From Bio-Inspired to Institutional-Inspired Collective Robotics

PTDC/EEA-CRO/104658/2008

Task 2: Institutional approaches to the micro-macro link within human societies.

(Institutional Economics for Institutional Robotics)

Progress Report

Porfírio Silva

7 September 2010

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

1 Preface

> "The time has surely gone in which economists could analyze in great detail two individuals exchanging nuts for berries on the edge of the forest and then feel that their analysis of the process of exchange was complete, illuminating though this analysis may be in certain respects."
> Ronald H. Coase, 1991 Nobel Prize in Economic Sciences, Prize Lecture

Collective Robotics, as a field of research, can also opt for the once-dominant strategy of the sciences of the artificial: dealing only with toy problems can give us the pleasant illusion of progress. Researchers can choose to deal primarily with systems with a very small number of robots and simple tasks - and keep talking of such systems as Artificial Societies. To do so, may be Collective Robotics could limit its inspiration to soccer or to colonies of social insects.

If, however, Collective Robotics wants to get closer to the real diversity and complexity of human societies, it really has to expand its inspiration. Despite the failures of economic science dealing with human societies, it still offers an impressive array of methods to study social phenomena. That is our general motivation to study some Economics' contributions to the study of society. And complexity - the need not to avoid trying to understand complexity - motivates us to choose, among all, institutional approaches to the economic world. Because the social world of human beings can take so many diverse shapes as different institutional environments are able to produce.

Contents

2 Introduction

Within the project 'From Bio-Inspired to Institutional-Inspired Collective Robotics', a team of researchers from Instituto de Sistemas e Robótica (Instituto Superior Técnico) and Instituto Gulbenkian de Ciência (Fundação Calouste Gulbenkian) seek to study and formalise laws that govern collective systems. One important methodological aspect of the project is to bring together theories, ideas and inspiration from institutional economics and cell biology under a common formal framework for large robot populations modelling and analysis. Mathematical modelling will frame the interaction between so disparate approaches. The first three tasks of the project are preparatory and integrative in nature. They will combine contributions from Biology (epigenetic models of the micro-macro link within multicellular organisms) and Economics (Institutional Economics' view of the micro-macro link within human societies) into a unified Bio-Institutional-inspired framework to analyse and synthesise collective systems, focusing fundamental properties of collectives.
This report is the initial contribution from Task 2 - 'Institutional approaches to the micro-macro link within human societies' - to the preparatory phase (tasks 1 to 3) of the project.
Task 2 expected direct contribution to the project is to make available an understanding of how Institutional Economics (and other institutional approaches within social sciences) explains the dynamics of the micro-macro link within human societies. [More on the micro-macro link problem: Section 3.1.] The micro-macro link is to be approached from the two sides of the link. (1) From the macro side, the task is to understand institutional environments. Three different, while related, problems, must be considered: the ontology of the institutional realm; a typology of institutional devices; a typology of institutional failures. (2) From the micro side, the task is to understand means to understand institutionalized individuals as individual agents able both to conform and to modify the institutional setup.
Main specific contributions to answering these questions are summarised in the following, with links to specific sections where more detailed analysis is given.

1. INSTITUTIONAL ENVIRONMENTS

Ontology of the institutional realm

(1) A basic distinction between "brute facts" and "institutional facts" must be recognized. A brute fact is a fact whose existence owes nothing to the observers (Mount Everest has snow and ice near the summit; hydrogen atoms have one electron). A social fact is a fact involving collective intentionality (a pack of hyenas hunting a lion is a social fact, all the hyenas behaving in a responsive way both

to each other and to the lion). Institutional facts are a subset of social facts. The creation of institutional facts involves a mechanism by which a group decides to assign some function to some type of objects, where the function is not explainable by the material (physical, or chemical or biological) features of the object, and must be activated by the ongoing cooperation (agreement, acceptance) of individuals within that group. The assignment of status functions needs a sophisticated representational apparatus (symbolic language). Status functions are vehicles of power in human society. We accept status functions and in so accepting, we accept a series of obligations, rights, responsibilities, duties, permissions, and so on. All these are deontic powers. Deontic relationships provide reasons for action that are independent of desires. Institutional facts are objective facts in the sense that they are not a matter of the preferences of any particular individual. The fact that the piece of paper in my pocket is a ten euro bill does not depend in any way on my subjective preferences, even if my agreement or acceptance is part of the collective agreement and acceptance that is essential to the existence of that institutional fact. [More on this: 3.2.2.]

(2) To understand how institutions work, we need to consider the multilevel character of the institutional realm. Rules are a basic element of any institutional environment. Distinguishing the different levels of rules in any institutional setting helps to recognize the wide variety of strategies that are open to individuals within an action situation. It is useful to distinguish at least three institutional levels (rules of three levels). Operational rules directly affect day-to-day decisions made by participants in a specific setting. These can change relatively rapidly. Collective-choice rules affect operational activities and results through their effects in making policies (procedures to be used to change operational rules) and in determining who is eligible to be a participant. These change at a much slower pace. Constitutional-choice rules determine how, and who, and within which limits, can change collective-choice rules. These change at the slowest pace. If needed, for analytical purposes, the existence of more basic rules (metaconstitutional rules) can be assumed; we can add more basic levels until one gets directly to constraints from the biophysical world (natural constraints are not institutional constraints). The participants in action situations at different levels can be the same individuals or they may differ. [More on this: 4.1.2.]

(3) To be able to exploit all action possibilities they enjoy in an institutional setting, agents need to recognize the multilevel character of their institutional environment and know its specificity. Otherwise, they will be stuck in a single tier world (the operational level). Self-organizing and self-governing capabilities of the agents depend on being able to act in multilevel environments, so being able to

change rules that impact (in an indirect way) the operational level of immediate action. [More on this: 4.1.3.]

(4) Another dimension of the multilevel character of the world is the fact that no action arena can be understood without taking into consideration the exogenous variables that affect the mechanics of that action arena. Three clusters of variables represent this exogenous influence: the biophysical world; the more general attributes of the community (culture); rules (institutional rules broader in scope than the specific action situation). Institutions can only serve efficiently its objectives if they fit its (natural and social) environment. [More on this: 4.2.1.]

Typology of institutional devices

(1) Several writers in Economics gave some definitions of institutions. Especially interesting aspects of these definitions are: institutions shape social actions over long periods of time; institutions connect the past with the present and the future; institutions organize repetitive and structured interactions among members of social groups; there are both formal and informal institutional forms. Institutions are systems of embedded social rules, rather than rules as such. Institutions cannot be taken as equivalent to prescribed patterns of correlated behaviour: behaviour can cease and rules continue in force. [More on this: 3.2.1.]

(2) Institutions can be seen as coordination artefacts. Different coordination artefacts depend on different combinations of physical and cognitive opportunities and constraints they offer to coordinated agents. Sometimes, the effectiveness of coordination depends on the agents' capability to recognize those opportunities; but this is not always so. [More on this: 6.1.]

(3) Institutional environments are about mediated interaction. One of the most powerful features of institutions is that they have the means for mediated interaction. Money and property are classical examples of large scale mediated interaction. [More on this: 6.2.1. and 6.2.5.] Severe limitations of purely direct interaction are shown: by a critic of the spontaneous order hypothesis, the aggregation problem, and the principal/agent problem. [More on this: 6.2.2, 6.2.3. and 6.2.4.]

Typology of institutional failures

(1) We talk of "institutional failures" where existing institutions cannot pro-

vide enough coordination to agents be able to get satisfactory outcomes from their aggregated actions. We have an interdependent situation where a common or collective interest cannot be adequately advanced by individual unorganized action alone. We have independent action where agents act without taking into account neither the effects of their actions on the choices and actions of other agents, nor the aggregate effect of all agents' actions on outcomes. Now, institutional failures can be produced when agents are obliged to independent action in interdependent situations. Given the omnipresence of dependence (and mutual dependence) relationships among agents in social settings, interdependent situations are a basic fact of human societies. Bounded autonomy of agents is related to dependence and interest relationships. [More on this: 4.1.1.]

(2) Economic agents facing an undesirable state of the world, and struggling to change it, can try to design ex ante mechanisms to govern all ex post eventualities in a desired future state of the world. The problem is that such a move must prove ineffective in most complex multi-agent situations, because designing a solution without reliable information would be hard in every case. And several fundamental problems with information must be taken into the picture. First, it is always costly to get an accurate benefits-cost analysis (we cannot simply assume that agents know action-outcome linkages). Second, agents have limited capabilities to avail and weight all available information in an objective manner. Third, since agents can behave in a strategic manner, other agents cannot assume to be able to predict their behaviour just as function of the objective opportunities offered by the situation. Fourth, agents sometimes hold information to prevent other agents from having a complete picture of the situation. Because of the fundamental problems related to incompleteness and asymmetry of information, most complex social processes must be taken as sequential, incremental, and self-transforming processes. Institutional devices ignoring the fundamental challenge of incomplete information must, soon or later, reveal themselves as examples of failed institutions. [More on this: 5.1.1. and 5.1.2.]

2. INSTITUTIONALIZED INDIVIDUALS

A set of guiding principles to the modelling of institutionalized agents

(1) The internal world of individual choice is much more complex than the "rational egoist". Four internal variables affect individual choices: expected benefits; expected costs; internal norms; individual discount rates. Individuals jointly produce outcomes in the external world. The external world not always confirms expectations. Found outcomes impinge on future expectations concerning benefits

and costs of actions. Internal norms affect choices - and are affected by norms held by others. Discount rates are affected by the range of opportunities that an individual has outside a particular situation. [More on this: 4.3.1.] Note that agents cannot be assumed to know action-outcome linkages (how specific sets of actions lead to specific outcomes.)

(2) To cope with complex and uncertain (natural and institutional) environments, where mediated interaction (based on the assignment of status functions) impinges heavily on direct interaction, agents need sophisticated representational capabilities. Internal models of the external world (and of the internal world of other agents) are part of the representational apparatus of individual agents. Shared mental models (ideologies) influence and are influenced by institutions. A revised model of the internal world of individual choice is given to include mental models. [More on this: 6.2.5.]

(3) Populations of institutionalized individuals are heterogeneous. Discount rates, and how norms are perceived by different individuals, are sources of heterogeneity within a population. Individual discount rates depend on the acknowledge range of opportunities that may or may not be available to an individual outside a particular situation. [More on this: 4.3.2.] Internal norms are influenced by norms shared by other relevant individuals. The individual sensitivity to shared norms (internal benefits or costs of obeying or breaking a prescription) varies from one individual to another within a population. [More on this: 4.3.3.]

(4) Habits and routines are behaviour generating mechanisms of individuals with bounded rationality and bounded autonomy in uncertain and complex natural and social environments. To understand the establishment and workings of habits and routines we need to understand how they relate to deliberative capabilities of individuals. [More on this: 4.3.4.]

(5) Self-organizing and self-governing capabilities of agents play a role in solving coordination problems within collectives. Self-organizing capabilities of agents are not "spontaneous" or "emergent" anonymous properties of a system. Talking of agents able to self-organizing is about agents recognizing the need of coordinated action, analyzing the situation and mobilizing themselves to act in a coordinated way, and crafting better rules related to local setting. [More on this: 4.1.3.]

(6) Self-organizing capabilities of agents can be improved with better knowledge on how to change the current situation into a desired direction. A recipe for creating (modifying) situations will be of help. [More on this: 4.1.1. and 4.1.2.]

Task 2 more generic contribution to the project relates to the need of a better understanding of how can we model, in a principled way, fundamental differences between a range of scenarios of collective interaction that we intuitively recognize as diverse, but without a clear definition of the difference at stake. Institutional approaches fit very well this need of the project. An important aspect of the institutional approach, as opposed to the neoclassical approach, is that it allows taking into account a more diverse range of situations in the real economic world. Thinking only in terms of perfect competition context leaves out most of the real situations. Thinking about how different institutions create different situations is much more realistic. For example, some economists think in terms of stock and product exchanges as examples of perfect or near-perfect competition. But these exchanges regulate in great detail the activities of agents (what can be traded, when it can be traded, the terms of settlement, and so on), while in most economic situations agents are not so specifically constrained by rules. [For generic examples in Economics, see 4.3., in fine; 5.2.1.]

The basic and more general approach we can take to differentiate collective systems is to focus on action situations and use the seven clusters of variables (proposed by Ostrom) to describe and analyze them: (1) the set of participants (who may be either single individuals or collective actors), (2) the positions to be filled by participants,(3) potential outcomes,(4) the set of allowable actions (including the choice not to act) and the function that maps actions into realized outcome, (5) the control that an individual participant has in regard to this function, (6) the information available to participants about actions and outcomes and their linkages, and (7) the costs and benefits - serving as incentives and deterrents - assigned to actions and outcomes. [More on this: 4.2.1.]

Within populations of heterogeneous agents, different distributions of different types of agents can impact the effectiveness of some control mechanisms at collective level. Heterogeneity factors affecting the individuals' conformity to norms, or their eventual adoption of opportunistic behaviours, can have huge consequences at the system's level. We can experiment with this kind of dynamics taking individual discount rates or individual delta parameters as indexes of fundamental economic heterogeneity among individuals. [More on this: 4.3.2. and 4.3.3.]

It can prove useful to take a more specific route to an understanding of fundamental differences between scenarios of collective action, the Transaction Costs approach. A taxonomy of transaction costs can be of help to compare scenarios. [More on a basic definition of transaction costs and taxonomy: 5.2.2.]

