Available online at www.sciencedirect.com
SCIENCE DIRECT
Cognitive Systems Research 7 (2006) 307–323
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

The cognitive and behavioral mediation of institutions: Towards an account of institutional actions

Action editor: Ron Sun

Luca Tummolini *, Cristiano Castelfranchi

University of Siena, Piazza San Francesco 1, 53100 Siena, Italy
Institute of Cognitive Sciences and Technologies (ISTC-CNR), Via San Martino della Battaglia 44, 00185 Rome, Italy

Received 1 April 2005; accepted 7 November 2005
Available online 30 March 2006

Abstract

The aim of this paper is to provide an analysis of institutional actions from the standpoint of cognitive science. The notion of *constitutive rules* have been proposed to describe the conceptual nature of institutions. In this paper it is extended to cover specific processes of 'recognition' that provide the agents with additional *artificial powers*. The power of doing an institutional action is considered as a special kind of artificial power. It is argued that institutional actions achieve their effects thanks to a *cognitive* and *behavioral* mediation of a collective of agents. Individual actions are *seen* and *treated as* (count as) institutional actions by the involved participants even if, in fact, institutional actions are collective actions. When human behavior becomes institutionalized, it acquires special *conventional powers* to bring about effects in the social world. A model of such conventional empowerment of an agent is proposed and is identified in a sort of collective permission. Finally it is argued that institutions are a specific kind of coordination artifacts. In particular, the importance of institutional roles as artifacts that assign conventional powers is investigated.

Keywords: Coordination; Institutions; Institutional actions; Power; Collective behavior; Multi-agent systems; Constitutive rules

1. Introduction

Institutions are usually conceived as normative systems that structure social interactions. It is especially in the economic literature that institutions have been scientifically approached by means of the game-theoretical apparatus to provide models of how institutions can evolve from the independent interactions of individual agents. A general property of economic models is to focus on institutions as the 'rules of the game' (North, 1990), the set of constraints that evolve (or are centrally issued) to regulate agents' interactions. Part of these rules are in the interest of the individual agents themselves (as in the case of coordination games, Lewis, 1969) while others are needed to solve cooperation dilemmas that, if left to individuals, would not be solved (as in the prisoner dilemma or in mixed coordination games).

While different ways of modeling the normative component characterize different disciplines, its centrality in the understanding what institutions are, is undisputed.

However it is part of a renewed interest in the foundations of institutions in social philosophy to stress also their intrinsic conceptual or constitutive nature (Searle, 1969, 1995; Tuomela, 2002). What is specific to institutions (as opposed to mere regulating conventions) is that they are also defined by constitutive rules (Rawls, 1955; Searle, 1969). These rules create a new level of activities by defining that "*X* counts as *Y* in context *C*" as in the case of "this piece of paper *counts as* 'money' in Europe". By regulating this new level, institutions constrain and influence the concrete practical actions of the agents.

* Corresponding author.
E-mail addresses: luca.tummolini@istc.cnr.it (L. Tummolini), cristiano.castelfranchi@istc.cnr.it (C. Castelfranchi).

1389-0417/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2005.11.014

For example, many economic models have been proposed to explain how the institution of money can evolve (see Hodgson, 2002 for a review). However these models assume that the agents are already acquainted with all the practices that will be regulated by the evolved institutions. In general, what is missing is that the most basic function of considering something as money, is precisely to enable the agents to do something new which is *paying, pricing, and saving* and whatever we *can* do inside the institution of money. Moreover, it is a characteristic of these institutional actions that have normative consequences, viz. if I have paid for this commodity, I have the "right" to claim its use (Searle, 1969).

As it is generally acknowledged, an institution is a solution to coordination problems of a collective, but what seems to be special in the case of institutions is that such coordination is obtained thanks to the *constitution* of a new level of actions that can be done. The coordination is mainly achieved with the creation of deontic 'enablements and requirements' that are the opportunities and constraints that influence the agents' interactions. In this paper, we will try to disentangle the cognitive and behavioral mediation of institutions. Institutions will be seen as a specific kind of 'coordination artifacts' that is man made products with the *function* of coordinating a collective of agents. Their peculiarity being that they achieve this result by means of deontic mediators that enable multi-agent actions. While we acknowledge the fundamental role of the deontic dimension, in this paper we are particularly interested in the *conceptual* one (the mediating role of conceptual schemas). Hence the institutionalization process will be considered as a specific kind of conceptualization. In agreement with Searle (1995), we will consider institutional actions (i.e., the action of 'paying') as prior to the institutional objects (i.e., 'money') and so we will provide an account of how this kind of actions are constituted. Our main thesis will be that institutional actions are always multi-agent (or collective) actions. Finally, we will provide an account of how an individual is empowered by the collectivity in executing an action which is a collective action (*conventional power*). Such empowerment, it will be argued, is due to a form of unintentional collective permission.

2. The cognitive nature of constitutive rules

Much of the contemporary philosophical debate on the nature of institutions has a declared ontological aim. It is claimed that institutional facts like 'being the president of Italy' or 'being married to Mary' exist in the world but are different in their ontological status from brute facts like 'being a mountain' or 'being a water molecule' (Smith & Searle, 2003).

Differently, our interest is not so much in the ontology of social reality (how social and institutional facts exist) but in modeling how institutions are constructed and conduct their affairs through the minds and the actions of the involved agents (how institutions work) (Conte & Castelfranchi, 1995). The seminal work of John Searle is somewhere in the middle and, as a matter of fact, it has inspired authors across many disciplines. We agree with Searle that there is a 'primacy of the micro-level' where the individual agents *constitute* the institution by considering something *as* something else.

2.1. Constitutive rules as triadic relations

Rawls (1955) has been the first to introduce the distinction between two different conceptions of rules. The summative conception of rules refers to rules that emerge or are issued in order to regulate already existing actions. The practice conception, differently, relates to rules that create the possibility for a new action by creating a new description for the action. This second notion of rule has been properly named by Searle *constitutive* (Searle, 1969).

From the perspective of a cognitive scientist, rules of the kind "*X counts as Y* in *C*" seem to regulate a cognitive activity, viz. the proper application of a concept. In other words, a constitutive rule describes, albeit very abstractly, a 'recognition' process. Because such rules are used to describe the constitutive nature of institutions, the institutionalization process turns out to be a specific case of conceptualization of an entity in the world.

The application of a concept in fact can be represented in form of a rule that associates a specific set of stimuli ('something such and such') *X* with a linguistic label *Y*. This model however is too simplistic also for an abstract account because it does not properly identify the underlying cognitive mediation. The *Y* term in the relation collapses two different entities: the Cognitive Type[1] (CT) and its label. A more appropriate formula to express such a relation is that "*X*, *seen as* a token of a CT, counts as *Y*, in *C*" (see Fig. 1). *Counts as* relations are *triadic relations* where the set of stimuli *S* are interpreted through a conceptual schema or cognitive type CT, and such a schema can also be associated with a linguistic label.[2] The relation between the stimuli and the schema is a *token-type* relation.

2.2. 'Institutionalization' is also a kind of conceptualization

It is a possible mistake to treat counts as relations between two terms as a signification process between two

[1] We borrow from Eco (1997) the expression *Cognitive Type* to refer to the set of representations that characterize a specific type. As emphasized already by Johnson-Laird (1983) such representations can be of different formats, from images, to propositions, to sensorimotor ones.

[2] In our account, what really matters is not the *label* but the *concept*. It is the concept that gives meaning to the stimuli and we react to this meaning. One might claim that a label is necessary for building a concept. However this is another issue. The label is also necessary for having the concept more or less shared in a community and for the 'negotiation' process about our coordinated cognition. For a computational model of the reciprocal influence between conceptualization and labeling see Mirolli and Parisi (2005) and for the sharing of categories by means of label use and communication see Steels (2003).

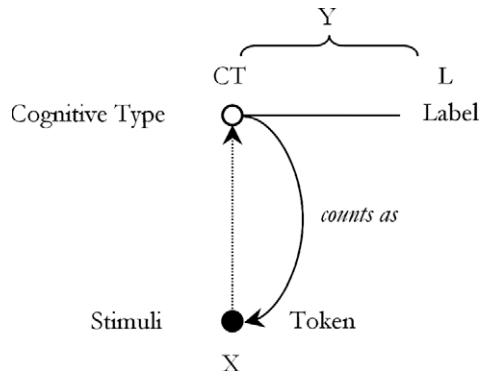

Fig. 1. The triadic *counts as* relation as a simplified model of recognition.

different entities, i.e. two distinct actions. Take this example from Jones and Sergot (1996): in a department the signature of the secretary (*X* term) *counts* as the signature of the boss (*Y* term). While this a perfect possible case of institutional action that should be appropriately modeled, it is quite different from the relation between the action of writing his own name in certain conditions done by the boss that counts as 'signing'. When something *stands* for something else, there are two distinct actions and one signifies the other. It is often argued that the agents in the relation can be the same. In such a case, one of the actions in the agent's repertoire signifies another and the two are distinct actions.

However there is a different form of signification such as the one of typical of perception. Perception is always a matter of inference. It is at least since the late forties (Bruner & Goodman, 1947) that it is argued that perception is "necessarily the end product of a process of categorization [...] in which organisms move *inferentially* from cues to category identity". Such a 'perceptual semiosis' is something different from the signification process where something *stands for* something else. The 'smoke' stands for the 'fire' only after having been recognized. Once a perceptual judgment is drawn, it can be propositionally articulated so that also the presence of 'fire' is derived. The two judgments are always distinct (see Fig. 2).

