiano Castelfranchi (1994): Mind is Not Enough: The Precognizive Bases of Social Interaction, in: Jim Doran (Hrsg.), Simulating Societies. The Computer Simulation of Social Phenomena, London: UCL-Press, S. 267-286.

Conte, Rosaria/Mario Paolucci (2002): Reputation in Artificial Societies. Social Beliefs for Social Order, Boston: Kluwer Academic Publishers.

Davis, Randy (1980): Report on the Workshop on Distributed AI, in: Sigart Newsletter 73, S. 42-52.

Davis, Randy/R. G. Smith (1983): Negotiation as a Metaphor for Distributed Problem Solving, in: Artificial Intelligence 20(1), S. 63-109.

Durfee, Edmund H./Victor R. Lesser/D. D. Corkhill (1989): Trends in Cooperative Distributed Problem Solving, in: IEEE Transactions on Knowledge and Data Engineering I(1), S. 63-83.

Durkheim, Emile (1982): The Rules of Sociological Method. Edited with an Introduction by Steven Lukes, New York u.a.: Free Press.
Esser, Hartmut (1993): Soziologie. Allgemeine Grundlagen, Frankfurt/Main u.a.: Campus.
Esser, Hartmut (2000): Soziologie. Spezielle Grundlagen, Bd. 3: Soziales Handeln, Frankfurt/Main u.a.: Campus.
Giddens, Anthony (1990): The Consequences of Modernity, Stanford Ca.: Stanford University Press.
Habermas, Jürgen (1987): Theorie des kommunikativen Handelns, Vierte, durchgesehene Auflage, 2 Bde, Frankfurt/Main: Suhrkamp.
Haddadi, Afsaneh/Kurt Sundermeyer (1996): Belief-Desire-Intention Agent Architectures, in: N. R. Jennings (Hrsg.), Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons, S. 169-210.
Hewitt, Carl E. (1986): Offices are Open Systems, in: ACM Transactions on Office Information Systems 4(4), S. 271-287.
Hewitt, Carl E. (1991): Open Information Systems Semantics for Distributed Artificial Intelligence, in: Artificial Intelligence 47, S. 79-106.
Jennings, Nick R. (1996): Coordination Techniques for Distributed Artificial Intelligence, in: N. R. Jennings (Hrsg.), Foundations of Distributed Artificial Intelligence, New York u.a.: John Wiley & Sons, S. 187-210.
Jennings, Nicolas R. (1993): Coordination: Commitment and Conventions: The Foundation of Coordination in Multi-Agent Systems, in: Knowledge Engineering Review 8(3), S. 223-250.
Kant, Immanuel (1956): Kritik der reinen Vernunft, Hamburg: Meiner.
Krohn, Wolfgang (2003): Das Risiko des (Nicht-)Wissens. Zum Funktionswandel der Wissenschaft in der Wissensgesellschaft, in: Stefan Böschen/Ingo Schulz-Schaeffer (Hrsg.), Wissenschaft in der Wissensgesellschaft, Wiesbaden: Westdeutscher Verlag, S. 97-118
Lesser, Victor R./D. D. Corkhill (1983): The Distributed Vehicle Monitoring Testbed: A Tool for Investigationg Distributed Problem Solving Networks, in: The AI Magazine 4, S. 15-33.
Luhmann, Niklas (1975): Einführende Bemerkungen zu einer Theorie symbolisch generalisierter Kommunikationsmedien, in: Niklas Luhmann (Hrsg.), Soziologische Aufklärung 2. Aufsätze zur Theorie der Gesellschaft, S. 170-192.
Luhmann, Niklas (1984): Soziale Systeme. Grundriß einer allgemeinen Theorie, Frankfurt/Main: Suhrkamp.
Luhmann, Niklas (1990): Die Wissenschaft der Gesellschaft, Frankfurt/Main: Suhrkamp.
Luhmann, Niklas (1997): Die Gesellschaft der Gesellschaft, Frankfurt/Main: Suhrkamp.
Marsh, Stephan (1994): Trust in Distributed Artificial Intelligence, in: Eric Werner (Hrsg.), Artificial Social Systems, 4th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW '92, S. Martino al Cimino, Italy, July, 29-31, 1992, Selected Papers, Lecture Notes in Artificial Intelligence 830, Berlin u.a.: Springer, S. 94-114.
Martial, Frank von (1992): Coordinating Plans of Autonomous Agents, Berlin u.a.: Springer-Verlag.
Marx, Karl (1971 <1890>): Das Kapital. Kritik der politischen Ökonomie, Erster Band. Nach der vierten, von Friedrich Engels durchgesehenen und herausgegebenen Auflage, Hamburg 1890, Berlin: Dietz Verlag.
Mead, George Herbert (1967): Mind, Self, and Society. From the Standpoint of a Social Behaviorist, 14. impr., Chicago u.a.: Univ. of Chicago Press.
Menger, Carl (1969 <1883>): Ueber das exacte (das atomistische) Verständnis des Ursprungs jener Socialgebilde, welche das unreflectirte Ergebnis gesellschaftlicher Entwickelung sind, in: Carl Menger (Hrsg.), Untersuchungen über die Methode der