Three dimensions are important to analyse transactions: asset specificity; the disturbances to which transactions are subject; the frequency with which transactions occur. The three attributes of principal importance for describing governance structures are incentive intensity; administrative controls; contract law

regime. [More on these attributes of governance structure: 5.2.4.] [More on Asset Specificity: 5.2.6.]
Sometimes, more than the individual behaviour, the structure of the situation itself is the main factor causing the observed results. Some environments can be favourable to optimizers, but this is not always the case. At least three features of the environment need to be set at favourable values in order to make the situation optimizers-friendly: complexity, motivation, and information. Most economic situations cannot be characterized by low complexity, strong motivation, and cheap information. The same kind of analysis can be of use for other kinds of collective systems. [More on this: 6.2.5.]
Beyond the above mentioned specific contributions to the Project, this report must be read as a whole: as a contribution to a vision on collective systems from the vantage point of research on human societies. More specifically, this report intends to contribute with an institutional vision of sophisticated collective systems, where massive numbers of agents (with bounded rationality and bounded autonomy) are both constructive within and constructed through institutional environments - networks of (unavoidably) incomplete institutions (coordination artefacts of a specific kind) combined with mental models of the (natural and social) world, shared by the agents themselves. The elements of an institutional approach given in this report intend to be a contribution to resist, on solid grounds, any biologically driven reductionist approach to institutional realm.

Porfírio Silva

3 Why Institutions? Why Economics? Why Institutional Economics?

In this chapter we motivate the use of Economics, and specifically the appeal to Institutional Economics, to inspire new approaches to Robotics Collective. The theoretical reasons given for this trip will, of course, have implications for the proposed itinerary.
In Section 3.1. we introduce the fundamental problem of social order, or the micro-macro link problem. In Section 3.2, some definitions of institutions, from some writers in Economics, are given; we endorse a fundamental approach to institutions based on the ontological status of the institutional realm; and present a case study that illustrates the meaning and potential of this ontological approach. In Section 3.3., the appeal to Institutional Economics, instead of other approaches to the economic world, is motivated.

3.1 The micro-macro link and the problem of social order

Epstein and Axtell start his **Growing Artificial Societies** with the question: "How does the heterogeneous micro-world of individual behaviours generate the global macroscopic regularities of the society?" [Epstein and Axtell, 1996, p.1]. This is a classical problem of social sciences, and of social philosophy: the micro-macro link problem, or the problem of social order. To understand the micro-macro link within human societies, it is important to recognize it is a three faceted issue, involving agents' actions, agents' cognition, and external forces and structures [Conte and Castelfranchi, 1995, pp.9,47,142]. Within this scenario, cognition and emergence combine to make the complexity of the social world.
On the one hand, macro-social phenomena may emerge, unintentionally, from micro-interactions. Notwithstanding, emergence is not the only possible explanation of the origins of macro-phenomena. Often, macro-phenomena are deliberately set up (institutional building). It is worthy to note that the direction of emergence is not necessarily from micro to macro. Macro-phenomena may unintentionally feed back into micro phenomena. Bounded rationality combines with bounded autonomy to give place to emergent phenomena: there are deliberately planned actions but they may produce unintended effects beyond reach of the agents' understanding or awareness.
On the other hand, macro-social phenomena not only directly emerge from behaviours; they also derive from the agent's cognitive representations and states. Unlike mere reactive systems, socially responsive systems react not only to (physical) actions of other systems, but also to what they believe their intentions are.

In cognitive systems, behavioural responses are mediated by modifications of the system's beliefs, capable of inducing modifications of the system's goals. If we can talk of "behavioural shaping" in reference to processes by means of which agents acquire behavioural dispositions (habits and routines), we can also talk of "cognitive shaping" in reference to processes by means of which agents acquire beliefs and goals from external social sources.

The micro-macro link problem, being at the crossroads of a true constellation of problems related to the functioning of collectives, deserves to be put at the heart of a research effort looking for an improvement of the current situation, where most multi-robot systems model extremely poor social environments.

In recent years, the predominant inspiration for collective robotics modelling analysis and design has been originated from biology. In the case of swarm robotics (SR) [Bayindir and Sahin, 2007], inspiration is taken from studies of the self-organizing capabilities displayed by social insects such as ants. Drawing inspiration from these types of natural systems, SR concepts have been applied to swarms of robots with limited sensing and actuation capabilities, performing relatively simple cooperative tasks such as foraging, coverage or odour tracking. Despite the reasonable success of SR in relatively simple applications, there is no systematic method to design individual behaviours at the micro level, including their interaction actions for the desired collective behaviour at the macro level to emerge; in fact, the emergent nature of the collective behaviour is a principle that precludes goal- or performance-oriented design. Economics-inspired experiences within multiagent systems suggest that merely emergent processes and simple local interactions between individuals sometimes lead to inefficient solutions to collective problems (Caldas, 2001). Furthermore, some interesting problems of social order concern, not only reactive agents, but also agents endowed with some deliberative capabilities and some autonomy.

Looking at another field of scientific research, Economics can provide some insights on how to deal with large collective systems. Institutional Economics (IE) [e.g., Hodgson, 2000] takes institutions - coordination devices deliberately set up by agents or evolved out of interaction - as key elements of any sophisticated society. Building on IE main direction, we have suggested Institutional Robotics (IR) as a new strategy to conceptualize multi-robot systems, taking institutions as the main tool of social life of robots with bounded rationality and bounded autonomy [Silva and Lima, 2007].

The central aim of this paper is to explore, in more detail, fundamental concepts from Institutional Economics, with a view to make it easier for robotics' practitioners using them to implement and control systems of multiple robots. Providing an understanding of how Institutional Economics (IE) explains the dynamics of the micro-macro link within human societies, this working paper aims at contribut-

ing to a more sophisticated implementation of the micro-macro link in Collective Robotics.

3.2 What are Institutions?

3.2.1 Some definitions

It may seem simple to say what institutions are. However, dealing with such a complex reality, this apparent simplicity has to be misleading. We will present some definitions proposed by economists. In some of them we will emphasize particular aspects particularly relevant.
One "old" definition, from G. von Schmoller:

> An institution is "a partial order for community life which serves specific purposes and which has the capacity to undergo further evolution independently. It offers a firm basis for shaping social actions **over long periods of time**; as for example property, slavery, serfhood, marriage, guardianship, market system, coinage system, freedom of trade."(Schmoller [1900, p.61], quoted by Furubotn and Richter [1997, p.6])

From Douglass North, 1993 Nobel Prize in Economic Sciences, one of the leading withers in New Institutional Economics:

> "Institutions are the humanly devised constraints that structure political, economic and social interaction. **They consist of both informal constraints** (sanctions, taboos, customs, traditions, and codes of conduct), **and formal rules** (constitutions, laws, property rights). Throughout history, institutions have been devised by human beings to create order and reduce uncertainty in exchange. (...) They evolve incrementally, **connecting the past with the present and the future**; history in consequence is largely a story of institutional evolution in which the historical performance of economies can only be understood as a part of a sequential story." [North, 1991, p.97]

From Elinor Ostrom, 2009 Nobel Prize in Economic Sciences, who has deeply researched the institutional environments of natural resources with communal management:

> "**Institutions can be defined as the sets of working rules** that are used to determine who is eligible to make decisions in some arena, what actions are allowed or constrained, what aggregation rules will be used, what procedures must be followed, what information must or must not be provided, and what payoffs will be assigned to individuals dependent on their actions (...). **All rules contain prescriptions that forbid, permit, or require some action or some outcome. Working rules are those actually used, monitored, and enforced when individuals make choices about the actions they will take** (...)." [Ostrom, 1990, p.51]

> "Broadly defined, institutions are the prescriptions that humans use to organize all forms of **repetitive** and structured interactions including those within families, neighborhoods, markets, firms, sports leagues, churches, private associations, and governments at all scales." [Ostrom, 2005, p.3]

From the introductory essay in a recent anthology of New Institutional Economics contributions:

> "**Institutions are the written and unwritten rules, norms and constraints that humans devise to reduce uncertainty and control their environment.** These include (i) written rules and agreements that govern contractual relations and corporate governance, (ii) constitutions, laws and rules that govern politics, government, finance, and society more broadly, and (iii) unwritten codes of conduct, norms of behavior, and beliefs. Organizational arrangements are the different modes of governance that agents implement to support production and exchange. These include (i) markets, firms, and the various combinations of forms that economic actors develop to facilitate transactions and (ii) contractual agreements that provide a framework for organizing activities, as well as (iii) the behavioral traits that underlie the arrangements chosen." [Ménard and Shirley, 2005, p.1]

Geoffrey Hodgson, one of the leading representatives of heirs of "Old Institutionalism" ...

> ... defines institutions as "**durable systems of established and embedded social rules that structure social interactions, rather than rules as such**. (...) institutions are social rule-systems, not simply rules." This definition is intended to exclude misleading definitions, those taking institutions as "prescribed patterns of correlated behavior". Defining institutions as behavior would mislead us into presuming that institutions no longer existed if their associated behaviors were interrupted. The British monarchy does not cease to exist when the members of the royal family are all asleep and no royal ceremony is taking place: royal prerogatives and powers remain, even when they are not enacted. It is these powers, not the behaviors themselves, which mean that the institution exists [Hodgson, 2006, pp.2-3] .

> On the other hand, **this definition is intended to cover "systems of established and prevalent social rules that structure social interactions**", like "language, money, law, systems of weights and measures, table manners, and firms (and other organizations)" [Hodgson, 2006, p.2], **as well as "the informal basis of all structured and durable behaviour", informal basis that requires the presence of non-deliberative mechanisms like habits and routines** [Hodgson, 2006, p.13]. Habits that are persistent, shared, and prevalent within a group are the basis for costumes; organizational meta-habits, existing on a substrate of habituated individuals in a social structure, are routines; habits and routines are essential parts of institutional dynamics [Hodgson, 2007, p.111] .

From this diversity of definitions, two lessons can be learned. First, each definition of institutions emphasizes certain aspects of reality, but at least in some cases, it seems doubtful whether the definition is logically prior to the examples or vice versa. And second, they all lack a clear indication of the fundamental mechanism underlying institutions in human societies. To overcome this lack, we need to understand the fundamental ontology of institutional reality.

3.2.2 A fundamental ontology of institutions

We need an understanding of the basic structure of institutional reality in order to capture the essential aspects underlying the workings of social and economic institutions. We can try to restrict ourselves to the most complex aspects of institutions (the intricate features of courts or governments), just to avoid recognizing

that those high level realities rest on much more basic mechanisms and processes. Our proposal at this point is to address a basic point of ontology: what are the most fundamental foundations of institutions? What are institutions from an ontological point of view?

John Searle's research on the construction of social reality [starting with Searle, 1995] is useful to answer that question and to enlightening the most fundamental mechanisms of institutions. The cornerstone of his exercise is the distinction between brute facts and institutional facts. A brute fact is a fact whose existence owes nothing to the observers (Mount Everest has snow and ice near the summit; hydrogen atoms have one electron). A social fact is a fact involving collective intentionality: if I am a violinist in an orchestra I play my part in our performance of the symphony; the orchestra playing the symphony is not the same thing as a certain number of performers each playing his own part. A pack of hyenas hunting a lion is a social fact, all the hyenas behaving in a responsive way both to each other and to the lion. Institutional facts are a subset of social facts. The creation of institutional facts involves a process by which a group decides to assign some function to some type of objects, where the function is not explainable solely by the material (physical, chemical, or biological) features of the object, and must be activated by the ongoing cooperation (agreement, acceptance) of individuals within that group. Institutional facts are objective facts in the sense that they are not a matter of the preferences of any particular individual. The fact that the piece of paper in my pocket is a ten euro bill does not depend in any way on my subjective preferences, even if my agreement or acceptance is part of the collective agreement and acceptance that is essential to the existence of that institutional fact.

Searle [2006] presents a clear systematization of this approach to institutional reality, as a specifically human reality, based on three elements.