There is perceptual semiosis when a perceptual judgment is inferentially drawn from something *to the same something*, and not to something else (Eco, 1997). The relation between a *token* and a *cognitive type*, when some stimuli are interpreted through a pre-existing conceptual

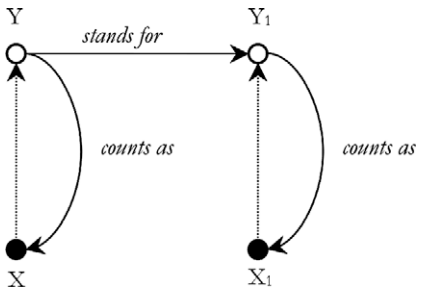

Fig. 2. A simple model of perceptual semiosis and of signification.

schema, hence when they are *recognized*, is an inference. A perceptual inference is the process of recognizing something as a token of a specific type. In this relation we say that the token *counts as* the type.

In Section 3, we will propose our theory of action and we will argue that an action is always a (supra-)action composed of a vehicle action and an external delegated event. The relation between the two is that the latter (the vehicle action) *is part of* the former (the supra-action) and between the two there is a cognitive mediation just as the one sketched above.

An institutional action is recognized when an *X* is *considered as* a token of the type *Y*, as when involved participants see the action of writing one's own name on a paper as the institutional action of 'signing'. The institutionalization process establishes a code that specifies how an action in certain context should be interpreted or, equivalently, establishes the sufficient conditions for the application of institutional concepts.

So far so good but to say that there is signification in the process of institutionalization does not seem to be enough for an adequate explanation. From our account it follows that a similar code is established every time that a stimulus is recognized.

To emphasize the peculiarity of institutional actions, Searle argues that by defining that "objects that are designed and used to be sat on by one person *counts* as chairs" (Searle, 1995), we do not adopt a constitutive rule because we are not ascribing the status to an object and, with it, a function. The function of the chairs depends on their physical features that are there independently of any human agreement. However, at the same time, Searle argues that functions are never intrinsic and always observer relative. In his theory, functions are always ascribed by humans to the external material world.

In the remaining parts of this section we intend to provide an alternative account of the functions of artifacts in order to stress more the similarity with institutions than their differences. The reason why this is important for our analysis will be clarified in Section 5.

2.3. The double empowerment of tools and artifacts

In Conte and Castelfranchi (1995, p. 125) a precise notion of 'function' has been introduced. A function is "an external goal placed on a system that results in a *transformation* of the structural properties of the system". The external goal is usually internal to another goal-governed agent, but what is specific relative to weaker notions like 'use' and 'destination' is that the physical characteristics of the system are caused by the external goal.

Even having stated that the goals modify the shape of the system, if we assume that external goals are always internal to some other agents then our definition could be considered quite similar to the one advanced by Searle. However, we also claim that there exist external goals that are not internal to any agent at all.

Examples of these stronger notions of external goals are the biological and social finalities or functions (Castelfranchi, 2001; Conte & Castelfranchi, 1995). Finalities are selecting effects that modify an organism's characteristics in a way that will render it more *adaptive*.

Let x be an entity instantiated in a sequence of distinct repetitions $(x_1, \ldots, x_n)$. A sequence of repetitions is defined as a set of occurrences of the same entity linked in such a way that each is produced by the preceding, and produces the following occurrences (if any) in the sequence, thanks to whatever mechanism of reproduction. Let B_x be the set of behaviors or characteristics of x, *a finality* or *function* can be defined as follows:

(1) some items in B_x produces effects unintended by, and unknown to, x;
(2) any item in x that produces the unintended effect is *functional*, if that effect acts through *a causal feedback loop* on the mechanism of reproduction, favoring x's reproduction, and as a consequence, that of the item themselves.

Such an effect is no longer a simple one among the others but is a *finality* of the behavior or characteristic in question. It has selected and shaped that behavior or morphology to be as it is. It needs a set of reproduction mechanisms and a feedback mechanism to select some variations or to reinforce the corresponding behavior.

The set of reproduction mechanisms can be the most various from the anticipatory mental representation of a future effect of one's own action, to reinforcement learning mechanisms, to natural or artificial selection.

Our notion of function is intended to be general enough to cover both functions that originate from intentional behavior of agents like designers of artifacts (the external goal on the artifact is internal to some agent) and the biological and social finalities (where the external goal is not internal to any agent at all).[3]

Consequently, since our notion of function is relative to a *goal*, we need also a more general notion of it beyond the mental one. A goal can be defined as a *sieve* used to select the (morphological or behavioral) properties of a system such that at time t_1 such properties are not the result of chance, but of its preceding history (t_0) that consists in the rejections resulting from the sieve sifting the alternatives of a given property.

Adopting this general definition, that covers both internal and external goals, we argue that functions are not observer relative at all.[4] We claim that artifacts have 'intrinsic' functions that have shaped their characteristics.[5]

However, if we focus only on artifacts that must be employed by an agent to achieve their function (not automatic or autonomous), a sort of dependence on the involved agents is still there.

Something is *usable* as a chair even if it is not *recognized* as a chair, however this possibility of being used as a chair will never be actualized if somebody does not recognize it as a chair. In this recognition process, the agent needs to infer a possible *use* of the artifact. Notwithstanding its intrinsic function, if the agent does not acknowledge its use, the artifact seems to lack the status of being fully a chair.

The 'use' of something is a weaker notion relative to the notion of 'function'. We consider the 'use' as the simplest notion of external goal. Even if something does not have a function, it can still have a use. When an agent has a goal to achieve, and something 'can' give him the power of achieving it because of its existing (morphological or behavioral) properties, then it has a use (and that something becomes a *tool*).

Let B_T be the set of behaviors or characteristics of T, T has a use U if and only if there exists at least an agent x with a goal G such that a non empty sub-set of B_T is a sufficient condition for x to achieve G.

We say that such a tool (and of course a proper artifact) provides the agent with the *physical power of* achieving at least one of his goals (for a recent analysis of power relations see Castelfranchi, 2003).

However, to effectively have the power of achieving the goal, even an artifact with a function needs to be recognized as *usable* which means that the agent is able to infer how the artifact is supposed to be used (rules of use). In this sense, every artifact, to physically empower

[3] According to Vermaas and Houkes (2003), this account could be classified as a *non-intentionalist reproduction etiological theory*. Vermaas and Houkes argue that, as far as the ascription of function to technical artifacts is concerned, etiological theories fail in their account. While there is no space to argue for the adequacy of our theory to their requirements, we think that our account is tenable and left a precise defense for future work. They also claim that reproduction theories must be intentionalist to account for artifacts. However, if we consider institutions as artifacts it is clear that, while they are man made 'social objects' they need not necessarily be intentionally created.

[4] With this definition of 'goal' as a 'sieve' we are able to cover also the case in which a novel artifact is invented by a solitary designer (the fourth desideratum in Vermaas & Houkes, 2003). In this case, the anticipatory representation of its intended use is a sieve that selects the physical characteristics (means-end reasoning) among a set of possible anticipated alternatives. Then, once a model is chosen, the continual iteration of the test between the intended structural characteristics and the actual ones represents the 're-production mechanism' by adapting the building actions. Actually the artifact is first produced in the designer's mind and then re-produced in the physical reality.

[5] For the aim of this paper the actual account of functions of artifacts is enough. However it is of course a very strong simplification to state that the function of an artifact is the goal of the designer had in building it, and that only such goal has causally modified its morphology. The history of technology is full of cases in which the function of an artifact has emerged after cycles of trials and errors of social use (Basalla, 1988; Ziman, 2000). A proper model needs to explain this mechanism in which the social 'destination' of an artifact to a recurrent use evolve in a new 'function'. For the technical notion of 'destination' in this framework see Conte and Castelfranchi (1995).

the agent, always depends on the agents that are evaluating it.

How agents reason about the use of tools and artifacts, how they ascribe a use, is an important issue both in cognitive science and cognitive neuro-science. An important debate divides those giving a central role to the perception of affordances, which actions are 'afforded' by the object (Gibson, 1979), and those contending the prominence of the ascription of the intention of the object's creator. For the aim of this paper it is not necessary to propose a model for this kind of reasoning.[6] What is relevant is that when something is considered as a *chair*, this recognition activates appropriate learned motor schemas. Chao and Martin (2000), for example, show that tool recognition activates the left ventral premotor cortex and that this is necessary for a proper categorization. The agents that recognize a tool or an artifacts react having at least a disposition to adopt the appropriate actions. Those motor schemas are part of the cognitive type of such tools and artifacts (see Borghi, 2004 for a review).

From this perspective, to say that "objects that are designed and used to be sat on by one person *counts as* chairs" is something more that a simple description. A sort of empowerment is present also when agents appropriately recognize a tool or an artifact. Conceptualization of artifacts is a form of empowerment.