Sozialwissenschaften und der politischen Ökonomie insbesondere, Gesammelte Werke, hrsg. mit einer Einleitung und einem Schriftenverzeichnis von F. A. Hayek, Bd. 2, 2. Aufl., Tübingen: Mohr, S. 171-183.

Merton, Robert K. (1957): Priorities in Scientific Discovery, in: American Sociological Review 22(5), S. 635-659.

Merton, Robert K. (1968): The Matthew Effect in Science, in: Science 159(3810), S. 56-63.

Parsons, Talcott (1963a): On the Concept of Influence, in: Public Opinion Quarterly 27(1), S. 37-62.

Parsons, Talcott (1963b): On the Concept of Political Power, in: Proceedings of the American Philosophical Society 107(3), S. 232-262.

Parsons, Talcott (1975): Social Structure and the Symbolic Media of Interchance, in: P. Blau (Hrsg.), Approaches to the Study of Social Structure, New York, S. 94-120.

Parsons, Talcott et al. (1951): Some Fundamental Categories of the Theory of Action. A General Statement, in: Talcott Parsons/Edward Shils (Hrsg.), Toward a General Theory of Action, Cambridge Mass.: Harvard University Press, S. 3-29.

Rosenschein, J. R./M. R. Genesereth (1988): Deals Among Rational Agents, in: Les Gasser (Hrsg.), Readings in Distributed Artificial Intelligence, San Mateo, Ca.: Morgan Kaufmann Publishers, S. 227-234.

Schulz-Schaeffer, Ingo (2000): Vergesellschaftung und Vergemeinschaftung künstlicher Agenten. Sozialvorstellungen in der Multiagenten-Forschung, RR 3, Hamburg: Technikbewertung und Technikgestaltung, TU Hamburg-Harburg, Research Reports

Schulz-Schaeffer, Ingo (2001): Enrolling Software Agents in Human Organizations. The Exploration of Hybrid Organizations within the Socionics Research Program, in: Nicole J. Saam/Bernd Schmidt (Hrsg.), Cooperative Agents. Applications in the Social Sciences, Dordrecht u.a.: Kluwer Academic Publishers, S. 149-163.

Schulz-Schaeffer, Ingo (2002): Innovation durch Konzeptübertragung. Der Rückgriff auf Bekanntes bei der Erzeugung technischer Neuerungen am Beispiel der Multiagentensystem-Forschung, in: Zeitschrift für Soziologie 31(3), S. 232-251.

Shoham, Yoav (1993): Agent-oriented Programming, in: Artificial Intelligence 60(1), S. 51-92.

Shoham, Yoav/M. Tennenholtz (1992): On the Synthesis of Useful Social Laws for Artificial Agents Societies (Preliminary Report), in: (Hrsg.), AAAI-92. Proceedings Tenth National Conference on Artificial Intelligence, Menlo Park, Ca. u.a.: AAAI Press/The MIT Press, S. 276-281.

Spiegel Online (2003): eBay-Sicherheitsloch: Wie ein Träumer das Bewertungssystem aushebelte, in: Spiegel Online, 19. Februar 2003, 9:08, URL: http.//www.spiegel.de/netzwelt/netzkultur/0,1518,236673,00.html.

Wellman, Michael P. (1992): A General-Equilibrium Approach to Distributed Transportation Planning, in: Peter Szolovits (Hrsg.), AAAI-92. Proceedings of the Tenth National Conference on Artificial Intelligence, Menlo Park u.a.: AAAI Press/The MIT Press, S. 282-289.

Zuckerman, Harriet/Robert K. Merton (1971): Patterns of Evaluation in Science: Institutionalization, Structure and Functions of the Referee System, in: Minerva: A Review of Science, Learning and Policy 9, S. 66-100.