First, collective intentionality. Collective intentionality is a capacity of human beings (and of many other species) to engage in cooperative behaviour and sharing of attitudes with con-specifics. "Intentionality", in the sense philosophers use the word, describes the feature of mind by which mental states are directed at or about objects and states of affairs in the world. Thus, for example, if I have a belief it must be a belief that such and such is the case. Besides individual intentionality (which we can describe by forms like "I desire", "I believe", "I intend"), we are also capable of collective intentionality (which we can describe by forms such as "we desire", "we believe", "we intend"). Collective intentionality can take the form of intentional collective action (I am playing the violin part as part of the orchestra playing the symphony) or other forms, like a collective belief (the church congregation reciting the Nicene Creed is expressing a common belief that is an identity mark of the community). In Searle's terms, a social fact is any fact

involving collective intentionality of two or more human or animal agents.
Second, status functions. Humans, and some animals, have the capacity to assign functions to objects. If an individual can use a stump as a chair, a group can use a log as a bench. Here, the assignment of function is supported on physical features of objects. Humans have the capacity to assign functions to objects where the physical features of the objects are largely irrelevant to the assigned function. It seems that this capacity separates humans from all other species. In this case we speak of status functions. Money, as a function, does not depend on the material chosen for banknotes or coins (although material has some practical relevance, related, for example, to be easy to transportation and hard to counterfeit). In fact, "electronic money" is close to complete dematerialization: money without currency, just magnetic traces on computer disks organized in some specific ways - because of the status function they serve. A border may have once been indicated by a wall with gates, but the border can persist after the removal of the wall as a physical obstacle and its replacement by some signs. Money and borders, as well as many other institutions and institutional facts, are created and exist thanks to acts of collective intentionality: collective assignment and recognition of status functions.
The general form of the assignment of a status function ("constitutive rule") is "X counts as Y in context C". Money is an institution in which a certain kind of piece of paper, produced under certain circumstances, is taken as currency and performs a function that can be described as "general equivalent of exchange". Marriage is an institution in which certain words, uttered by the right person in the circumstances envisaged, serve as the beginning of a certain kind of relationship between the people involved, implying specific rights and duties. In any case, the Y term must name something more than the sheer physical features of the object named by the X term. In the formula "X counts as Y in context C", the Y can be people (e.g., chairperson, wife, priest), objects (e.g., bills, certificates, licences), and events (e.g., elections, wars, weddings). The nesting of institutions can be represented by successive iteration of the fundamental formula for constitutive rules, where the X term at one level can be the Y term at other level. For example, usually only a citizen of a given country can become president of that country.
Third, deontic powers. Status functions are vehicles of power in human society. We accept status functions and in so accepting, we accept a series of obligations, rights, responsibilities, duties, permissions, and so on. All these are deontic powers. If I have a property, I have a certain authority over it, and I have an obligation to pay some taxes. There is nothing like this in the animal kingdom. In human societies, we have a set of deontic power relations. Obligations and permissions are reasons for action, if we can recognize them. And, importantly, deontic rela-

tionships provide reasons for action that are independent of desires. To recognize that I am the owner of this site gives people some reason to act a certain way, those reasons not being based on any of their desires.
The general form of the assignment of a status function can be different if we rather want it to express deontic powers assigned by collective intentionality. So, we can have: "We accept (S has power (S does A))", where S is an individual or a group and A is an action. More specific forms can be "We accept (S is enabled (S does A))" or "We accept (S is required (S does A))". The two forms for the assignment of status functions can be linked. A particular example is: "X, this piece of paper, counts as a five euro note" would be in part replaceable by "We accept (S, the bearer of X, is enabled (S buys with X up to the value of five euro))".
So, on this account, institutions are all a matter of the assignment of status functions by collective intentional acts, so creating deontic powers representing reasons for action that are independent of desires. Searle insists on this being a human specific phenomenon: "Suppose I train my dog to chase dollar bills and bring them back to me in return for food. He still is not buying the food and the bills are not money to him. Why not? Because he cannot represent to himself the relevant deontic phenomena. He might be able to think 'If I give him this he will give me that food'. But he cannot think, for example, now I have the right to buy things and when someone else has this, he will also have the right to buy things." [Searle, 1995, p.70] The essential thing that discriminates the dog from the human is that only the human has the powerful representational capabilities that language allows.
It is worth to note that agents don't need to be aware of the details of the workings of institutional ontology to behave adequately within an institutional environment. Most people never reflect on the underlying mechanisms of money, just recognizing banknotes and coins and how to use them in a day to day basis. In some special circumstances, experts can be called into the scene to clarify some issues (for example, when someone is caught using counterfeit notes and insists that he was just using notes withdrawn from an automated teller machine).
From Searle's point of view, it is useful to consider institutions as processes: even objects (like currency) are just the continuous possibility of an activity (a standing possibility of paying for something). This priority of process over products explain why institutions are not worn out by continued use, but each use of the institution is in a sense a renewal of that institution, because of being a renewed expression of agreement and acceptance.
One crucial point is that institutions allow direct and immediate interaction being replaced by indirect and mediated interaction of a much more sophisticated kind. With the deontic apparatus associated, for example, to property or marriage, we no more have to rely on direct interaction with things and other people in order

to sustain the arrangements and we can maintain them in the absence of the original physical setup. People can remain married even though marriage is originally about cohabitation and they now have not lived with each other for years. People can own property even though property is originally about physical possession and now the property is a long way away from them. And one essential point here is that indirect mediated interaction needs representations (we will deal with this problem within the last chapter of this report).
It may seem difficult to accept Searle's view on the ontology of institutional reality, at the same time resting in the material features of the world and independent of them, relying on human mental acts. However, a well known episode of the Portugal's twentieth century history illustrates wonderfully some aspects of the Searle's proposal.

3.2.3 Case Study: The Alves Reis affair

Between February and March 1925, more than one hundred and fifty thousand counterfeit banknotes of escudo 500 (escudo was the Portuguese currency before the country adhered to the euro zone) have been put into circulation in the European continental part of Portugal [1]. These counterfeit banknotes had a remarkable peculiarity: they were strictly identical to the corresponding genuine banknotes. They were made of the same type of a special paper, had been printed with the same type of ink, and by the same processes used with the government authorised banknotes. In fact, they had been printed on the same security printing firm in London, by the same rotary press, and with the same plates of their legal twins.
How had this happened? A group of swindlers had persuaded a security printing firm of London, with large past experience of printing Portuguese banknotes, that they had got a mandate from the Bank of Portugal, and from the country's government, to print a second lot of notes of escudo 500 with the effigy of Vasco da Gama, of the variant called "plate 2", using precisely the same plates, the same numeration of the notes, and the same seals from the governor and directors of the Bank of Portugal the firm had previously used to a similar emission of banknotes. Supposedly, the notes of this second emission were to circulate in the Portuguese African colony of Angola, with a superimposed impress of the word "Angola", to be added later.
Alves Reis, the leader of the operation, and his blind collaborators, had managed to mislead the English firm with an impressive array of forged contracts and authorisations. Notwithstanding, the counterfeit banknotes were, as physical objects, in all respects perfectly equal to their legal twins. Even managers and experts

[1] For the basic information about the series of events here summarized, see [da Mota, 1996] and [Wigan, 2004].

from the Bank of Portugal had several opportunities to testify, answering checking requests from suspicious people, that they were perfectly good notes.
The hasty conversion of the counterfeit money into genuine notes, with recourse to various types of massive transactions, triggered a deluge of brand new Vasco da Gama 500 escudo notes (the 200.000 notes Reis got printed were equivalent to 0.88% of Portugal's nominal GDP at the time), spreading a wave of suspicion about its origin. Several local agencies of the Bank of Portugal, after analysis of specimens, quickly denied any reason for concern. Given the increasing refusal of the public to receive such notes, the Bank of Portugal in a statement belies the circulation of counterfeit notes of escudo 500. Only at a later stage, when the monetary authorities conducted a large scale checking, they discovered that, among six thousand notes examined, four of them were duplicates (they had the same matriculation number). They recognized that something had to be wrong - but this recognition did not improve their ability to distinguish the counterfeit from the good notes. On December 1925, the Bank of Portugal Governor called a meeting of the Board, which found that it was impossible to distinguish between good and bad notes. The Board therefore decided to withdraw all notes of escudos 500. Once again attesting the complete material identity between the two sets of notes, the Bank of Portugal exchanged any escudo 500 notes for escudo 1000 notes, admittedly unable to distinguish between the different emissions. Counterfeit and legal banknotes were strictly equal in everything related to its material features.
Reis conceived the swindler after reading a speech in the Parliament by former Prime-Minister Francisco Cunha Leal, revealing an irregular method of issuing national currency notes, whereby the Bank of Portugal secretly had notes printed, and neither recorded such operations in the books, nor did it get the required approval by parliament.
This anomalous situation arose from the dismantling of preceding rules about metal-backing of banknotes. Since 1891, banknotes were not fully convertible to gold or silver. However, by a 1906 legal imposition, the Bank of Portugal was required to have 20% of notes backed by gold and 100% of notes backed by silver. In April 1918, these metal-backing constraints had been relaxed, allowing the Bank of Portugal to oblige the deficit financing needs of the government. Since the banknotes no longer were convertible, the only expenses involved in issuing currency were the cost of printing. Previously, the Bank was authorized to issue notes to the amount of twice its share capital. Up to 1924, the Bank issued notes in excess of 100 times its share capital. Now, Alves Reis asked himself, if the Bank of Portugal itself can follow irregular issuing methods, why would he refrain from doing the same?
At a point, Reis, using the money he earned with the counterfeit notes, established his own bank, both as an instrument by which he could dispose of his illicit

currency, and as a vehicle to take control of the Bank of Portugal. The Bank of Portugal, although bound to specific law provisions of law, was a private firm - and was the sole institution capable of initiating proceedings against counterfeiters of its banknotes. Reis started buying Bank of Portugal shares on the Lisbon stock exchange.
In fact, what made Reis' counterfeit banknotes different from legal banknotes was not in the notes themselves, in any of its material qualities. The unique difference between them was an institutional difference. Reis no longer was about to counterfeit notes; he was about to fake institutions. Something that, if Searle is right, is only possible for human beings empowered with his representational tools provided by symbolic language.
If the above proposed ontological approach to institutions is mainly correct, institutional reality has a dual aspect. It is built on the material world, and, at the same time, it is fundamentally different from its raw material. And there is nothing magic about this: it is just the representational capabilities of human beings producing yet another ontological level, the institutional level of social reality. The Alves Reis affair is an extraordinary illustration of this. And the basic mechanism of collective agreement and acceptance, "constitutive rules" for status functions, and deontic powers, is - we suggest - strong enough to allow an understanding of the fundamentals of institutional realm.

3.3 Why Economics? Why Institutional Economics?

The field of Economics, and social and political philosophy before, have been trying for centuries to understand how large collective systems (human societies) work. Even if successive failures of Economics in predicting or explaining social dynamics made us suspicious about its basic assumptions, it still offers an impressive array of methods to study social phenomena. Robotics, so far mostly restricted to small sized systems of multiple robots, and/or systems targeted to relatively simple applications, can perhaps learn something from a science of societies with massive numbers of relatively complex members (human societies, human beings).
The field of Economics has already provided some insights on how to deal with large collective systems. Market-based multi-robot coordination is an example [Dias et al., 2006]. Inspired by market mechanisms, researchers have proposed systems like MURDOCH [Gerkey and Mataric, 2002] and TraderBots [Dias et al., 2004] to achieve flexible allocation of subtasks using auctions between robots. In these systems, robots act as agents trying to maximize their individual profits. Every time a task is auctioned, robots must pay a price to obtain it. Once the task is completed, a payment is done to the robot who won the auction. Nevertheless, to accomplish the task, that robot has to expend some resources for which

it must also pay a price. In these systems, tasks and resources are considered as commodities than can be compared in value and traded among robots. The underlying assumption is that with every robot trying to maximize its individual profit, team coordination and efficiency will be improved. Meanwhile, basic assumptions of utility-driven approaches are challenged by results of some experiences within multiagent systems (MAS) suggesting that merely emergent processes, and simple local interactions between individuals, sometimes lead to inefficient solutions to collective problems [Caldas, 2001]. Institutional Economics (IE) proposes alternative assumptions [see, e.g., Hodgson, 2000]. We will explore those alternative assumptions within the text you are reading.

A note on terminology. The term "institutional economics" is applied to at least two different economic approaches: "American institutional economics" or "old institutional economics" refers to the heirs of Economics' writers as Thorstein Veblen, John R. Commons, Wesley Mitchell, and Clarence Ayres; "new institutional economics" refers to the tradition of work stemming primarily from the transactions cost approach of Ronald Coase, Oliver Williamson, and Douglass North. While "Old institutional economics" rejects most assumptions of orthodox ways of doing Economics, "new institutional economics" accepts in principle all the tools contributed to the field by neoclassical economics - something that has not preventing practitioners of this approach to strongly depart from the received view of economic world [Rutherford, 2001]. Throughout this paper we combine contributions from the "old institutionalism" and "new institutional economics", and we will not worry too much about that distinction. The point is that, even if we recognize the importance of the distinction, we still preferred some eclecticism, a trend that also exists within the field of economics.

Hodgson [1988] gives one of the best structured explanations of the reasons for preferring the institutionalist approach instead of the neoclassical approach to economics. According to Hodgson the core assumptions of neoclassical economic theory are:
(i) rational maximizing behaviour of economic agents;
(ii) no chronic information problems: no strong uncertainty about the future, no substantial ignorance about the structures and parameters of the world, no divergent cognition of individual phenomena;
(iii) theoretical preference for stable equilibrium states and insufficient attention to historical processes of transformation and social dynamics.
Hodgson draws a critique of this conception in three main points: first, a critique of methodological individualism; second, a critique of the maximization hypothesis; third, a critique to the rationalist conception of action. Let us say a few words about each of these points in turn.

First, we have a critique of methodological individualism [Hodgson, 1988, chapter 3].

Methodological individualism is a methodological position characterized by the assumption that all explanations of social phenomena have to be couched in terms of statements about individuals. The individual is taken as given. Methodological individualism sees the individual facing the external world and reacting to it through the perception of it constraints and opportunities - but does not pay enough attention to external factors influencing individual action, perception and purposes. Links from individual action to collective phenomena are conceived within the limits of the "compositive method": wholes must be explained by the properties of its elements, but parts suffer no relevant influence from the whole. The micro affects the macro, but not vice versa, feedback mechanisms being ruled out. The individual utility function is regarded as both immutable and beyond dispute. The crucial feature of methodological individualism is its statement of the primacy of individual purposefulness, its failure to give an adequate account of the formation of purpose itself, and its refusal to take into due consideration how institutions are involved in the moulding of individual preferences and purposes. Institutional Economics argues that the socio-economic and institutional environment has a significant effect in the kind of information individuals receive, and on individuals' cognition, preferences, and thereby much of their behaviour.