We can consider this process as a double empowerment. By recognizing artifacts and tools as usable – somehow acknowledging their 'rules of use' – (1) and by using them (2) agents become able to produce new physical effects in the world. Because the former kind of empowerment is mediated by the agent itself we consider it a process of *artificialization*. Even if the process is individual, it is the agent that produces it by means of his (individual) cognitive mediation, and so in the end he acquires additional *artificial powers*.[7]

In this paper, the scope of *constitutive rules* will be extended to the establishment of the sufficient conditions for the applicability of concepts that somehow artificially empower the agents.[8] A more appropriate term for our purposes would be *considering as*. When somebody considers a X as a Y, he *sees* and *treat* it *as* a Y. This double cognitive and behavioral mediation is of fundamental importance for our account.

This process is very similar to the Searlian idea that in constitutive rules there need to be a function that is there because a status has been recognized. Besides our account of tools and artifacts is very close to the idea that there is a primacy of actions over objects. The fact that "the object is the continuous possibility of the activity" (Searle, 1995) applies similarly well both to institutions and to tools and artifacts more generally: the knife exist only relative to the need of enabling 'cutting' actions, and such actions exist only thanks to the artifacts.

To the eyes of the agents, the tie between the artifact and the enabled actions is so strict the ordinary language allows expressions such 'the goal of the knife is to cut'. This common conception treats the action as intrinsic to the artifact. 'Cutting' is something that pertains to 'knifes' *as if* it were a power of the artifact to cut and not of the agent using it. In Section 5, we will argue that a similar mis-attribution is present also in the recognition of actions done in institutional roles.

However we agree with Searle that the basic mechanism that is present in institutions is slightly different from the one sketched above for usual tools and artifacts. The power of achieving the function is not simply relative to the physical features of the term X and the artificial power added by recognition of a use by the agent. Something more is needed, a specific form of *artificial power* (*conventional power*) must be provided to the agents that act in the institution. In Section 4 we will advance a model of this conventional power based on a form of collective permission.

To conclude, our claim is that to institutionalize something is to assign a *specific* kind of artificial powers to it by means of a conceptualization that can be expressed by counts as formulas.

3. From physical to institutional actions

Even if we have claimed that we are not inspired by an ontological aim, a sort of ontology is always implied when advancing a conceptual analysis such ours. To make it explicit, our ontological statement is that institutional actions, like paying, marrying, promising, having precedence and so on and so forth, are concrete physical actions in the physical world. What is special is that (1) they always are *collective actions* or *multi-agent actions*. What makes them different from other collective actions like moving a table, cooking or dancing together is (2) a specific *cognitive and behavioral* mediation necessary for their success. Moreover we will argue that (3) there is also a specific coordination function which is absent in collaborative activities oriented to the achievement of a usual practical goal. In what follows we provide a model of institutional actions. We will characterize different actions from the most basic physical ones to the social, communicative, and collective ones. In the end we will

[6] For a model that integrates these approaches see Barsalou, Sloman, and Chaigneau (2003).

[7] The individualistic characterization of this process is not of course the only possible. If we assume that an agent can try to understand what is the intended use of an artifact (what the designer intended it for) this process become a social empowerment. In fact there is a social dependence between the designer and the user and the former one has sort of *power over* the latter. By designing the artifact so that it is more easily usable the former is 'practically permitting' to the latter the action. For the technical notion of *power over* see Castelfranchi (1990, 2003).

[8] That the counts as formulas express sufficient conditions for the applicability of concepts is a statement that is common also in other approaches (Jones & Sergot, 1996; Pörn, 1977). However these authors do not share the artificial empowerment condition. As we will see in Section 3 this is not limited to the recognition of tools and artifact.

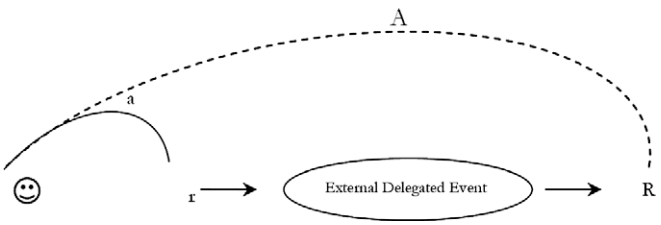

Fig. 3. A model of physical/practical action.

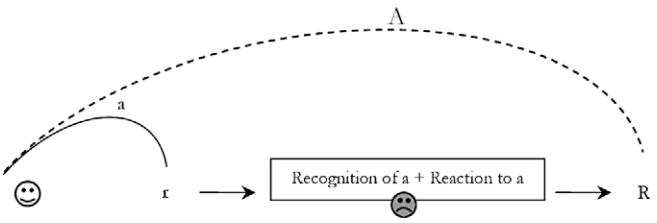

Fig. 4. A model of 'scandalizing'.

argue that institutional actions are a specific case of collective actions done by a group.

3.1. Physical actions

In our model an action A is composed of a vehicle action a and of an external delegated event E on which the agent relies to achieve its intended results. The action A can be named *supra-action* to distinguish it clearly from the vehicle one.[9] However to avoid awkward technical expressions in the rest of the paper we will refer to supra-actions simply as actions with a uppercase literal to identify them.

A classical example is that the agent intends to turn on the light in the room. This action is done by doing the vehicle action of flipping the switch. In our model the agent relies also on some external events such as the fact that the electric circuit once open will bring it about that the light is on (see similarly also Pörn, 1977).

The vehicle action a is a sequence of bodily movements done intentionally, in the example it is the sequence of movements done by the agent to flip the switch. The result r of these movements brings it about a specific external event (the circuit is open). Such an event is a causal complex that contributes to the achievement of the result R of the action A (the light is on). We say that such an event is delegated because the agent is relying on the event in order to achieve the intended result R (Castelfranchi, 1998). The agent intends to do A, by *intending to* do a and *intending that*[10] such event E obtains, in order that R is achieved (see Fig. 3).

The example of turning on the light can be misleading because it seems to be a complex case relative to more basic actions. However our claim is that for every action done in a context the agent always relies on the external world. The example is simply used to offer a clear case where the relation between the vehicle action and other events is evident.

A minimal model of action implies that the agent believes that his action A is necessary and sufficient to achieve R (while the vehicle action is only necessary) and that he can obtain R by doing A.[11]

This second belief can be named *Power of* belief. To do A the agent needs to believe he has the power of achieving his goal R by doing his action A. Having such physical power means that the agent is *able* and *in condition* to achieve R by doing A. The theory of power is complex and very articulated. In Castelfranchi (2003) such a theory has been proposed. For the aims of this paper it is sufficient to recall that the *subjective* power (believing to have a power) is, for cognitive agents, a necessary condition for the *objective* one. An agent has an objective power when, given his goals, he has the internal (skills, motivation, etc.), and external resources (tools and artifacts or whatever) and the conditions for action are such that if he acts, he achieves those goals.

However, if he, for any reason, wrongly believes that he lacks the power of A, then he will not be able to intend to do A. Consequently, also his objective power somehow disappears.[12]

3.2. Simple social actions

The simplest notion of socially mediated action is when what is needed for the action A to achieve its intended result R is simply an internal reaction (i.e. a feeling) of another agent (see Fig. 4).

An example of this kind of social action is 'to scandalize somebody'. To scandalize an agent has to bring it about a feeling of scandal in somebody else, viz. a feeling of being

[9] Another possible term could be *macro-action* but this term is used to refer to pattern of actions that are developed out of more micro-patterns, for example routinized procedure can be executed without planning of more basic or primitive actions (McGovern & Sutton, 1998). The macro-action is intended as a single action but is composed of more elementary building blocks not intentionally executed. In this sense every macro-action can be a vehicle action.

[10] Several authors (Sellars, 1967; Vermazen, 1993; Grosz & Kraus, 1996) have proposed that we should distinguish between at least two kinds of intentions: *intention to do something* and *intention that something be the case*. While it is common to say the one intends *to* go to the movie tonight, it is also possible to formulate the intention *that* you come with me. The real meaning of having an intention that something is the case is precisely that the agent believes that he has to do and can do something and that intends to do whatever it is necessary to make it the case that p. We extend this notion so that not only other agents but also events and artifacts can be objects of intention that.

[11] The same action can be the vehicle of different supra-actions (for example flipping the switch can be 'firing the bomb'), and the same action can be supported by different vehicle actions (for example I can turn the light on simply by snapping).

[12] For a initial formal model of this theory of power using an algebraic approach see Boella, Sauro, and van der Torre (2004).

provoked and offended by something which is against the accepted morality. An agent can scandalize without intending it and in this case this is not an 'action' done by the agent. However he can decide to intentionally scandalize an observer. If a nice girl Mary intentionally walks naked in her house knowing that an old nosy lady is looking at her, she intends to scandalize somebody. This means that she intends to do a vehicle action *a* (walking naked) to do an action *A* (scandalize the old lady). In this case, to achieve her intended result, the lady should recognize her action *a* and react internally to it (feeling the scandal). Notice that to achieve this result it is not necessary that the observer recognizes the action *A* that she is unintentionally contributing in bringing it about. The minimal contribution that the other agent has to do is to react to the understood vehicle action. If this recognition fails, the action of 'scandalizing' does not obtain. A similar action is also to caress somebody. It is possible in fact to caress somebody only if he can feel some sensations as a consequence of a gentle movement of an hand. It is part of the action of caressing that the other agent feels such sensations.

A slightly different case is that of 'handing something to somebody'. John handing a flower to Mary is not simply John bringing it about that Mary is close to a flower. If John with a movement of the arm brings the flower close to Mary, he has not handed the flower to Mary. To really 'hand' it to Mary, the vehicle action must be recognized by Mary has a token of the action concept 'handing', which means that Mary should acknowledge the intention of bringing it about that she has the flower by means of his movements. By understanding that it is John's intention that Mary is close to the flower, Mary is contributing to the accomplishment of action *A* (see Fig. 5).

Actions like handing necessarily rely on communication. There is a message in the practical action of bring it about the case that somebody is close to something when it is a case of 'handing'. By doing the practical action of making Mary close to the flower, John is also emitting a message without employing any conventional code.[13]

In this case, to bringing it about that *R*, it is a necessary condition that the involved agent recognizes in the vehicle action *a* the supra-action. *A* is artificially created also with her contribution. It is important to notice that the involved agent also believes that the other has the power of executing the action *A*.

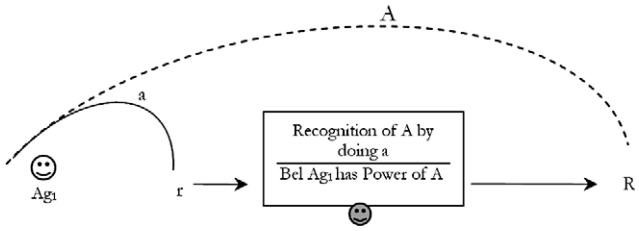

Fig. 5. A model of 'handing'.

3.3. Intermezzo: the co-power of agents

In more complex forms of socially mediated actions the recognition of the supra-action *A* (cognitive mediation) is necessary but not sufficient to bringing it about that *R* is the case. There needs to be also a behavioral mediation. In order that *R* might be realized (and so *A* might be executed), the involved agents should act. A typical case for this kind of actions is when agents act *together*. Consider the case of the supra-action *A* 'assembling a car together'. It does not make any sense of course to say that an individual agent is 'assembling the car together' alone. Acting together is an action done by a collective which means that each involved agent should do his *share* in *A* with appropriate attitudes so that *A* is done[14].

It is not necessary for this paper to provide a complete model of such actions[15] and arguing for or against a specific account of the required attitudes being them collective intentions (Searle, 1995), we-intentions (Tuomela, 1995) or shared intentions (Bratman, 1992). Whatever the most tenable account turns out to be, an intentional collaborative activity is by definition caused by the specific intentional attitudes of the participants towards the collective action *A*. In fact, the action of assembling the car together cannot be done by a single agent but it is done by the collectivity in question. None of the single agents has the power of doing the 'assembling the car together' alone, but the collectivity has it.

To define power of *A* in such a situation we need to introduce the notion of *co-power* (Castelfranchi, 2003). An agent can individually lack the power of assembling the car. But he *can* assemble it *together* with other agents. There is a co-power of assembling the car by means of the individual powers of doing one's own share. In this case, only cooperation (agents sharing the same end), or, more generally, the

[13] We name this kind of tacit and unconventional communication processes *behavioral implicit communication*. There is this form of communication whenever usual practical actions (like eating, walking, sitting etc.) or their traces (footprints) are intentionally done *also* for communicating something (making somebody else believe something) without employing any special codified mark. Think of the case of a scout leaving his safe footprints in a mined field for his fellows. For the general theory of behavioral implicit communication see Castelfranchi (2005) and for an initial formal model see Omicini, Ricci, Viroli, Castelfranchi, and Tummolini (2004).

[14] In the analysis of the concept of acting together it is usually avoided to consider joint-action concepts like kissing or quarreling where it is part of the concept that the action is done together (see for example Bratman, 1992). However it is important to notice that agents have developed action concepts that transcend their individual boundary of action and that directly refer to a collective. In particular, it will be argued that institutional action-concepts are solutions to coordination problems that embody multi-agent actions in single-agent action concepts.

[15] See Grosz and Kraus (1996) and Tuomela (1995) for two alternative detailed models.

combination of the individual powers in the same plan for the same end, allows the fulfillment of this end. This power is the *power of the Collective*. The collective can even have more power than the trivial sum of the powers of its members in the sense that the set of goals that the agents are able to achieve if they act together could be larger.

3.4. Complex social actions (intentional collaborative activity)

In collaborative activities, there is more than one agent executing the necessary vehicle actions $\{a,b,\ldots,n\}$ and each of these vehicles are part of the collective action A. However for each agent the actions of the others are part of the external delegated event on which each of them relies. Take the case of assembling the car together. The external delegated event here is both the assembly line that clearly has a role in the activity (i.e., to pass the semi-finished product to the next worker) and, each agent's share in the activity *seen as* a contribution to the accomplishment of A. The real vehicle actions when a set of agents is acting together are their *shares* in the supra-action.

In fact, a simple practical reasoning schema (I) for a cooperative collective action is:

(1) Ag_1 intends that R.
(2) Ag_1 believes that unless A is done, R will not be achieved.
(3) Ag_1 believes that unless everybody does his share of A, R will not be achieved (co-power belief).
(4) Ag_1 intends that everybody does his shares of A.

Is the fact that everybody does his own share equal to doing the actions $\{a,b,\ldots,n\}$? For us this is not sufficient. In fact each agent should do the necessary vehicle actions $\{a,b,\ldots,n\}$ as his share in A. To do a (painting the body of the car) as one's own share in A (assembling the car together) is to do a for a specific reason, that is because it is necessary for A.[16] But this is not enough. It is also necessary that the other involved agents *recognize* or *see* such action *a as your share in A*. To do one's own share is similar to the simple social action of handing in the previous example. It is a supra-action with identical structure. It is in fact a communicative action, when one is doing his share in a group activity is also emitting a message. If vehicle action b is 'painting the body of the car' and supra-action B is 'painting the body of the car *as a share* in assembling the car together', the agent can do B only if the others recognize B in b. We argue that the identification of this kind of supra-action is important for two distinct reasons.

The first is that while acting together, each agent delegates to the others their shares of the Multi-Agent plan, each relies on her fellow. As we have seen in schema (I) each agent *intends that* the others do their part in the plan, both in the sense that he can decide to help them if they are having troubles (Grosz & Kraus, 1996) and in the sense that he can reproach them if they disconfirm his expectation.

As in the case of the supra-action 'handing', each agent believes that the other agents have the power of doing the supra-action, viz. doing their shares.

However in our example of 'assembling the car together', each individual agent has the power of the vehicles $\{a,b,\ldots,n\}$ as shares but the power of A is a co-power of the collectivity and they are aware of it.

When acting together, each agent to do his share must recognize in the vehicle actions of the others that they are doing their share of A, viz., that they are acting together.[17]

In fact also the following schema (II) seems appropriate:

(1) Ag_1 intends that R.
(2) Ag_1 believes that everybody is doing $\{a,b,\ldots,n\}$ as their shares.
(3) Ag_1 believes that unless everybody does his share of A, R will not be achieved (co-power belief).
(4) Ag_1 intends to do his share a as a vehicle of A.

Assuming that each agent has the goal R, his belief that other are doing their shares is a reason to do his own. Each of them is aware that A is a multi-agent action, and that their own shares are necessary for the achievement of R. R is the artificial effect of the collective action of all the involved agents. The complexity of acting together lies in this complex cognitive and behavioral mediation that can be summarized in (see Fig. 6):

(1) the real vehicle actions $\{a,b,\ldots,n\}$ are the actions necessary to the achievement of A done as shares in A;
(2) hence such actions are supra-actions that need to be recognized to be realized;
(3) given the fact that the agents recognize the shares they have reasons to do their own shares in A;
(4) their acting accordingly to this recognition bring it about that R and consequently also A is accomplished.

Seeing the doing of one's own share as a cognitively mediated action is important also for a second reason. It allows in fact the minimal definition of what a group is for the aims of this paper.

[16] This instrumental characterization of the process is enough for describing a minimal case of 'acting together' without group identification or for groupness (Tuomela, 1995). This minimal case should correspond to the acting together in the I-Mode in Tuomela's taxonomy.

[17] The communicative function of the vehicle actions 'doing one's own share' is precisely to let all the agents know that they are acting together. The common knowledge that they are acting together is one of the reason to keep acting together until R is achieved. There is a communicative function and not a communicative intention because the agent does not necessarily realize that they are also communicating. But because they do so, they keep on acting and reproducing the mechanism.

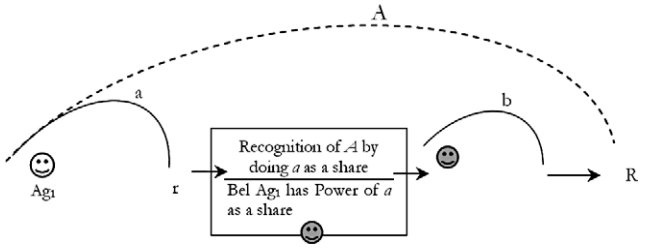

Fig. 6. A model of doing the action A together by doing a and b.

Given an agent Ag that is executing an action a, we define a *Group* at time t as the non empty set of all the agents G whose recognition of a at t and the action $a(\rho)$ at $t+1$ bring it about the case that the result R of the collective action obtains. We write $a(\rho)$ to stress the fact that the action is a function of the output of the recognition process, that is the action a associated to ρ. Each time an agent has the goal R and recognizes in the action of another one a doing of one's own share in the accomplishment of a supra-action A such that R is achieved, the two forms a minimal group.

This notion of group is however more general because it intends to cover also the kind of collective action typical of institutional actions which is different from the 'acting together' type.