Second, we have a critique of the maximization hypothesis [Hodgson, 1988, chapter 4].

The "rational economic man", a core assumption of neoclassical economic theory, is an agent that, through a rational calculation, takes into account all relevant information, and maximize something (expressed as a single value), usually called "utility". Usually this is linked to another assumption, that the economic agent has a consistent ordering of preferences, this preference ordering being transitive (if A is preferred to B, and B is preferred to C, then A is preferred to C) and irreflexive (for any good A, A is not preferred to A). This assumption about stable and consistent preferences ordering is not supported by any empirical evidence. The core of the maximizing hypothesis seems completely unrealistic, given what is known about how human beings deal with paradigmatic situations. For example, according to some calculations, there at least 10000 different products in an average supermarket; there are over 43 trillion possible initial positions of the Rubic's Cube. In all situations of this kind, we don't ever try to optimize: we use simple procedures to get out of the difficulty with reasonable outcomes - and typically

these procedures are sub-optimal.
The primitive form of the maximization hypothesis takes such behaviour as conscious and deliberate. An evolutionary variant leaves on one side the question of whether or not the individual or the firm is deliberately maximizing, and just assumes that economic agents that survive have been maximizing and that was the reason for their survival. The fundamental problem with this move is the absence of any specified mechanism responsible neither for the sustainability of the optimal behaviour through time nor for spreading it to other agents, thus failing to give a credible evolutionary explanation.
There has been also attempts to replace the strong maximization hypothesis by the bounded rationality hypothesis of Herbert Simon: due to the weight of uncertainty and incompleteness of knowledge that bears upon decision-making, and the limited computational capacity of the human brain, firms and consumers must not be maximizing, but rather "satisficing" , that is, trying to attain acceptable minima. This move could perhaps be sufficient to solve the problem of too little information to reason upon in order to decide how to behave; but it seems insufficient to deal with the problem of too much information to assess.
Another fundamental problem with this hypothesis is the assumption that "the end justifies the means". The determination of ends is taken as exogenous to economy; economics supposedly just deals with the calculation of means. But this ignores all the social and moral reasons that can make the picture more complex within real social settings. Not all ends are acceptable, and it is arguable whether it is the case or not; not all means can be mobilized to pursue legitimate ends; the factors affecting this kind of weighting are not all economic in nature.
Note that in orthodox economics the notions of global rationality and equilibrium are intimately connected: neoclassical maxima, where equilibrium can be reached, are attained through global calculation.

Third, we have a critique of the rationalist conception of action [Hodgson, 1988, chapter 5]).

Not all of action in the economic realm is driven by rational calculation: there is a large class of actions which are relevant to economic and which arise in a different manner, for example being influenced by unconscious and subconscious mental processes. This is not to suggest that no actions are dominated by reason; this is to state the need of a theory of action which does not rely largely or exclusively on rationalist mechanisms. We just need to mention advertising to recall at which point symbolic, not rational, dimensions interfere with economic behaviour. Even perception involves unconscious computational processes. A large share of our day to day behaviours depends on habits and routines, not in rational and

fully conscious deliberation. Beliefs, attitudes and values also have a guiding effect on our reading of the reality and response to it. Social norms constrain what we take as acceptable behaviour, and so channel our action courses. Information does not enter mind as raw data. Many economic theorists write as if the information was a fluid of undifferentiated sensory data entering the head of an individual. Yet, things do not work that way: information is accessed through a cognitive framework, which depends largely on culture and institutions. Our conceptual apparatus, which filters and organize sense data, is a result of interaction with many other individuals within society. Fortunately, we are not exclusively rational: given our scarce computational resources, we are lucky in that not all of our mental processes are at the same level of rationality. Saving mental resources with habits, routines, and norms, gives us the chance of using fully deliberative and rational capabilities within narrow and strategic domains - so excluding of all economic action being fully rational.

All these lines of criticism converge on the need to consider the existence and role of institutions in economic and social life - because there are no individuals like atoms shaped in complete independence from social interaction. Writers in economics in the neoclassical tradition also have recognized that institutions matter in the economic world, but, as economic models have become increasingly abstract, institutions have been left out of the mainstream picture, technical sophistication of the theory not encompassing the institutional complexities of the actual world [Furubotn and Richter, 1997, p.1].

Furthermore, the idea of the neutrality of institutions is strongly entrenched in neoclassical theory. At any time, the property-rights configuration existing in an economy is determined and guaranteed by a governance structure (a system of rules plus the enforcement mechanisms). That's the reason for Institutional Economics studying institutions as integral elements of a general economic model (endogenization of institutions). Of course, neoclassical economists do not ignore the existence of institutions. The existence of political, legal, monetary, and other institutions is recognized, but they are regarded as neutral for economic outcomes. Institutions are taken as neutral in the sense that a specific economic problem is seen as explainable without reference to institutional differences. For example, whether goods and services are exchanged by the use of money or otherwise, is a question we do not need to address to model the situation. Several economic problems show in which sense neoclassical models take institutions as neutral. In that sense, it does not matter, for example [2]:

- how production is organized - by the prices mechanism across markets or within a hierarchically organized firm;
- whether the factors of production are owned or rented by their users;

[2] All examples from [Furubotn and Richter, 1997, pp.9-10].

- who - individuals or society - hold property rights of the productive factors;
- whether or not the ownership and control of a firm are separated;
- whether transactions are undertaken singly as transactions between "faceless strangers" or are repeated frequently between the same parties;
- whether a good is supplied by a monopolist or by a large number of independent firms;
- whether an economy is based on the operation of decentralized individuals or on a command structure acting as a central agent.

Because of the idea of the neutrality of institutions, neoclassical theory will not be able to discriminate between certain economic situations that are, in fact, quite different: a money-using economy cannot be distinguished from a barter economy, a capitalist economy cannot be distinguished from a socialist economy. How incredible it may sound now, the "market socialism" of Oscar Lange was a theoretical construction based on the idea that all the neoclassical apparatus could apply to a socialist society, with no worries about the wholly different institutional environment as contrasted to an idealized capitalism. It is the idea of the neutrality of the institutional framework that permitted such a theoretical endeavor. This is the strange world of costless transactions , "as strange as the physical world would be without friction"[3], as Stigler once said.

Another interesting aspect of the neoclassic approach is its inadequate account of power within society. Neoclassical economics considers a society in which exchanges, property rights and decisions (or the decision makers themselves) are not at all exposed to the use of physical force or other forms of compulsion (except for the force of the state). Notwithstanding this nave view, in the actual world power exists and influences matters: not only state power (recognized at some extent by orthodox economics), but also pressure groups and coalitions with the purpose of improving the welfare of their members at the expenses of others in the system. (In the neoclassical paradigm, coalitions operate just to produce Pareto improvements of their members within an environment of voluntary association and voluntary exchange.) Because, in the neoclassical framework, force is supposed to be perfectly monopolized by the state, there is no room to consider strikes, boycotts, political and social pressure or resistance.

An important aspect of the institutional approach, as opposed to the neoclassical approach, is that it allows taking into account a more diverse range of real situations. Thinking only in terms of perfect competition contexts leaves out most of the real situations. Thinking about how different institutions create different situations is much more realistic - and it is one of our main motivations for studying the contribution of Institutional Economics. Let's see how Coase [2002] sees the point: "Stock and product exchanges are often used by economists as examples of

[3]See the section 5.2., on Transaction Costs Economics.

perfect or near-perfect competition. But these exchanges regulate in great detail the activities of traders (and this quite apart from any public regulation there may be).What can be traded, when it can be traded, the terms of settlement, and so on are all laid down by the authorities of the exchange. There is, in effect, a private law. Without such rules and regulations, the speedy conclusion of trades would not be possible. Of course, when trading takes place outside exchanges (and this is almost all trading) and where the dealers are scattered in space and have very divergent interests, as in retailing and wholesaling, such a private law would be difficult to establish, and their activities will be regulated by the laws of the State. It makes little sense for economists to discuss the process of exchange without specifying the institutional setting within which the trading takes place, since this affects the incentives to produce and the costs of transacting."

Having motivated the recourse to Institutional Economics to inspire new approaches to Collective Robotics, the next chapter will introduce the basic elements of a general framework of institutional environments.

4 From prisoners in a dilemma to institutional agents

This chapter presents the basic elements of a general framework of institutional environments. It is explained why it is important to understand the multilevel nature of institutional reality, and how this relates to the self-organizing capabilities of agents. A general model of action arenas is introduced, including the exogenous variables (biophysical and social world) influencing action situations and participants. A general recipe for creating situations is given: a grammar of regulatory rules that, at some extent, can be manipulated to modify specific aspects of a situation. A model of individual agents, the participants that are able to animate action arenas, is presented. The chapter ends with a reference to habits and routines as important links between individuals and institutions.

4.1 Self-organizing and self-governing individuals within a multilevel institutional realm

The tragedy of the commons [Hardin, 1968], the logic of collective action [Olson, 1965], and the prisoner's dilemma in game theory are related concepts in modelling problems that individuals face when attempting to achieve collective benefits. These models capture a thread of difficulties of collective action well represented by the free-rider problem and opportunism.

> "Whenever one person cannot be excluded from the benefits that others provide, each person is motivated not to contribute to the joint effort, but to free-ride on the efforts of others. If all participants choose to free-ride, the collective benefit will not be produced. The temptation to free-ride, however, may dominate the decision process, and thus all will end up where no one wanted to be." [Ostrom, 1990, p.6]

> "Opportunism - the deceitful behaviour intended to improve one's own welfare at the expenses of others - may take many forms, from inconsequential, perhaps unconscious, shirking to a carefully calculated effort to defraud others with whom one is engaged in ongoing relationships." [Ostrom, 2005, p.51]

The perhaps most important weakness of those models is that they assume relatively fixed constraints for action, beyond the reach of coordinated action of the agents. For example, the prisoners in the famous dilemma cannot change the constraints imposed on them by the district attorney; they are in jail; even if they

are in an **interdependent situation** (a common or collective interest cannot be adequately advanced by individual unorganized action), they are obliged to **independent action** (acting without taking into account neither the effects of their actions on the choices and actions of other agents, nor the aggregate effect of all agents' actions on outcomes). "Acting independently in this situation is the result of coercion, not its absence." [Ostrom, 1990, p.39]

4.1.1 Bounded autonomy of agents creates interdependent situations

It is important to recognize the puzzles independent action raises in interdependent situations, because there are no agents empowered with better than bounded autonomy. "The agents have bounded autonomy". What could this exactly mean? Let us try to contribute to an answer with the help of [Conte and Castelfranchi, 1995].

Autonomous agents must be capable of generating new goals as means for achieving existing goals of their own. However, agents, except for heavenly beings, are never self-sufficient. Autonomy is limited in several ways. An agent depends on a resource when he needs it to perform some action in order to achieve one of his goals. Beyond **resource dependence**, there is **social dependence**: an agent x depends on another agent y when, to achieve one of his goals, x needs an action of y. Agents can even treat other agents as resources. When two agents depend on each other to achieve one and the same goal, they are mutually dependent.

Dependences imply interests: a world state that favours the achievement of an agent's goals is an interest of that agent. Dependence and interest relations hold whether an agent is aware of them or not. Objective relations between two or more agents or between agents and the external world are those relations that could be described by a non-participant observer even if they are not in the participants minds. In that sense, **social interaction has objective bases**: there is social interference between two agents when the achievement of one's goals has some (positive or negative) effects on the other achieving his goals, be those effects intended or unintended by any agent [Conte and Castelfranchi, 1995, pp.20-26,46].

Limited autonomy of social agents comes also from influencing relations between them. By acquiring beliefs about their interests agents can acquire goals. An agent can have true beliefs about his interests, when they overlap with objective interests, and they can help setting goals and planning action. But an agent can also have false beliefs about interests, as well as ignoring some of his objective interests. Furthermore, there can be conflicting interests of the same agent (viz immediate vs. long-term interests). Now, an agent can adopt another agent's goals. If y has a goal g and x wants y to achieve g as long as x believes that y wants to achieve g, we can say that x adopted the y's goal. Goal adoption

can be a result of influencing: y can work to have x adopting some of y's goals. By influencing, new goals can replace older ones. An agent x can influence another agent y to adopt a goal g according to x's needs, even if that goal g is not an interest of y agent [Conte and Castelfranchi, 1995, pp.32,60-61].
So, the bounded autonomy of the agents comes from the relations of dependence and influencing holding among them and between them and the real world. And the bounded autonomy of agents creates interdependent situations among them, so asking for an understanding of how agents must go beyond independent action and engage on coordinated action. We need to understand how the capabilities of agents to change the constraints can lead to better outcomes.

4.1.2 A multilevel approach

The PD game is conceptualized as a non cooperative game (communication among the players is forbidden or impossible - or taken as irrelevant by the model, because it is simply not modelled) in which all players possess complete information (all players know the full structure of the game tree and the payoffs attached to outcomes). In a PD game, each player has a dominant strategy (the player is always better off choosing this strategy - to defeat -, no matter what the other player chooses). When both players choose their dominant strategy, they produce an equilibrium. However, the equilibrium in the PD game is not a Pareto-optimal outcome (we have a Pareto-optimal outcome when there is no other outcome strictly preferred by at least one player that is at least as good for the others). Thus, the equilibrium outcome in the classical PD game is Pareto-inferior. Other ways of dealing with this kind of collective dilemma are in need.
One traditional approach to solving these problems within human societies is to appeal to an external actor (a sovereign), assumed to have some kind of special status, knowledge and access to information, and authority to conceive and impose solutions. Rather frequently, however, the posited merits of a central authority depend on not paying due attention to the unrealistic underlying assumptions, concerning, for example, the accuracy of the information that actor is able to gather to play his role; monitoring capabilities and sanctioning reliability of a central authority; zero costs of administration of that special agent. Institutional Economics has a lot to say about these wrong assumptions and alternative ways of looking for collective coordination.
Other traditional approach to conceive solutions to coordination problems of multiple human agents is to let each individual agent behave on the basis of his vision of his self-interest and assume that social order will spontaneously arise in one or other way. Proponents of laissez-faire strategies often disregard evidence of the existence of types of situations where spontaneous order is unlikely to arise.
The institutional approach to social order does not rely uniquely either on the

merits of a "central authority" nor on the merits of "spontaneous order"[4]. A fundamental element of the institutional approach is a multilevel understanding of the coordination problems a sophisticated social group can face in different situations. The material/physical setting, as well as the prevalent social order, jointly defining the constraints of an action situation, can vary in multiple ways. Sometimes, the relevant constraints are not easily modifiable at relatively short time scales. However, some of the constraints are not fixed and can be changed, so representing opportunities to reframe coordination problems. In this respect, the most usual presentation of the PD game induces deceitful assumptions about the situation: usually people are not in jail, people can communicate, and can change the rules of the game. To understand these opportunities we need to understand the multilevel character of the institutional world.
According to [Ostrom, 1990, pp. 50-55, 60] and [Ostrom, 2005, p.58], it is useful to distinguish at least three institutional levels (rules of three levels):

- **Operational rules** directly affect day-to-day decisions made by participants in a specific setting. These can change relatively rapidly.

- **Collective-choice rules** affect operational activities and results through their effects in making policies (procedures to be used to change operational rules) and in determining who is eligible to be a participant. These change at a much slower pace.

- **Constitutional-choice rules** determine how, and who, and within which limits, can change collective-choice rules. These change at the slowest pace.

If needed, for analytical purposes, the existence of more basic rules (metaconstitutional rules) can be assumed; we can add more basic levels until one gets directly to constraints from the biophysical world (natural constraints are not institutional constraints).
It is worthy to note that the participants in action situations at different levels can be the same individuals or they may differ. The sets of individuals acting at operational, collective-choice, or constitutional-choice level are, rather frequently, sets with different elements. In other situations, individuals who make operational choices also make constitutional choices. In any case, the individuals acting at

[4]Within this approach, cognition does not preclude emergency. To form goals and establish plans to their achievement, agents must be cognitive. Unlike reactive systems, socially responsive systems react not only to (physical) actions of other systems, but also to what they believe their intentions are. However, bounded rationality combines with bounded autonomy to give place to emergent phenomena: there are deliberately planned actions but they may produce unintended effects beyond reach of agent's understanding or awareness [Conte and Castelfranchi, 1995, pp.47,142].

different levels remain of the same kind. "Thus, one should use a similar conception of the individual when thinking about operational and institutional choices." [Ostrom, 1990, p.193] Wrong conceptions about the kind of interactions agents must be able to engage in, can blur the fundamental sameness of participants at all levels. (Participants are heterogeneous, but we find heterogeneity at all levels of action.) Institutional-choice situations, both collective-choice and constitutional-choice situations, affect the rules used in operational situations. This is why agents unable to understand and to act at institutional-choice situations have limited capabilities to pursue their own goals in complex social settings. Limiting agents to immediate interaction (operational level) is to promote the social impairment of agents.

4.1.3 Self-organizing and self-governing

The above mentioned multilevel ontology of the institutional realm is closely related to an important opportunity agents enjoy in their trying to solve coordination problems.

> "(...) self-organizing and self-governing individuals trying to cope with problems in field settings go back and forth across levels as a key strategy for solving problems. **Individuals who have no self-organizing and self-governing authority are stuck in a single-tier world.** The structure of their problems is given to them. The best they can do is to adopt strategies within the bounds that are given." [Ostrom, 1990, p.54, our emphasis]

Ostrom permanently places self-organizing capabilities at a high place in any endeavor to solve coordination problems within collectives. It is, however, worthy to underline that, in Ostrom's writings, self-organization is not a spontaneous or emergent propriety of a collective system. Self-organization is not, in no way, a spontaneous process. Self-organization is a deliberate process and a targeted result of purposive and persistent efforts of individuals highly motivated to avoid some otherwise "natural" (but undesirable) harm or loss. When Ostrom talks of people self-organizing, she is talking about people recognizing that the sustainability of some common resource is at risk; about people analyzing the situation and mobilizing themselves to act in a coordinated way to improve their situation; about people crafting better rules related to local settings. She is not talking about any kind of "emergent" solution coming into earth "spontaneously" or without huge purposive effort from the participants in the concrete situation [Ostrom, 2005,

p.220]. In the same direction, [Ostrom, 1990, p.14] had already wrote that "*getting the institutions right* is a difficult, time-consuming, conflict-invoking process."

4.2 A general recipe for creating situations

Within a systematic multilevel approach to institutional environments, and given the impossibility of dealing simultaneously with all levels and all spatial and temporal scales, Ostrom [2005] takes "action arenas" as its main subject. "Action arenas exist in the home; in the neighbourhood; in local, national, and international councils; in firms and markets; and in the interaction among all of these arenas with others." [Ostrom, 2005, p.13] An action arena is the interaction space of an "action situation" and a plurality of participants. Participants do not have institutional interaction outside action situations; action situations without participants are empty forms; participants animate action situations; action situations frame participants' behaviour. To understand the interaction, we need to study both action situations and participants. We will concentrate first on action situations [following Ostrom, 2005, chapter 2]; the next section will be about participants.

4.2.1 The internal structure of action situations

An action situation can be described and analyzed characterized using seven clusters of variables:

1. the set of participants (who may be either single individuals or collective actors),
2. the positions to be filled by participants,
3. potential outcomes,
4. the set of allowable actions (including the choice not to act) and the function that maps actions into realized outcome,
5. the control that an individual participant has in regard to this function,
6. the information available to participants about actions and outcomes and their linkages, and
7. the costs and benefits - serving as incentives and deterrents - assigned to actions and outcomes.

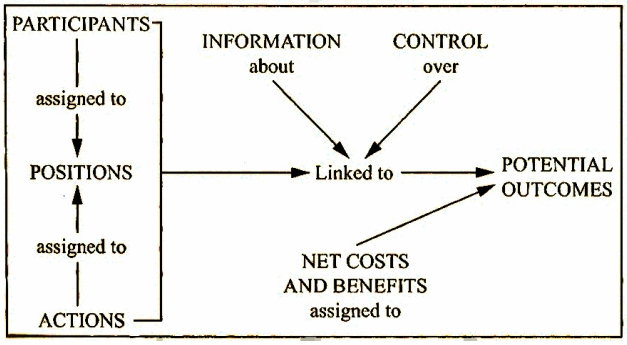


Figure 1: The internal structure of an action situation, in (Ostrom 2005)

Participants, individuals or organizations, are assigned to **positions**. In these positions, they choose among **actions** in light of their **information**, the **control** they have over **action-outcomes linkages**, and the **benefits and costs** assigned to actions and **outcomes**.

Positions are a kind of "anonymous slots" (buyer, seller, judge) into and out of which participants move. The number of positions is frequently fewer than the number of participants. Sometimes, the same participant can occupy some different positions at the same time. Positions are links between participants and actions. To each particular position corresponds a set of actions the holder of the position is authorized to take.

Potential outcomes must be appreciated both from an objective or external point of view (the **physical results** of specific chains of actions by participants; the **material rewards** assigned to actions and results by payoff rules) and from a subjective or internal point a view (the **valuation** placed by a participant on the combination of physical results and material rewards). As to the distinction between physical results and material rewards, it can be exemplified by a situation when a given amount of goods produced by a firm (physical results) will be shared in a unequal way by different participants (owners, managers, workers), so giving place to different material rewards for different participants. How each individual value his share of the results corresponds the subjective term "valuation".

Action-Outcome linkages. Participants can have strong or weak influences on the outcome by knowing the linkages from control variable to state variables and choosing whether to act or not to act in some direction. If, as it usually is the case, the state variable is also changeable by other means (as a result of some physical process, or of some action of other agents), the effect of the participant

choice can be less than decisive. Risk and uncertainty are linked to unknown or nondeterministic transformations of actions into outcomes. Risky or uncertain action- outcome linkages involve one-to-many relationships between actions and outcomes. In risky situations, the agent can know the objective probabilities linking actions to outcomes; in uncertain situations, there an essential indeterminacy. "Outside of formally organized large-scale markets, few interactive situations are likely to have one-to-one relationships between actions and outcomes." (p.48)

Control is about power a participant has in a situation. "The 'power' of an individual in a situation is the value of the opportunity (the range in the outcomes afforded by the situation) times the extent of control." (p.50)

Information. Participants in an action situation may have access to complete or incomplete information. Complete information is an assumption that each participant could know the full structure of an action situation, including number of other participants, the positions, the outcomes, the actions available, the actions-outcomes linkages, the information available to other players, and the payoffs for them. Aside this distinction, information can be perfect or imperfect. Information is perfect if each participant can know, not only their own past actions, but also the possible actions of all other players before they make any move. When information is less than complete, the question of who knows what at what juncture becomes very important. The difficulty to gather information links to the possibility of opportunism: "The opportunism of individuals who may say one thing and do something else further compounds the problem of incomplete information." (p.51)

Costs and Benefits. In addition to the physical results of a chain of actions, there is also the question of what rewards or sanctions will be distributed to participants depending on their participation. We need to distinguish physical results, material (external) rewards, and (internal) valuation. External rewards may be assigned on action variables (how many hours the worker clocks in), on outcome variables (how much of a product is produced), or on a combination of both. In economics and game theory the internal (subjective) value assigned by participants to the achievement of an outcome is referred to as utility. "Utility is a summary measure of all the net values to the individuals of all the benefits and costs of arriving at a particular outcome. (...) For simplicity, many analysts assume that subjective utility is positively associated with the net value of the external rewards. In economics, theorists normally assume that utility is monotonically associated with profits, for example. (...), this assumption is reasonable to make in many but not all situations. (...) Measuring intrinsic valuation is ex-

tremely challenging." (p.53)

4.2.2 The exogenous variables: biophysical and social world

Institutional aspects of an environment do not work in the void: the perhaps most challenging aspects of the action situation come from upstream and downstream: on the one hand, the exogenous variables characterizing the material and social world; on the other hand, participants' behaviour and internal world.

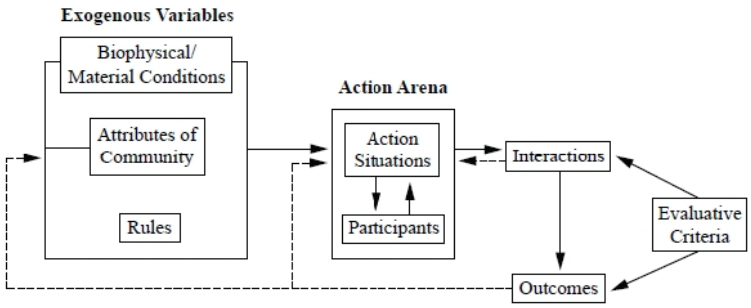


Figure 2: A framework for institutional analysis, in (Ostrom 2005).

The **exogenous variables** affecting the structure of an action arena include three clusters of variables:

(1) the attributes of the *biophysical world* that are acted upon in these arenas:

> "(...) some of the variables of an action situation (and thus the overall set of incentives facing individuals in a situation) are also affected by attributes of the biophysical and material world being acted upon or transformed. What actions are physically possible, what outcomes can be produced, how actions are linked to outcomes, and what is contained in the actors' information sets are affected by the world being acted upon in a situation. The same set of rules may yield entirely different types of action situations depending upon the types of events in the world being acted upon by participants. These 'events' are frequently referred to by political economists as the 'goods and services' being produced, consumed, and allocated in a situation as well as the technology available for these processes." (p. 22)

(2) the structure of the more general *community* within which any particular arena is placed:

"The attributes of a community that are important in affecting action arenas include: the values of behavior generally accepted in the community; the level of common understanding that potential participants share (or do not share) about the structure of particular types of action arenas; the extent of homogeneity in the preferences of those living in a community; the size and composition of the relevant community; and the extent of inequality of basic assets among those affected. The term *culture* is frequently applied to the values shared within a community. Culture affects the mental models that participants in a situation may share. Cultures evolve over time faster than our underlying genetic endowment can evolve." (pp. 26-27)

(3) the *rules* used by participants to order their relationships:

"(...) rules [are] shared understandings by participants about enforced prescriptions concerning what actions (or outcomes) are *required*, *prohibited*, or *permitted*. All rules are the result of implicit or explicit efforts to achieve order and predictability among humans by creating classes of persons (positions) who are then required, permitted, or forbidden to take classes of actions in relation to required, permitted, or forbidden outcomes or face the likelihood of being monitored and sanctioned in a predictable fashion." (p.18)

4.2.3 A general recipe for creating situations

Given that rules affect the structure of an action arena, and also that rules can be less difficult to modify in desired directions (at smaller time scales) than the biophysical world or even the community culture, we can think of changing an action arena by modifying the rules more directly affecting some of its elements. If we take the elements of an action situation and modulate the rules, as exogenous variables that affect particular elements, we can think of creating situations - and of a "general recipe for creating situations" [Ostrom and Crawford, 2005, p.183].