The next paragraph is precisely devoted to outline a model of institutional actions.

3.5. Institutional actions as an unintentional collaborative activity

In the preceding sections we have shown that for some socially mediated actions, it is a necessary condition that the involved agents recognize in the vehicle action a the supra-action A. This is necessary for example 'to hand something to somebody else' or to 'assembly a line together'. Alternatively such a process of recognition could be described as the fact that 'a counts as A in C'. In our opinion, the description with the constitutive rule in such cases is particularly appropriate also against Searle's requirements. In fact it is only thanks to this cognitive mediation that the supra-action of 'handing something' is realized and achieve its intended results which also include appropriate beliefs in the participants' mind.

The case of 'assembling the car together' is more complex because it involves also a behavioral mediation. Agents should overtly act accordingly to the recognition of A in order to realize the effect that is necessary to identify the action A. What is special when agents act together is that they recognize their *Aing* during the execution.

Another difference between the simple and the complex case is that when somebody is recognizing a 'handing', he believes that the agent has the power of the supra-action. While in the case of 'assembling the car together' the agents only believe that each of them has the power of doing their own shares. As we have seen, they believe that A is a co-power of the group.

What is the peculiar mechanism of institutional actions instead?

Take the institutional action of 'marrying'. If Paul, the priest, is marrying John and Mary such action can be executed only if there is a set of agents recognizing it. The involved agents see the vehicle actions of the priest *as a token* of the Cognitive Type 'to marry' (cognitive mediation). While necessary, this condition however is not sufficient.

As in the case of 'acting together' for such an action to have physical effects in the world (and so to exist), it is also necessary that the set of involved actors *act accordingly* to this interpretation (*treat as*). Because they have recognized that the institutional action A has already occurred, they consequently believe the R obtains, viz. that John and Mary are married (mutual belief that R is true).

Either the belief that the supra-action has been accomplished or simply its reactive interpretation (disposition to behave), trigger the agent to *treat* the vehicle action a as A and so act on this basis (behavioral mediation).

All the involved agents (both the priest and the participants) have an expectation on the success of A (achievement of R), on their mutual recognition of A and on the consequent reactions of all the others. Moreover each of them also feels the expectations of the others on his own behavior, and all these expectations are reasons for the agents to behave as expected (Sugden, 1998; Castelfranchi, Giardini, Lorini, & Tummolini, 2003). By acting on the basis of their mutual expectations that R is true, they physically behave in the way that makes R true.

This acting accordingly is coordinated and produces the effects of action A. In the end the belief that the agent Ag_1 has the power of A is true and the expectations on the result and on the others are validated. This is the mechanism of self-fulfilling prophecies (see Fig. 7). As in the case of 'acting together' the supra-action is a multi-agent action that is done by the collective. Differently, all the involved agents believe that the agent that is executing the vehicle action has *also* the power of executing the supra-action.

There is a trivial interpretation for the claim that every institutional action is a multi-agent action. An opponent could easily argue that it is obvious that a wedding is a multi-agent collaborative activity. Many agents are covering many institutional roles (a priest, a husband, a wife, two best men, etc.), and all of them are necessary to achieve the intended results that agents John and Mary are

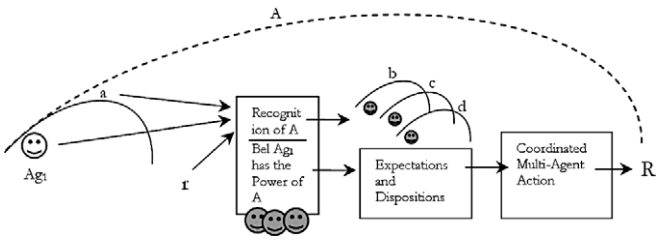

Fig. 7. A model of doing the institutional action of 'marrying'.

married. However this is only what the agents in the group believe that they are doing, it is their *subjective* interpretation of the event.

Our claim is that every individual action in the institution is actually a multiagent action where the collectivity unintentionally collaborate so that the individual agent achieves his intended results. Differently from the previous case, this collaborative action is not intentional. While during the ceremony each agent acting in the institutional role cooperates intentionally with the other, they are also unintentionally cooperating in a very different way.

No single agent in the collective has the power of executing the institutional action by himself. In a sense that will be clarified later, this power is given by all the others. Even without any acknowledgment of this process, the whole group is *helping* the agent in doing the action by doing the necessary actions that achieve the intended result.

In our account, the supra-action of marrying Mary and John is done by the whole group G of involved agents. So given an agent Ag that is executing a vehicle action a, we have defined a *Group* at time t as the non empty set of all the agents G whose recognition of a at t and the action $a(\rho)$ at $t+1$ bring it about the case that the result R of the collective action obtains. It is this group that is doing the institutional action, not only the agent Ag_1.

As it is clear the notion of 'group' in institutional actions is *local, context dependent and dynamic*. If an agent pays a sum to another in a desert the group is the collection of the two. If it does so in a shop all the involved participants that react in some way to this actions are part of the group. It is of course part of the group the agent himself that is executing the institutional actions.

Differently we can also define the *Community* C of agents as the non empty set of all the agents that *can* recognize that institutional action. Such a concept of community identifies all the agents that *potentially* can recognize the institutional action and that *potentially* can act accordingly. This means that a community is formed by all the agents sharing certain Cognitive Types of institutional actions.

From this, it follows plainly that a Group is a subset of the Community sharing that institutional Cognitive Type. The concept of Community, with its potential involvement in the institutional action, is crucial to understand how an institutional action can last much more than the simple time of its execution.

There are cases of institutional actions whose duration is very short. If, approaching a roundabout, I take the 'precedence' and you acknowledge this, I pass and from there on the action is completed. Differently, take the case of the priest that is marrying Mary and John. Being married is supposed to last forever. From our account it follows that such an action is done by the all the Group of the involved agents. In fact, if moving out of the church nobody acknowledges their roles, the marrying action has never occurred. The fact that a Community exists, viz. that any potential new bystander is able to recognize their being married and to react appropriately, is in fact a way of marrying them again and again.

4. Institutionalization is artificial empowerment by permitting

We have claimed that doing an institutional action A is a co-power of the group G. However the involved agents in G believe that Ag_1 has the power of A and for this reason, assume that A has been executed. From the subjective point of view, agents believe that one of them has such a power.

The cognitive and behavioral mediation of institutional actions is necessary to achieve the result R of an institutional action A. But because in the end the result intended by the single agent is true, a single agent has in fact the power of doing an action which is a multi-agent action. How this is possible?

4.1. The co-power of preventing an institutional action

To disentangle this problem we need, first of all, a minimal account of what a practical permission is. Face-to-face permission is a social relation between at least two agents x and y relative to a possible intentional action A of y. It implies:

(1) that y depends on x as for A (and x having power over y as for A);
(2) that x adopts y's goal (at least in a passive form: not preventing it);
(3) that there is a social commitment of x to y not to contrast y.

It creates 'rights' for y and correspondent 'obligations' for x. It empowers y enabling an action that was where impossible before. It requires also some either explicit or implicit communication (to ask for/to give) since it is based on mutual beliefs between x and y about the previous conditions.

It is crucial to notice that when x permits an action to y, x has a power of preventing y from doing A.

There are two clear distinct kinds of permissions that have been noticed in the literature[18] (Alchourron & Bulygin, 1981; Makinson & van der Torre, 2003) that in our terms can be stated as:

(1) Permitting by abstaining from preventing (passive permission or let).

[18] There is also a third kind which is permitting by actively impeding that an agent z negatively interferes with agent y. This example is a sub-case of consenting where the impediment is not already established but could be. This three party relation highlights the hierarchical structure between the agents (Bulygin, 1986).

(2) Permitting by removing an impediment (active permission or consent). Agent y is impeded or prevented from doing the action and agent x actively remove the impediment.

We have seen that relative to the institutional action A, it is the group of the involved agents that has the objective co-power of doing it. But because the action is intended by the single agent Ag_1 when they behaviorally mediate its vehicle action (acting accordingly to the recognition of A) they are giving him also the power of doing it. Since it is the group that *can* do the action, it is the group that has the co-power of preventing the single agent Ag_1.

The conventional/institutional empowerment is obtained thanks to a *functional* permission of the kind *consent*. The functional effect of the acting accordingly is that they positively interfere with Ag_1 reinforcing his *belief* of having the power of doing the action (Ag_1 believes that he *can do A*, he has the power of A). It is not simply the belief that he has the power of doing A that empowers the agent but also the fact that they act on this basis. This, together with the fact that the expected result is produced, is evidence for the agent that he has the power and so it provides additional strength to this belief. Moreover because everybody acts on the basis of their mutual belief that a single agent has the power of doing A is reinforced (see Fig. 8). In this model Ag_1 is the priest and Ag_2, Ag_3, Ag_4 are the involved participants. Before and beyond an institutional normative system, it is the participants that empowers the Priest in doing the action. If the participant Ag_4 does not act appropriately on the basis of the belief that the priest can marry Ag_2 and Ag_3 and that they are married as a result of Ag_1's action they in fact are not married because the priest actually lacks the power of marrying them. The most basic functional effect is precisely related to this propositional content "The priest is empowered to marry". All the participants believe and assume that he has such power but are unaware of the fact that they (as a group and a community) are empowering him by consenting him doing the action, by actively conforming to his expectations and so contributing to the achievement of the intended result.