Crawford and Ostrom [2005] proposed a general syntax of the grammar of institutions (grammar of rules). That grammar considers two basic kinds of rules: **generative rules**, of the form "let there be an X" (like rules that create organized bodies or positions); **regulatory rules**, with a syntax elucidated in the following. The general syntax of the grammar of regulatory rules includes five components:

A - ATTRIBUTES - any values of participant-level variables describing to whom the institutional statement applies (examples: more than 18 years of age, female, employer); a blank slot here (the default value) means "all members of the

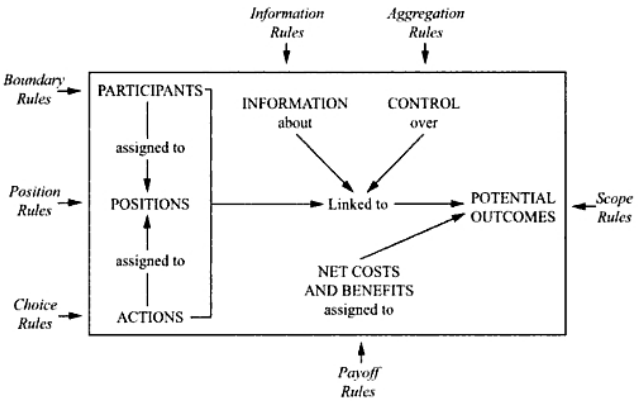


Figure 3: Rules as exogenous variables directly affecting the elements of an action situation, in (Ostrom 2005).

group";

D - DEONTIC - one of the deontic operators "may" (permitted), "must" (obliged), or "must not" (forbidden), assigned to actions or to outcomes (restricted to what is presumed to be physically possible); note that the three deontic operator are interdefinable;

I - AIM - the particular amount of action or outcomes to which the Deontic is assigned; to influence behaviour, both the AIM and its negation must be physically possible;

C - CONDITIONS - describes when, where, how, and to what extent an AIm is permitted, obligatory, or forbidden; the default value is "all times in all places";

OR ELSE - the consequences for not following a rule, for example a possible sanction for violation; the consequence must be the result of collective action (a collective decision); a mechanism for implementing the stated consequence must be in place, built-in some other rule (without the establishment of positions with the obligation and the authority for monitoring and sanctioning, the occurrence of "or else" in a phrase does not constitute an OR ELSE clause); the responsibility of monitoring the conformance of others must have a price, affecting the opportunities of those in charge.

This is the general form of a regulatory rule:

ATTRIBUTES of participants who are OBLIGED, FORBIDDEN, or PERMITTED to ACT (or AFFECT an outcome) under specified CONDITIONS, OR ELSE.

With this syntax, we can more accurately distinguish between different institutional statements. In particular, Crawford and Ostrom propose to distinguish between **strategies**, **norms**, and **rules**:

All regulative rules can be written as:

[ATTRIBUTES] [DEONTIC] [AIM] [CONDITIONS] [OR ELSE]

All norms can be written as:

[ATTRIBUTES] [DEONTIC] [AIM] [CONDITIONS]

All shared strategies can be written as:

[ATTRIBUTES] [AIM] [CONDITIONS]

Now, if we want to create a situation, by changing rules as mentioned above, it is useful to classify rules according to the element in the action situation that they most directly impact. Using the AIM of the rule for classification gives us:

- position rules: create positions;
- boundary rules (also called entry and exit rules): affect how individuals can become members, how they are assigned to or leave positions, either in a voluntary or a compulsory basis; also affect how one situation is linked to others, by the way of linkages of positions;
- choice rules: affect the assignment of particular action sets to positions: what an individual in a position must, must not or may do in such and such conditions, so affecting rights, duties, liberties and their relative distribution among members, so affecting power; one particular type of choice rules, "agenda control rules", define who can propose particular actions, and so limiting the alternative actions available to the group;
- aggregation rules: affect the level of control that individual participants exercise at a node in a decision process, for example allowing a single participant to take a decision or establishing how multiple participants must proceed to have a collective decision (who can/must participate, how to weight votes, how to recognize when a valid decision has been taken);

- information rules: affect the level of information available to participants, about the overall structure of the situation, the current state of individual variables, the previous and current moves of participants; establish channels of information and the rights and duties related to communicating information;

- payoff rules: assign external rewards or sanctions to particular actions or outcomes (e.g., pay schedule assigning salaries to participants in different positions);

- scope rules: affect which outcomes must be affected within a situation.

Two cautionary notes are worthy for users of recipes for creating situations. First, rules operate together, as a configuration - and rather frequently theres is no easy way to predict its combined effects. Second, in the real world, inconsistencies can arise: "The partitioning of actions can (...) be complicated by complex sets of rules that may be inconsistent in their ordering of actions with different rules assigning different DEONTICS to the same action. One rule may forbid an action, while another rule requires that same action." [Ostrom and Crawford, 2005, pp. 200-201]

4.3 Individual Agents

4.3.1 The internal world of individual choice: much more complex than the rational egoist

Much of the contemporary work on the micro-macro link problem can be said to embody these assumptions:
1. Individuals possess as much information about the structure of a situation as is contained in the situation itself.
2. Individuals assign a complete and consistent, internal evaluation to outcomes that is a monotonic function of an individual's own net external payoff. (Classical utility theory did not make this assumption, but it had to be used to have some specific assumption on where utility come from.)
3. After making a complete analysis of the situation, individuals choose an action in light of their resources to maximize expected material net benefits to themselves given what others are expected to do.
These assumptions describe a rational egoist [Ostrom, 2005, p.101]. Institutional Economics found it necessary to use some alternative assumptions to face the strong unrealistic flavour of this picture of real agents in real situations in the real world. The model [Ostrom, 1990] suggested do the internal world of individual

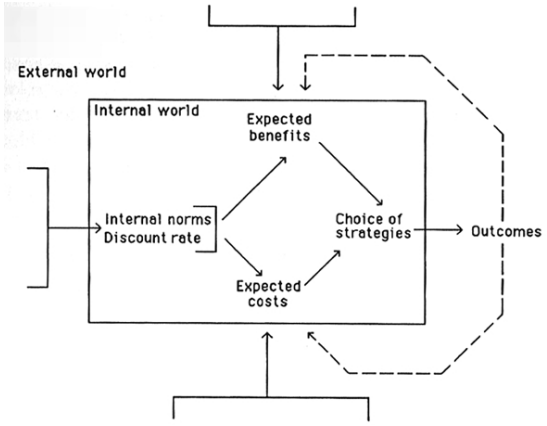


Figure 4: The internal world of individual choice, in (Ostrom 1990)

.

choice is a step into that direction.

Four internal variables affect individual choices:

- expected benefits
- expected costs
- internal norms
- discount rates

These four internal variables affect individual's choices. Individuals jointly produce outcomes in the external world. The external world not always confirms expectations. Found outcomes impinge on future expectations concerning benefits and costs of actions. Internal norms affect choices - and are affected by norms held by others. Discount rates are affected by the range of opportunities that an individual has outside a particular situation. Discount rates **(how opportunities are perceived)** and norms **(how norms are perceived)** are sources of heterogeneity within a population (see below).
It is crucial to note that, in contrast to the three dominant assumptions above mentioned, agents cannot be assumed to know the action-outcome linkages (how specific sets fo actions lead to specific outcomes). The difference between "objective" and perceived circumstances is crucial due to information incompleteness (see chapter 5).

4.3.2 Individual diversity within a population of social agents (1): Discount rates

One important aspect of the "internal world" of the individual agent is the source of variety represented by individual's discount rates. We will first introduce the concept of discount rate in financial terms, and then appreciate its importance to understand the workings of institutional environments.
The discount rate is needed for the computation of a present value: the current value of a future payment (or series of future payments), discounted to reflect the time value of money and other factors.
Let's start with a more familiar concept: interest rates. For example, EUR 200 today is worth more than EUR 200 two years from now, because we can put it in the bank and earn some interest. So, interests payed by banks are a reason to prefer the same amount now than in a year from now.
Interest rates are to compute future values. Say that

present value = 200

a = number of periods = 2

r = interest rate in a reference period = 5% a year

So,

$$200 \times (1,05)^2 = \text{future value in n periods with rate r}$$

Or, more generally,

$$\text{Future value} = \text{Present value} \times (1 + \text{interest rate, in a given period})^{\text{a = nb. of periods}}$$

Time preference is a more fundamental reason than bank payed interests: income received in the future is worth less now than income received now.
Now, let's use the same reasoning, except in reverse, to answer the question: How much would we need today to have EUR 200 in one year?
To calculate the present value of a future (income) amount **a** years in the future:

Present value = (Future value) / $(1 + \text{interest rate})^a$

When an interest rate is used in reverse like this, to calculate how much we need now to have a certain amount later, economists use the term discount rate rather than interest rate. So:

Present value = (Future value) / (1 + discount rate)^a

Example. The present value of EUR 200 in **a** years, at a 5% discount rate:

a	6	5	4	3	2	1	0
EUR	149,24	156,71	164,54	172,77	181,41	190,48	200,00

Note this fundamental relationship: when the discount rate goes up, present values go down; when the discount rate goes down, present values go up.
In financial terms, the discount rate measures the opportunity cost of capital (how much interest you could earn on your money if you put that money away). The application of the concept to social dynamics also relates to different opportunities enjoyed by different sets of participants.
Now, suppose we want to understand how a group of individuals act as interdependent appropriators of a natural (or man-made) resource system that is sufficiently large as to make it costly (and in some cases infeasible) to exclude one potential appropriator from improvements made to the resource system. Some examples are communal tenure in high mountain meadows and forests; irrigation systems in semiarid regions with extreme variation in rainfall from year to year; an inshore fishery, with a traditional small operation of a reduced number of local fishers in small boats. Renewable resource systems have problems of sustainability: to the resource to be sustained over time, the average rate of withdrawal must not exceed the average rate of replenishment. For some cases, investments made in maintenance and repair can improve sustainability. This kind of resources has subtractive attributes: the fish harvested by one boat are not there for someone else. Crowding effects and overuse are chronic in this kind of resource systems. These are reasons to ask for coordinated action of appropriators, given the risk of opportunistic behaviour.
How different individuals discount future benefits in different ways is a critical element of the dynamics of this kind of situation. Different discount rates depend on several factors, all related to different time horizons and different opportunities enjoyed by the agents. Time horizons are affected by whether or not individuals expect that they or their children will be present to reap the benefits of the re-

source, as well as by opportunities they may have for more rapid returns in other settings. "In a fishery, for example, the discount rates of local fishers who live in nearby villages will differ from the discount rates of those who operate the larger trawlers, who may fish anywhere along a coastline. The time horizons of the local fishers, in relation to the yield of the inshore, extend far into the future. They hope that their children and their children's children can make a living in the same location. More mobile fishers, on the other hand, can go on to other fishing grounds when local fish are no longer available. Discount rates are affected by the levels of physical and economic security faced by appropriators. Appropriators who are uncertain whether or not there will be sufficient food to survive the year will discount future returns heavily when traded off against increasing the probability of survival during the current year." [Ostrom, 1990, pp.34-35]

Discount rate can impact relational variables. Where the population in a location has remained stable over long periods of time, and individuals have shared a past and expect to share a future, this affects expectations. For example, within communities organized around agricultural resources, people expect their children and their grandchildren to inherit their land: their discount rates are low. This affects the importance of reputation: within such stable populations, reputation as reliable members of the community is important for individuals, because reputation impacts their opportunities [Ostrom, 1990, p.88].

An interesting point is that changing the discount rates can change the whole logic of the situation. If the boundaries of a common natural resource (a fishery, for example, or an irrigation system), or the specification of those authorized to use it, are not clearly defined, local appropriators face the risk of outsiders taking undue benefits of their efforts to preserve the resource. Local appropriators can, for example, coordinate patterns (reduce) appropriation and provision, to stay at sustainable levels of exploration and to prevent destruction of the resource. This effort from local appropriators is a meaningful move because of their long term expectations about the resource: their discount rates are low. If outsiders, without long term links to the local community, act in a destructive way (fast appropriation), local appropriators can abandon their commitment to the prudent behaviour and start withdrawing units as fast as they can. The temporal horizon has changed, the long term expectations vanished, and immediate consumption becomes the only alternative. This situation pushed the discount rates of appropriators toward 100%. The dominant strategy of all participants is now to overuse the resource. The situations becomes like a one-shot prisoners' dilemma [Ostrom, 1990, p.91].

Another interesting point is that different discount rates affect not only collectively managed resources, but also the management of private owned resources. Even private owners of a share of a scarce resource will overuse the resource if their discount rates are high [Ostrom, 1990, p.219].

If, for modelling purposes, we need to take into account different expectations, time horizons, and opportunities among participants in an action arena, while not being able to model specific reasons for that at micro-level, discount rates can be useful to model distributions of different profiles within a population.