To say that the conventional empowerment is the result of act of permitting done by all the participants seems strange because the involved agents are unaware of doing this action, of giving this power. However this basic mechanism is always present even in cases when the power is assigned by the normative system itself.

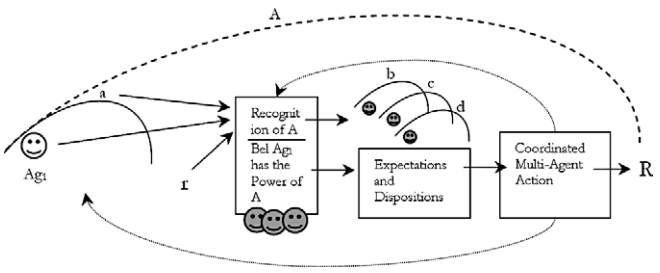

Fig. 8. A model of conventional empowerment.

Makinson (1986) has argued that 'having the power is not the same as being permitted to use it'. The case it the one of a priest that while having the power to marry two people is not permitted to do it when they are of different religions and one of the two is not converted to the relevant one. If the priest marries them, they are married even if he was not permitted to do it. Actually, for us, this example shows that there should be a conflict in the acknowledgment of the power of doing an action. It is not the place here to argue for an adequate account of this situation. It is sufficient to say that the most basic and fundamental empowerment is the one given by the involved actors. There is a complex hierarchy of levels which involve both formal and informal authorities. The formal authority has the conventional power of assigning conventional powers (the lawmaker). The informal one assigns these powers simply by their behavioral mediation, by acting accordingly to a given recognition (the involved participants).

The latter is actually more fundamental because it is the real support also for the first one. But due to the fact that *usually* the informal authority does not know of having such an objective power, it is not able to influence the formal one.

The mechanism is stable also because, as we have seen, it is a co-power of the involved participants not a single power of any of them. None of them individually by deciding not to accept the power of the priest is able to prevent all the other from empowering him.

4.2. The puzzle of unintentional institutional actions

A problem with our model could arise from the fact that we have only considered cases of *intentional* institutional action. We have tried to disentangle the complex collective activity that underlies a single institutional action. We have argued that the belief on the conventional power is essential for the agent to execute the institutional action. However, this model seems to apply to the case of intentional institutional actions only.

Differently, it is evident that there are a number of institutional actions that we perform that are non-intentional. Think of the case of murdering. Killing a man and murdering can be two very distinct kind of actions. Some instances of the first, in certain conditions, are cases of the second. Others are not, a soldier in battle is not murdering even if he is killing thousands of men. There are very different collective behaviors when we know that a specific soldier has killed an enemy (or even an innocent civilian) with the respect to the case that a killer has been discovered in our block.

However it is very common that in both cases the agents while intending to kill somebody *do not* intend to perform an institutional action that is nevertheless performed.

Since in our model for an action to exist a movement of the body must be intentionally performed, how is it possible that an unintentional institutional action does exist in the very first place? Moreover since an agent to perform an action needs the power to do it, how can such power be acquired if the agent does not intend to perform the institutional action at all?

The problem is even more deep if we assume that even the vehicle action can be done unintentionally. If I accidentally cause the death of somebody else, I am committing a specific kind of homicide. In this case the vehicle is not even an action of mine but, we would say, simply an event the I have caused. Nevertheless, a real institutional action has been performed.

Moreover, as we have seen an agent can acquire a conventional power of doing an institutional action A through a sort of collective permission of the kind consent. Participants do not give a permission but simply *permit* the action by their acting accordingly. However the most basic issue that still needs to be tackled is how, in fact, the involved agents come to believe that he already has such power in the first place. This belief is in fact necessary to activate the self-reinforcing mechanism. Solving this issue will be necessary also to address the previous problematic questions. To address these problems, an understanding of the *artifactual* nature of institutions is necessary. To this task is devoted the last section.

5. Why marriage is more similar to a table than a chair: institutions as coordination artifacts

It is a platitude to argue that institutions seem to be something external to the agents that enact them while actually they are man made products evolved or designed to coordinate agents' activities. In this last section, we will argue that institutions are artifacts of a specific kind, coordination artifacts, that exist and are maintained by the collaborative (almost always unintentional) activity of the collective.

Relative to other coordination artifacts institutions can be distinguished for a specific way in which they achieve this coordinating result.

5.1. Coordination in interference situations

Agents that live in a common world are agents whose actions can *interfere* one with the other. Interference is the most basic social notion in which the goals that an agent is willing to achieve are favored (positive interference) or hampered (negative interference) by the actions of other agents (Castelfranchi, 1998).

For the aims of this paper, coordination between the agents in interference situations is the combination of their actions so that a non-empty set of Goals is achieved. Coordination is not every arrangement of actions but the set of the 'good' ones. There is a value implicit in coordination that highlights that coordination is always evaluated against a set of goals.[19] Hence coordination in social interaction can be represented as the execution of a specific multiagent plan that solves an interference problem.

5.2. A definition of artifacts

As we have argued in Section 2, we claim that, given a population P of agents, A is an artifact if and only if the following conditions are satisfied:

(1) there exists at least one agent x in P such that x brings it about that A;
(2) there exists at least one agent y in P such that A has a use U for y;
(3) U is one of the motivating results for x to bring it about that A or y using A at t bring it about that x bring it about that A again at some point in the future.

In other words, we identify an artifact as a kind of result of an action, as something which is done (is brought about) by an agent x in order to (intentionally or functionally) be used by another agent y.[20] It is too general to identify as an artifact anything that is a man made product, a child is a consequence of the actions of her parents but she is not an artifact. It must be done for a use, so that if is done to be sold to some other people it acquires an *artifactual* nature.

This notion is also intended to be general enough to cover cases that often are not accounted for. In fact also traces of actions can be artifacts if they are done to be used by others, so that footprints left on a mined field for the followers are artifacts.

In the next paragraphs we intend to provide a first analysis of the use of artifacts for coordination purposes.

5.3. Tools and artifacts for coordination

Traces of actions can be the most basic kind of coordination artifacts. Consider the case of some people walking in a park. The unintended effect of their walking is to modify the shape of the grass in a way that is visible to others. Assuming that they prefer a beautiful meadow to one that is completely trampled, they begin following the traces of the others. After some time this self-reinforcing effect create a real 'path' which is obviously an artifact. Even if it is not designed explicitly by anyone, the function of this path is *also* to coordinate the agents.

If an artifact, given its physical properties, is used for coordination purposes but was not done for it, it simply

[19] The goals to be satisfied can be also the one of an external agent that designs the activity of others so that his goal is achieved. If the goals are those of the involved agents, such goals can be parallel and complementary (coordination in collaboration), opposite (coordination in conflict) or shared (coordination in cooperation). For a notion of parallel goals see Sellars (1967) and Conte, Miceli, and Castelfranchi (1991).

[20] Agents x and y can be the same agent.

has a coordination use. Relative to this use, it is a tool and not an artifact. It is very common that agents exploit the physical properties of artifacts to ease their coordination. Think of the case of Mary that wants to pour some wine to John. Mary is carrying the bottle whose function is to preserve the wine and to pour it. John has a glass whose function is to keep the wine to drink it. During the process of pouring the wine, Mary exploits some of the physical features of the bottle for coordination. The observable inclination of the bottle and its direction are used to communicate to John where to put the glass.[21] The function of the bottle is still to pour the preserved liquid but its physical shape is exploited for coordination. It is simply a use for coordination.

To really be a coordination artifact, it needs to have the function of (shaped for) coordinating the agents' actions. In the case discussed above, if the design of the bottle has been influenced by this possible use then it is a real function.[22] A table is in fact also a coordination artifact. Its physical shape has been designed to support things that are laid on it. This is the main practical use the table is made for (practical function). However, because more than one agent can eat at the same table, it must also be made for avoiding interference problems so that agents can sit at appropriate distances. Moreover this is evident also by the fact that tables are designed with many shapes. Rectangular tables convey a sort of hierarchy. The agents that sit at the heads of it are more visible by all the others and conventions can evolve such that in a family the head of the table is assigned to the head of the family. Differently, a circular table is considered more egalitarian in its disposition.

The coordination function that some artifacts can have is a specific kind of *social function* (Castelfranchi, 2001). Following the definition provided in Section 2, a social function is the set of effects of an action (intentional or not) that favors or hampers the goals of other agents and, *because of this interference*, somehow reinforces the mechanism that have produced those effects. It is a coordination function when the achievement of the coordination reinforces the actions that have solved the interference problem.[23]

[21] Communication through artifacts, like in the example, is a case of stigmergic communication. See note 13 for the notion of *behavioral implicit communication.*

[22] It is very difficult to discern mere 'uses' from 'functions' in concrete cases because one should be able to identify the precise design and use history of the artifact. However from a conceptual point of view the distinction is sharp.

[23] Our notion of coordination artifact is quite different from the notion of coordination mechanism advanced by Schmidt and Simone (1996). The interest in the use of artifacts for coordination purposes arose mainly in literature on Computer Supported Collaborative Work. However in these contexts, the artifact is mainly considered as a "permanent symbolic construct" that objectifies a pre-designed protocol of actions (p. 165). Examples of this kind of artifact are checklists and printed procedures. The role of artifacts that are not symbolic such as a table or a rotunda is completely underestimated. It is not the symbolic representation of a protocol that concerns us, but its materialization in the physical affordances (opportunities) and constraints.

5.4. Physical and deontic 'opportunities' and 'constraints'

As we have seen practical and coordination uses and functions very often coexist. Even if we have distinguished artifacts that have practical functions but only coordination uses (i.e. the bottle and glass) from artifacts that have both practical and coordination functions (i.e. the emerged path and the table), it is out of the scope of this paper to offer an exhaustive typology of coordination artifacts.

For our argument however it is useful to identify different ways in which such artifacts support the agents in achieving coordination.

Proposition 1. *There exist some artifacts such that their physical opportunities and constraints and the recognition of their use by an agent are necessary and sufficient conditions to enable a single-agent action.*

The use of every artifact is always related to 'opportunities' and 'constraints' that create sufficient conditions for action execution. As we have seen, in some cases artifacts also create necessary conditions to execute an action, such as the 'cutting' that is possible if and only if the agent uses a sort of knife. It is not only the fact that we have developed an action concept for an agent using a knife. What is relevant is that those specific physical effects, such as having slices of meat, are not possible without the artifact. *Neanderthals* could only tear the meat to pieces while *Homo habilis* acquired the *power* of cutting. Their actions repertoires were different. It is the case also of the chair that, once its use is inferred, provides the agent with the physical and the artificial power of 'sitting'.[24]

In any case when a tool or an artifact is used for coordination, the set of its physical characteristics is such that some of them are affordances (or opportunities, in the sense that they enable or facilitate the execution of some action) while others are constraints (they create obstacles or impediments to the execution of some action).

Proposition 2. *There exist some artifacts such that their physical opportunities and constraints are sufficient conditions to enable a single-agent coordinated action.*

The most basic mechanism by which an artifact can coordinate several agents is, in fact, by imposing physical enablements and constraints that define the sufficient conditions for action. Only a sub-set of actions is possible and by executing it the agent is coordinated with the others. As an example take the corridor of a corral. Once open, the animals flow coordinated out of it. The wall of a house keeps separated people in the house from people outside it because of its physical properties. These basic

[24] This enabling feature of tools and artifacts is often considered as a defining one. Tools and artifacts are means by which humans have expanded their influencing sphere. They not only help humans in doing better what they can do otherwise, but also *empower* them in doing what was not possible before.

mechanisms do not even need to be recognized to be 'used'. Actually agents do not realize of using something, they just adaptively react to external circumstances.

Proposition 3. *There exist some artifacts such that their physical opportunities and constraints and the recognition of their use by an agent are necessary and sufficient conditions to enable a single-agent coordinated action.*

In the case of the table whose set of physical opportunities and constraints and the recognition of the possible use the table is needed to coordinate the agents. Agents need to know how to use the table to actually achieve the coordination (viz. they should not try to sit at the same place). The recognition of the coordination use can be expressed in form of the technical norms that the agents endorse if they want to reach their goals.

Proposition 4. *There exist some artifacts such that their physical opportunities and constraints and the recognition of their use by an agent and the set of 'cognitive opportunities and constraints' (deontic mediatiors) are necessary and sufficient conditions to enable a single-agent coordinated action.*

There exist also artifacts that are the combination of different kinds of opportunities and constraints to achieve coordination. Consider the case of a roundabout as a prototypical one. Its physical properties are such that drivers must necessarily slow down when approaching it and can only go left or right. This is of course one of the main intended functions of roundabouts. However it is not enough. Once arrived at the roundabout, drivers need to choose the appropriate direction, otherwise they would crash one with the other. To do this they need to know the convention or norm in force. From the perspective of this paper, this entails that some deontic mediators are represented in the agent's mind. For example the agent is 'obliged' to turn right. Obligations are considered as mental constraints that limit available actions.

It is part of our research program the grounding of the deontic dimension in the physical one. We consider obligations, prohibitions, permissions and the like as mental constraints and opportunities. It is out of the scope of this paper to specify how these mediators arise in the agent's mind and how they are represented.[25]

Proposition 5. *There exist some artifacts such that the set of cognitive opportunities and constraints (deontic mediatiors) are necessary and sufficient conditions to enable a single-agent coordinated action.*

Finally, the artifact can be completely dematerialized as in social conventions and norms. A social convention to drive on the left or on the right exists independently of any material coordination artifact and can evolve by itself and into a social norm (Lewis, 1969; Castelfranchi et al., 2003). The convention or the set of conventions is for us a case of normative system that regulate the actions of the collective (everybody drives on the same line) by regulating the single actions of the individuals (drive on the left).

This features of conventions make them similar to procedures or scripts. While in the case of procedures and scripts one course of action is preferred to possible others and is used as a means to coordinating the actions, the specificity of conventions lies in their arbitrariness. When a convention is in place a specific course of action is expected, but there exists at least another equally good to achieve the goal (Lewis, 1969).

5.5. Institutions as coordination artifacts

Institutions are artifacts that are oriented to achieve coordination in a peculiar way.

Proposition 6. *There exist some artifacts such that the recognition of their use by an agent and the set of cognitive opportunities and constraints (deontic mediators) are necessary and sufficient conditions to enable a multiagent coordinated action.*

This is the case of institutions as a specific kind of coordination artifacts that make them different both from material artifacts and from conventions.

It is a common property of economic analysis of institutions to consider conventions and institutions as identical phenomena. Traditional and assumed definitions of an institution as "the laws, rules and conventions that give a durable structure to social interactions in members of a population" (Bowles, 2004) tend to reduce institutions to conventions. The emergence of the convention of driving on the left or the set of conventions that regulate traffic in a community does not make driving on the left an institution.[26]

The main difference between conventions and institutions lies in the fact that while a convention enables a single-agent action that contributes to a multiagent coordination (when acting according to a convention, the agent has the power of acting in coordination with others), an institution provides individual agents with the special conventional power of doing an action which is a multi-agent coordinated action.

[25] Artifacts and traces of actions can also be used to communicate such mental constraints and opportunities. Consider the case of a broom that hinders the entrance of a toilet. The broom is put there on purpose to create an impediment for people who wants to enter. By understanding this intention, a person can understand also the implicit imperative of the cleaner not to enter. In this way a double constraint is created and one is complementary to the other. The broom by itself would be an easily removable obstacle.

[26] Actually there is a continuum the goes from conventions, to practices and to institutions (Tuomela, 2002). And more importantly, convention in Lewis' sense can *evolve* in institutions. 'Taking precedence' is an example of institutional action that evolves from a 'driving on the left' convention. In order to outline the basic mechanism of institutions, here we are more concerned in the opposition of the two extremes.

This feature that seems to be absurd, is obtained through a *cognitive mediation* that is absent in mere conventions.

5.6. Institutional roles as coordination artifacts

Both 'property' and 'marriage' in a society coordinate actions, viz. the access to scarce resources. This is their ultimate function. As we have seen in Section 3, the coordination of actions of the involved participants is a necessary condition to achieve the result of an institutional action A. Their coordination is both a prerequisite and a consequence of institutional actions. However in the end of that section, we have also acknowledged that to get the institutional machinery off the ground it is necessary that the involved participants believe that there is one agent in the group that has the power of doing the institutional action (subjective power). This is again a coordination problem. There should be a mechanism to coordinate the assignment of conventional power. In this section, we have also claimed that institutions are coordinating artifacts. In fact, they are a complex artifact that should be analyzed as a functional structure whose different parts contribute to the overall coordinating function. Also a simple artifact as a hammer is composed of different parts (the handle, the stick, the head) that contribute to the function of hitting.

What are the parts in the institutions that contribute to the coordinating function? What is used and by whom in an institution?

In Section 1, we have claimed that institutional actions, like paying, are prior to institutional artifacts, like money. A theory of institutions should provide an adequate analysis of the role of artifacts such as money, scepters, signatures and the like.

Institutional artifacts are results of institutional actions that are done to be used by agents. For the time being, our claim is that such external artifacts achieve their coordinating function by interacting with the most basic coordination artifacts of all.

Our working hypothesis is that such basic coordination artifact is the *institutional role* that the agent is playing (the priest, the owner) and that the collectivity, even if unaware of doing this, is *using* for coordinating their physical actions.

Institutional roles *enable* actions that otherwise would not be possible. From this respect the role of 'owner' is an artifact similar to the 'knife'. By using the knife an agent acquire the power of cutting. Similarly, once an agent has the role of 'owner', he is somehow empowered to exclude all the others from the use of a resource.

But who exactly is using this role is not so trivial. Is the agent that does the action of excluding all the others or is the community doing it?

As we have seen in Section 2, there is a possible mis-attribution that is evident when we say that the goal of the knife is to cut. We are implicitly acknowledging to the artifact the power of the action that only the agent is able to execute. We argue that a similar mis-attribution is in place when we acknowledge to an agent playing an institutional role the power of doing the action. As we have seen, institutional actions are always collective actions whose power is always a co-power of all the agents. As in the case of usual artifact (see Section 2.3), the appropriate actions of an agent playing that role are part of the concept.