4.3.3 Individual diversity within a population of social agents (2): Norms and delta parameters

A social order can be enforced by specific external effects that individuals, on the basis of past experience of the workings of monitoring and sanctioning mechanisms, expect to happen as consequences of some of theirs (and others') actions. Internal or subjective mechanisms can also contribute to enforce a social order. To understand agents in institutional environments, we need to include complex motivations in the modelling of their behaviour (see 3.2.2., about Searle on deontic relationships providing reasons for action that are independent of desires). These internal mechanisms of complex motivations can be broadly called "moral disposition" of agents.
Moral disposition of agents can be seen as willingness to follow shared rules. (An example, from experiments within MAS, is the rule "tell the truth".) In this sense, moral dispositions can be modelled by variables of the internal mechanism of individual agents, and, for artificial agents within robotic experiments, those variables can be manipulated in several ways. Moral variables can, for example, regulate the correlation between different behaviours (e.g., how much to contribute; what contribution to announce). Moral disposition is unevenly distributed over the set of agents, and we can experiment on the consequences of different distributions. The individual moral disposition can be conditional on the available information about the moral disposition of the group (and this can be linked to monitoring mechanisms).
This view on the role moral considerations can play within institutional environments can be enlightened if taken in conjunction with one element of the internal world of the agent, as modelled by [Ostrom, 1990]. Norms are expected to change the internal value participants place on an action or outcome in a situation. Norms can definitely change individuals' behaviour - but may not do so, depending on the relative size of the costs and benefits following or not following a norm represent to an individual agent. Crawford and Ostrom [2005] give a concrete translation of institutional statements into the payoff structure of individuals, using delta parameters. **Delta parameters are used to represent the intrinsic benefits or costs of obeying or of breaking a prescription in a particular situation.** This is necessary to take into account that not all individuals value the same way the effects of breaking or following a rule. Delta parameters are added

to an individual's payoff to represent the perceived costs and rewards of obeying or breaking a prescription. The delta parameters are defined as:

$\Delta = \delta^o + \delta^b$, where

Δ = the sum of all delta components

δ^o = the change in expected payoffs from *obeying* a prescription

δ^b = the change in expected payoffs from *breaking* a prescription

To be able to use the distinction between "internalized norms" (like guilt or shame for deviant behaviour) and "externally sanctioned norms" (including phenomena related to reputation), the mentioned rewards and costs can be divided into those arising from an external and those arising from an internal source of valuation:

$\delta^o = \delta^{oe} + \delta^{oi}$ and $\delta^b = \delta^{be} + \delta^{bi}$, where

e = changes in expected payoffs originating from external sources

i = changes in expected payoffs originating from internal sources

Participants can have different orientations regarding these effects: one can perceive costs of breaking a prescription (δ^{bi} or δ^{be}) to be high while other perceives them to be low.
Delta parameters can be used also to model individuals in charge of enforcing duties. A situation that includes an enforcing mechanism must specify the effect of the monitoring (or not monitoring) on the monitor, delta parameters (and possibly OR ELSE parameters) becoming part of the monitor's payoff formulas. For example: important social pressure to monitor or sanction can be modelled by large δ^{oe} and δ^{be} for those in charge; strong moral commitment on the part of the monitors to their responsibilities can be modelled by large δ^{bi} and δ^{oi}; the reward for monitoring and sanctioning must be high enough to offset the costs.
Recognizing the existence of, on the one hand, norms which effectiveness depends mainly on external enforcement mechanisms, and, on the other hand, norms which effectiveness depends mainly on the internal valuation of obeying or breaking a prescription, does not mean that external enforcement and internal "moral sense" are mutually exclusive. The recognition of the legitimacy of rules by participants will positively impact enforcement and compliance (higher level of compliance with less

enforcement efforts). With rules imposed primarily by force, individuals subject to these rules are unlikely to develop internal delta parameters associated with breaking or following them. Without a significant level of voluntary compliance to rules, cost of monitoring and sanctioning will be prohibitively high for all the system.
At least partially, the effect of legitimacy is linked to a feature of norms that deserves some attention: not always norms are subject to a computation of consequences. As Ostrom [1990, p.35] puts it: : "Norms of behaviour reflect valuations that individuals place on actions or strategies in and of themselves, not as they are connected to immediate consequences."
Norms can raise fundamental problems when used to deal with artificial agents. If agents must comply with norms automatically, they are not seen as autonomous any more. If they can violate norms to their own advantage (e.g. to maximize utilities), the advantages of normative approach evaporate and the normative framework does not stabilize the collective. Alonso [2004], recognizing this problem, argues for using rights and argumentation in MAS. He suggests that the concept of rights offers a middle way to escape the dilemma. Individuals have basic rights to execute some sets of actions (under certain conditions), but rights are implemented collectively. Agents are not allowed to inhibit the exercising of others' rights and the collective is obliged to prevent such inhibitory action. Rights are not piecemeal permissions; they represent a system of values. Nobody can trade with rights (even its own); rights are beyond utility calculus. Systems of rights do not eliminate autonomy. Because they are typically incomplete or ambiguous, some argumentation mechanism must be at hand to solve underspecification problems. How artful this suggestion can be, it cannot provide real free-will to programmed agents. (It is not even sure that human beings have "real" free-will.) Notwithstanding the difficulties any project of providing free-will to artificial agents will face, at collective level we can perhaps be able to model its effects experimenting with different distributions of delta parameters within a population of these agents.

4.3.4 Habits, routines, and Institutional Economics

Geoffrey Hodgson has given some attention to the role played by habits in a broad institutional approach to human nature and society (see, specifically, Hodgson, 2004, 2007). Habits are part of a more general aspect of the working of institutions. Institutions enable ordered thought and action by imposing form and consistency on human activities. That way, institutions both constrain and enable behaviour. They impose constraints: traffic rules do not let us drive freely through the streets of the town. They also facilitate coordination: traffic rules help

traffic to flow more easily and safely[5]. The constraints can open up possibilities, enabling choices that otherwise would not exist. The capacity to form habits has evolved to deal with uncertainty and complexity, to be able to respond to variable circumstances without analysing any detail each time. Habits are part of this process: institutions constrain our behaviour and we acquire habits consistent with the operation within these constraints; upon new habits of thought and behaviour, new (more conformist) preferences and intentions emerge, reinforcing already in operation institutions.

To understand the relationship between institutions and individual agency we need the notion of habit, habit being a core dispositional mechanism underlying actions and beliefs. According to [Hodgson, 2007, p.106]), "Habits are submerged repertoires of potential thought or behaviour, to be triggered by an appropriate stimulus or context." Habits are foundational to all thought and behaviour. Even deliberation, including rational optimization, relies on habits and rules.

Repeated behaviour is important in establishing a habit. However, habit and behaviour are not equivalent: the acquisition of a habit does not mean the necessary use of it all the time. The repeated behaviours, that favour the formation of habits, sometimes are triggered by innate dispositions, and often result from the propensity to imitate the way other people faces constraints in social situations. Many habits are unconscious. But instincts and habits are different mechanisms. Instincts are biologically inherited; habits are formed after birth, are socially learned, and are adaptable to changing problems. Compared to instinct, habits are relatively flexible - but, still not excessively costly - means of adapting to complexity, disturbance and unpredictable change.

Habits can give raise to rules. For a habit to acquire the status of a rule, it has to acquire some inherent normative content, to be potentially codifiable (made explicit), and to be prevalent among a group. Habits respond to reiterated circumstances and constraints they represent. Rules, as part of those circumstances and constraints, help to create habits and preferences. Habits, on the other side, reinforce the rules they conform to, enhancing their durability by channelling conformist behaviour. Producing conformism and normative agreement, habits mould agent's preferences.

Habits, removing some actions from conscious deliberation, help to economizing on decision-making. But this does not mean that institutions directly, entirely or uniformly determine individual preferences. In Hodgson's view, habit is not a deterministic device: habit is "a disposition, which, once acquired, is not necessarily realised in any future behaviour. Habit is a causal mechanism, not merely a set

[5][Hodgson and Knudsen, 2004] is perhaps the first computer simulation of extensive interaction between agents and structures, where the preferences of agents are altered by institutional circumstances just as institutions are developed by agents. This is an agent based simulation of the emergence of a traffic convention to drive on the left or the right side of the road.

of correlated events." [Hodgson, 2004, p.653] However, the author does not give a concrete explanation on how habits work differently from strict rules of behaviour. Habits are part of the preferences of each agent, they change with experience, but there is mutual influence between structure and individual agency. On the one hand, institutions depend for their existence upon individuals, and it is sometimes possible for individuals to change institutions. This could be described as 'upward causation'. On the other hand, institutions, by structuring, constraining and enabling individual behaviours, mould the dispositions and behaviours of agents, change their aspirations. This could be named 'reconstitutive downward causation'. This is not done directly, in a deterministic way, because reconstitutive downward causation does not operate on behaviour (action), but rather on habits (propensities or dispositions).

Organizations and Routines

Habituation is a social mechanism, in that it typically involves either the imitation of others' behaviours, or behaviour repeatedly constrained by others. One interesting extension of this approach is about the workings of habit-based behaviours within organizations [6].
Hodgson [2006] makes a terminological distinction between social structures, institutions and organizations:

- *Social structures* include all sets of social relations, including the episodic and those without rules, as well as social institutions. (An example of a social structure that is not an institution is a demographic structure.)
- *Institutions* are systems of established and embedded social rules that structure social interactions.
- *Organizations* are special institutions that involve (a) criteria to establish their boundaries and to distinguish their members from non-members, (b) principles of sovereignty concerning who is in charge and (c) chains of command delineating responsibilities within the organization. (Language is an example of an institution that is not an organization.)

So, organizations are bounded institutions with a relatively high degree of cohesion, necessarily involving the development of accordant individual habits. At the same time, their characteristics enhance the possibilities for more intensive interactions between individuals in the same organization, carrying a richer repertoire

[6] [Hodgson, 2007, pp.110-111]

of opportunities for creating habits. Routines are structures for habits within organizations.
A routine is a generative structure within an organization, with means to trigger conditional patterns of behaviour among of individuals within an organized group, involving sequential responses to cues.
Within a structured group of individuals, each has habits of a particular kind, related to his role as a member of the organization and the specific tasks he must carry out when playing such a role. Specific behaviours induced by these habits send cues by some of the other individuals within the organization, these behavioural cues triggering specific habits in others. The responses to these behavioural cues usually are facilitated by procedural memories that have been shaped by precedent events within the same organization (learning the customary way of doing things). Various individual habits sustain each other in an interlocking structure of reciprocating individual behaviours, both of the explicit and of the informal kind. Technological and physical artefacts, as well as social specifications (roles, for instance), are part of the environment of routines. The organizational environment usually is specific to a particular institution; it is not a general solution for all the organizations of the same kind or with the same goal. If one person leaves the organization and is replaced by another, the new recruit cannot just apply his general knowledge on the tasks he has to perform: he has to learn the habits that are required to maintain specific routines within that particular organization.

The establishment of habits

We should specifically concentrate on the cognitive mechanisms responsible for the forming and establishment of goal-directed habits. Psychology and Social Psychology can be of help here. The kind of details we endeavour to explore can be instantiated by the suggestions made by [Danner, 2007] on the role of inhibitory processes within the basic mechanisms of habit.
It seems that ample evidence exists nowadays that people are capable of rather complex behaviours and of automatically selecting and performing a specific goal-directed behaviour without considering all possible options that may also serve as means to attain their goal. (To take an example, in the morning I may grab my belongings and walk towards my bicycle to take my standard route to work, all without much conscious thought as I have carried out these behaviours on many previous occasions.) However, to be able to perform goal-directed behaviour in a habitual fashion, one must first form the habit.
Habits, goal-directed habits, as knowledge structures, are formed by repetitive selection of some means to attain some goal. People, when confronted with the

appropriate situational cues, are capable of engaging automatically in goal-directed behaviour, without being aware of activation and pursuit of the goal. Habits are shaped by the individual's personal history, because they are the result of having performed the same "means for a goal" selection many times before.
The process begins as a deliberative one. Goal attainment can often be realized using different unique means. At first encounters with a goal in a given situation, the individual reflect on all (or some of) the possibilities and potential consequences and decide on a choice of means. Repeatedly selecting and using the same means for the same goal, will reduce the need to exert conscious processes for goal pursuit.
Next instigation of the goal on subsequent occasions increases the probability of retrieving the means from memory that was previously selected. It is therefore likely that the goal will be pursued with the same means. A kind of a "cognitive preference" is formed to retrieve and select this means again for future pursuit of the same goal. Probably other alternative means remain available for the pursuing of the same goal, even if they are no longer considered as options by the individual. Those alternatives have been inhibited: the continuing presentation of several alternatives will interfere with the habit and reduce its usefulness. Through the inhibition of these alternatives, it is easier and more likely that the target means will be found and retrieved from memory. Hence, inhibitory control may be crucial to instigate habit formation by reducing the activation of the alternative means before they reach conscious thought. Inhibition enables the individual to override some memories by preventing these from entering awareness and disturbing working habits with the consideration of non habitual alternative means. Inhibitory control of interfering information that otherwise hinders current cognitive processes occurs without explicit intent to put the alternative means out of mind. For instance, when being asked how one wants to get to work, it is easier to explain a specific route when access to the memory of other routes is inhibited. Access to some memories is reduced to avert interference with ongoing memory processes. Furthermore, inhibition has been found to shield goal pursuit by reducing access to the mental representation of alternative goals.
Taking into account different mechanisms for selecting means for goals, another situation must be considered. People can form the intention to pursue a habitual goal with non-habitual means. But intentions to use non-habitual means for goal pursuit can be hampered by one's habits. The habit interferes on the intention of originating deliberative conscious selection of means. Inhibition can prevent this hindrance, this time reducing access to the habitual means.
Another interesting contribution from Danner is on the role of context stability in the formation and operation of habit. She writes that, despite different perspectives on habits, there is general agreement on the notion that environmental