Turning back to the Searlian constitutive rules, we have argued that they can be used to describe a recognition process which is also a process of artificial empowerment. Searle uses the "X counts as Y in C" formulas both to describe the fact that "a certain agent counts as a priest" and the fact that "a certain movement of body counts as a baptizing". We agree that there is an artificial empowerment in both situations but we consider the former as a more basic one. When the involved participants recognize (*seeing and treating as*) the 'priest' in Paul, the output of this process is the belief (and the actual fact from there on) that Paul has the power of marrying John and Mary. This belief is then necessary to recognize in his vehicle action the institutional supra-action A. The cognitive and behavioral mediation of the institutional action is also an artificial empowerment. The vehicle action could not achieve the intended result without this artificial mediation.

5.7. A (preliminary) solution to the puzzle of unintentional institutional action

If we accept the fact that considering 'an agent' as 'an agent in a role' is a sufficient condition for the group of agents to assign conventional powers to one of them, then a possible solution to the puzzle of unintentional institutional actions seems to be within reach.

We can in fact claim that when an agent is playing an institutional role the group of involved agent in the community assumes he has the relevant conventional powers. The basic mechanism can get off the ground in the moment that there exists a community that shares the Cognitive Type of an institutional role. It is our hypothesis the institutional roles are macro-concepts that group a set of institutional actions.

Consequently, if a role is built (the concept is formed and shared) including intentional vehicle actions or vehicle events as vehicles of institutional actions then the power of doing an institutional action is created. In such strange cases, the behavioral mediation that is done by the collective now (*treating* Mary *as an* assassin), completes the supra-action she has done maybe years before (killing her husband John).

It is also worth mentioning that institutional actions being always goal oriented (there always is a coordination function) are actions under every respect.

6. Conclusions

This paper has tried to provide a conceptual analysis of how institutions conduct their activities *viz*. coordinate the agents' activities.

In form of a conclusion is useful to summarize the main theses that we have defended:

(1) the process of institutionalization of human behavior is a process of artificial empowerment;
(2) an institutional action is a socially mediated multi-agent action;
(3) while an institutional action is a socially mediated multi-agent action (viz. a collaborative activity) it is not executed on the basis of a shared plan;
(4) an institutional action is regarded as a single agent action and the agent believes he has the power of doing the action;
(5) an institutional action has artificial effects that are added by the collectivity so that the intended result is obtained;
(6) the main result that the action is achieving is the coordination result. This result is not necessary intended but is the ultimate end of the action (function);
(7) the institutional action can be done intentionally or not by the agent but it is a real action relative to the function of coordinating the collectivity;
(8) the institutional role is the artifact that assign the conventional power to the agent. The collectivity by recognizing the role and by acting accordingly enable the institutional actions because they wrongly attribute the power of the action to the role. This is why having the role is having the power;
(9) the conventional power is always a power of coordinating a collectivity of agents.

Hopefully, this conceptual apparatus will be useful in analyzing the way in which institutions act by means of their participants. However a different question is *how* and *why* a specific institution has emerged in the first place and by which specific process it is temporally maintained. Our conceptual apparatus will be used to approach these questions in further work.

Acknowledgements

We thank Andrea Omicini, Alessandro Ricci, Mirko Viroli, Guido Boella, Leon van der Torre, Raimo Tuomela and Mark Bickhard for their comments on earlier drafts of the paper. This research has been partially funded by the Italian Miur Cofin Project "Fiducia e Diritto nella Societá dell'Informazione" and by the European Project MindRACES, Contract No. IST-511931.

References

Alchourron, C., & Bulygin, E. (1981). The expressive conception of norms. In R. Hilpinen (Ed.), *New studies in deontic logic* (pp. 125–148). Dordrecht: D. Reidel.

Barsalou, L., Sloman, S., & Chaigneau, S. (2003). The hipe theory of function. In L. Carlson & E. van der Zee (Eds.), *Representing functional features for language and space: Insights from perception, categorization and development*. Oxford: Oxford University Press.

Basalla, G. (1988). The evolution of technology. Cambridge: Cambridge University Press.

Boella, G., Sauro, L., & van der Torre, L. (2004). Power and dependence relations in groups of agents. In *Proceedings of the conference on intelligent agent technology (IAT 2004)*. Beijing.

Borghi, A. M. (2004). Object concepts and action. In D. Pecher & R. Zwaan (Eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.

Bowles, S. (2004). Microeconomics: Behavior, institutions and evolution. Princeton: Princeton University Press.

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review, 101*(2), 327–341.

Bruner, J., & Goodman, C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*, 33–44.

Bulygin, E. (1986). Permissive norms and normative systems. In A. Martino & F. Natali (Eds.), *Automated analysis of legal texts* (pp. 211–218). Amsterdam: Amsterdam Publishing Company.

Castelfranchi, C. (1990). Social power: a point missed in multi-agent, dai, and hci. In Y. Demazeau & J. P. Muller (Eds.), *Decentralized A.I.*. Amsterdam: North-Holland.

Castelfranchi, C. (1998). Modelling social action for AI agents. *Artificial Intelligence, 103*, 157–182.

Castelfranchi, C. (2001). The theory of social functions. Challenges for multiagent-based social simulation and multi-agent learning. *Cognitive Systems Research, 2*(1), 5–38.

Castelfranchi, C. (2003a). The micro-macro constitution of power. *Protosociology*, 18–19.

Castelfranchi, C. (2005). For a semiotic design of objects, environments, and behaviors. In M. Mattioda & S. Bagnara (Eds.), *From conversation to interaction via behavioral communication*. London: LEA.

Castelfranchi, C., Giardini, F., Lorini, E., & Tummolini, L. (2003). The prescriptive destiny of predictive attitudes: from expectations to norms via conventions. In Alterman, R., & Kirsh, D. (Eds.), *Proceedings of the 25th annual meeting of the cognitive science society*. Boston.

Chao, L., & Martin, A. (2000). Representation of manipulable manmade objects in the dorsal stream. *Neuroimage, 12*, 478–484.

Conte, R., & Castelfranchi, C. (1995). Cognitive and social action. London: UCL Press.

Conte, R., Miceli, M., & Castelfranchi, C. (1991). Limits and levels of cooperation. Disentangling various types of prosocial interaction. In Y. Demazeau & J. P. Mueller (Eds.), *Decentralized A.I-2* (pp. 147–157). Amsterdam: Elsevier.

Eco, U. (1997). Kant e l'Ornitorinco. Milano: Bompiani.

Gibson, J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin.

Grosz, B., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence, 86*(2), 269–357.

Hodgson, G. (2002). The evolution of institutions: an agenda for future theoretical research. *Constitutional Political Economy, 13*(2), 111–127.

Johnson-Laird, P. (1983). Mental models. Cambridge: Cambridge University Press.

Jones, A., & Sergot, M. J. (1996). A formal characterization institutionalized power. *Journal of the IGPL, 4*, 429–445.

Lewis, D. K. (1969). Convention: A philosophical study. Cambridge: Harvard University Press.

Makinson, D. (1986). On the formal representation of rights relations: remarks on the work of stig kanger and lars lindahl. *The Journal of Philosophical Logic, 15*, 403–425.

Makinson, D., & van der Torre, L. (2003). Permissions from an input/output perspective. *Journal of Philosophical Logic, 32*(4), 391–416.

McGovern, A., & Sutton, R. (1998). Macro-actions in reinforcement learning: An empirical analysis. Tech. Rep. 98-70, University of Massachussets, Department of Computer Science.

Mirolli, M., & Parisi, D. (2005). Language as an aid to categorization: a neural network model of early language acquisition. In A. Cangelosi, G. Bugmann, & R. Borisyuk (Eds.), *Proceedings of the ninth neural computation and psychology workshop*. Singapore: World Scientific.

North, D. (1990). Institutions, institutional change and economic performance. Cambridge: Cambridge University Press.

Omicini, A., Ricci, A., Viroli, M., Castelfranchi, C., & Tummolini, L. (2004). Coordination artifacts: environment based coordination for intelligent agents. In *Proceedings of the international joint conference on autonomous agents and multi-agent systems (AAMAS 2004)*. New York, USA.

Pörn, I. (1977). *Action theory and social science: some formal models*. *Synthese library* (Vol. 120). Dordrecht: D. Reidel.

Rawls, J. (1955). Two concepts of rules. *The Philosophical Review, 64*, 3–32.

Schmidt, K., & Simone, C. (1996). Coordination mechanisms: towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work, 5*(2–3), 155–200.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, J. (1995). *The construction of social reality*. New York: The Free Press.

Sellars, W. (1967). *Science and metaphysics: Variations on Kantian themes*. London: Routledge and Kegan Paul.

Smith, B., & Searle, J. (2003). The construction of social reality: an exchange. *American Journal of Economics and Sociology, 62*(1), 285–309.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science, 7*(7), 308–312.

Sugden, R. (1998). The motivating power of expectations. Tech. rep., School of Economic and Social Studies, University of East Anglia, UK.

Tuomela, R. (1995). *The importance of US: A philosophical study of basic social notions*. Stanford: Stanford University Press.

Tuomela, R. (2002). *The philosophy of social practices: A collective acceptance view*. Cambridge: Cambridge University Press.

Vermaas, P., & Houkes, W. (2003). Ascribing functions to technical artefacts: a challenge to etiological accounts of functions. *British Journal for the Philosophy of Science, 54*, 261–289.

Vermazen, B. (1993). Objects of intention. *Philosophical Studies, 71*, 223–265.

Ziman, J. (2000). *Technological innovation as an evolutionary process*. Cambridge: Cambridge University Press.