cues can trigger behaviour directly without the involvement of intentional processes. So, the context, not only frequency of past behaviour, plays a crucial role in strengthening habits. The point is the stability of the context. Danner identifies three factors that render a context stable: past behaviour has to occur always in the same location, at the same time and in the same situation. For example, one always drinks beer when socializing with one's friends (situation) in the pub (location) each Friday night (time). Hence, when these factors are always consistent when the behaviour is carried out, the context is stable. This line of reasoning suggests that people are more likely to rely on intentional processes when they rarely perform the same behaviour in the same context, or regularly perform the same behaviour in different contexts, as the context is either less strongly or less uniquely linked to the behaviour. For example, a person drinking white wine sporadically during the past four weeks at the same place (e.g., a pub) in the same social setting (e.g., being with friends on a Friday night) may rely on conscious intention to initiate the behaviour to a similar extent as a person frequently drinking white wine in the same period at different places (e.g., a pub, restaurant, at home) in different social settings (e.g., being with friends, having a business meeting, spouse's birthday party).
The findings of [Danner et al., 2008] indicate that frequency of past behaviour does not necessarily result in habitually driven behaviour. Specifically, they demonstrated that frequency of past behaviour moderated the intention-behaviour relations only when information about the stability of the context in which the behaviour has been performed is represented in a habit measure: intentions do not guide behaviour when it is frequently performed in a stable context (i.e., strong habit), while intentions are more likely to guide behaviour under conditions of either infrequent performance or unstable, variable contexts (i.e., in both cases there is a relatively weak habit). These findings are important as they show that the context in which the behaviour is performed plays a crucial role in the establishment of habits. Behaviour can be performed very frequently in a given time span, but as long as the context - that is, place, time and situation - in which the behaviour is executed always differs instigation of the behaviour is dependent on intentions. Similarly, although behaviour is always performed in the same context, intentional processes will still guide behaviour when performance of the behaviour only occurs occasionally. In both cases, people seem to be more prone to rely on their conscious thought and intent to produce the behaviour.

5 The Challenge of Incompleteness

This chapter analyses some of the fundamental challenges represented, either to the analysis of human societies or to the design of artificial societies, by the pervasive practical and theoretical consequences of incompleteness.
Incompleteness is about a basic feature of our epistemological condition as human beings. In the real world of collective action facing natural and social uncertain and complex environments, there is absolutely no situation where we could be able to collect all information, neither about current status nor about future values of relevant variables of the system we are in. Incompleteness is about how difficult it can be to search, organize and analyze information before it can be of any use for agents. Incompleteness is about how dependent any agent, or any group of agents, is on their peers to get informed about what is going on now and about what they can expect in the future - because the future in part depends on the other's actions. Incompleteness is about how fool it would be to try to design ex ante mechanisms to govern all ex post eventualities of any relationship we plan to engage in. In the real world, complex contracts are always incomplete, its implementation always face disturbances, which contingencies we are never fully capable of anticipating - so, the only reasonably way agents have to face such a world is to craft governance structures able to solve future impasses, not trying to have a God's view of the future.
Incompleteness, if taken at its fundamental meaning, must strongly impact any vision about the management and control of any kind of system. We are no less able to predict the exact behaviour of a man-designed complex "artificial society" than we are to predict the exact behaviour of a human group. If this is understood, we can concentrate on designing governance structures, rather than trying to anticipate all relevant details of a complex function describing the behaviour of a collective system.
Section 5.1. introduces the basic concept of incomplete information and, with the help of a case study, also introduces the related concepts of contingent strategies and of sequential, incremental, and self-transforming processes. Rudiments of incomplete contracts, as a practical application of incompleteness to decentralized economies, are also presented in this section, which ends with a generalisation of the concept: incomplete institutions.
Section 5.2. deals specifically with what can be seen as the main result the concept of incompleteness gave to institutional economics: transaction costs economics. The concepts of transaction and of transaction costs are framed by transaction costs economics, a field of research developed by the New Institutional Economics.

5.1 Incomplete Information, Incomplete Contracts, Incomplete Institutions

Considering the internal world of the individual choice (see previous chapter), it may seem a simple thing to predict individuals' decisions. To do so, we would "just" need to know the values of "summary variables": the values, for alternative scenarios, of expected benefits, expected costs, shared norms (norms shared by other relevant individuals influence internal norms) and opportunities (discount rates depend on the acknowledge range of opportunities that may or may not be available outside a particular situation). This easiness is, actually, a dream of some rationalist accounts of real action of real agents in the real world.

5.1.1 Incomplete Information

In practice, things are much harder than rationalist accounts can imagine. First, accurate measures for each summary variable are not freely available in the wild. For example, benefits-costs analyses depend on investing resources to obtain information - not to mention the fact that these analyses are often blind to benefits and costs that are not monetized. Second, individuals neither are attentive to all available information nor are prepared to weight that information in an objective and unbiased manner. Third, since other individuals can behave in a strategic, not in a straightforward, manner, one cannot compute their behaviour just as a function of the objective opportunities offered by the situation. So, any prediction assuming the availability of the values of summary variables, without knowledge of the situational variables affecting summary variables, is vacuous.
Some examples of how situational variables affect "summary variables" are as follows. In complex situations, even considering only the physical features of the environment, information is difficult and costly to obtain. Rather frequently, to gain an accurate image of the environment requires heavy investments. Moreover, when information about the physical environment depends on information about the behaviour of agents (e.g., fisher's activity on a fishery), strategic reasons can lead agents to withhold information. Some possible states of the world are not "facts" that exist independently of agents, and so information about the situation: the presence of some kind of organization based on voluntary cooperation can prove instrumental to obtaining and disseminating information, not due to intrinsic features of the information, but due to characteristics of the social setup framing information production and provision. **Not only information must be searched for, organized, and analyzed by some agents, but also the external world information is about is not always completely independent from the same agents' behaviour.** The concrete reachability of some future

states of the world often depends on agents' dispositions to behave in such and such manners. For example, the cost of transforming the status quo by adopting a new set of rules is not independent of the strategies individuals adopt along the transformation process: when individuals adopt confrontational strategies, transformation costs usually rise sharply.
All these ask for a deep understanding of **the epistemic and pragmatic consequences of incomplete information**. Because full assumption of incomplete information, as a basic feature of most actions situations, defies deeply-rooted rationalist accounts, it could be helpful to use a case study allowing us to be as concrete as possible in making this point. The case study examines the origin of a set of institutions to manage a series of groundwater basins located beneath the Los Angeles metropolitan area [Ostrom, 1990, pp.104-133].

5.1.2 Case Study: The Competitive Pumping Race

In such a semiarid region, groundwater basins, combined with surface supply systems, are extremely valuable. Not only they are sources of inexpensive and high-quality water, but also serve as natural storage vessels that can retain water for use during periods of peak demand. Groundwater basins can be destroyed by overextraction or pollution. If more water is withdrawn per year than the average level of replenishment (the "safe yield" of a basin), eventually the gravel and sand in the water-bearing strata will compact so that they cannot hold as much water as they formerly did. In coastal areas, if water level is drawn down below sea level, saltwater intrusion will occur and eventually affect the entire basin.
Before the institutional changes to be described here, water rights in the area were defined for two types of water users: overlying landowners (owners of overlying land, withdrawing water to use it on that land), and appropriators (withdrawing water to other purposes). Water rights of overlying landowners had already been changed by courts: they no more had absolute rights over the water they were able to withdraw, but, especially during a time of shortage, only rights for a proportionate share of the water, limited by the flow of water each would be able to put to beneficial use. Appropriators (like private and public water companies) were allowed to withdraw "surplus water", water not being put to beneficial use by overlying landowners. Among appropriators, the doctrine "first in time, first in right" would exclude mostly junior appropriators.
Prescriptive rights for water (acquired by open and adverse use) made the situation more complicated. A new appropriator taking water continuously for more than five years could lead to two completely different situations. If, going to court, he was seen as someone withdrawing surplus water during that period, he would be classified as a junior appropriator with water rights inferior to everybody else.

If the court decides that he had been withdrawing nonsurplus water (adverse use), he would be recognized as having perfected prescriptive rights and, as such, having acquired water rights superior to those of overlying landowners. Overlying landowners were more motivated than appropriators to launch court action to prevent appropriators from obtaining prescriptive rights. However, also for them, the decision about when to start litigation was highly risky. If the court ruled that the water being diverted by the defendant was surplus water, the plaintiff would pay the costs of litigation and receive no remedy. If, to avoid this situation, an overlying landowner waited too much to go to court, and the court ruled that the water being diverted was non-surplus water, depending on the time he has waited, he might find that the defendant had perfected a prescriptive right. Both for overlying landowners and for appropriators, the uncertainty about the physical world (the actual level of water in the basin) and about the behaviour of other people (the quantity of water actually withdrawn by all agents) fueled the uncertainty about the legal situation (because of the competing water doctrines and the legal consequences of the court recognizing or not the presence of a surplus). These kinds of information, so difficult to obtain, were sometimes only available as a result of litigation, because the court had to appoint a specialist to determine the situation at the basin and so some information got shared by all the producers. Without a change of institutions, pumpers were encouraged to overexploit, eventually leading to the destruction of the resource itself. This was the scenario during the first 50 years of twentieth century, but changed - due to collective action.

The change of institutions started at the Raymond Basin, where the city of Pasadena was by far the largest producer (its production equalled the production of the other 30 producers combined) and assumed for some years the strategy of the dominant player (undertaking independent actions that benefited other producers who were not contributing to the costs). On the context of legal proceedings from Pasadena against other producers, a report asked by the court showed that a level of dangerous overexploitation of the basin had been reached. As a consequence, the need of curtailing the pumping to the safe yield of the basin became common challenge to all parties. Given the legal uncertainty, and the insupportable costs of litigation that would result from a contested trial, serious negotiations resulted in an agreement signed by all but one of the producers. The parties invented the notion of "mutual prescription", meaning that all accepted that each producer's withdrawal of groundwater had been open, continuous, and notorious and was, because of the overdraft, adverse to the claims of all of the others, and, thus, each producer had prescribed against all the others. With this foundation, the parties agreed to share the cutback proportionately, instead of pursuing further legal procedures about water rights. A framework for future selling and buying of water rights was also set. The judge issued a final judgment based on the parties'

agreement and assigned a public division to serve as the official supervisor of the agreement (paid by the parties and by the state).
The overdraft came a decade later to West Basin than to Raymond Basin, thus giving the local producers the opportunity to build on the experience of the Raymond Basin. The West Basin is much larger than the Raymond Basin and had some disadvantages: a larger number of producers (around 500); the absence of a dominant producer; important asymmetries in the perceptions of the risks faced by different classes of producers. During 1943, nine of the coastal municipalities initiated action. A report they ordered came in 1944 and made evident that the salinity threatened the entire basin with destruction. A permanent association of water producers was created, a renowned engineer was recruited to identify alternative sources of water for the basin, and legal action was initiated to find a solution.
After four years of intense study, a report from a referee came as a bombshell: the safe yield of the basin was 30.000 acre-feet per year; by 1952, water withdrawals had reached 90.000 acre-feet per year. The situation as much more difficult than expected, but even supporters of proportionate curtailment opposed a two-thirds reduction in groundwater production, because of the economic effects of such a decision. The water producers' association created a committee of attorneys and engineers to find a reasonable solution. The committee proposed to use the Raymond Basin's concept of mutual prescription; to calculate prescriptive rights using data from 1949, a date at which additional parties had been added to the court case (63.728 acre-feet), instead of using 1944 data, the year immediately before the initiation of litigation (44.387 acre-feet); to implement a cutback of 25% to 30% of prescriptive rights, as an interim agreement that the parties could ratify immediately to achieve an actual cutback within a short time. The interim agreement was drafted as a contingent contract: a water producer, by signing the agreement, promised to curtail production to his own "prescriptive rights, 1949", in the event of holders of at least 80% of the total water rights had signed the agreement and it had been approved by the court. Thus, no one would be a "sucker". In two years, these conditions had been met, the agreement entered into force, and water levels rose immediately and continued to rise for several years.
The interim agreement was used for seven years, while the committee of attorneys and engineers continued its work at an intense pace (at least weekly meetings during most of 1957 and 1958). A final agreement, also as a contingent contract, was presented to the court and became legal for all parties in the case. All nonsignatories - including the city of Hawthorne, that continued to pump all the water it found necessary, saving a lot of money compared to signatories having to import water to compensate their cutbacks - were placed under legal order to reduce its groundwater production to the levels stipulated by the agreement. The city

of Hawthorne appealed the decision, first to the District Court of Appeals, and subsequently to the California Supreme Court, just to see the initial decision confirmed. So, the case closed 18 years after it had opened.
Water producers of the Central Basin, with a less pressing situation because of less deteriorated natural conditions, followed the general strategy of other basins, and got a final approval to its agreement from the court in 1965.
At a point in the process, West Basin and Central Basin water producers recognized that long-term regulation of their problems will not be achieved by the agreements already signed and in preparation. First, agreed cutbacks were insufficient to lead to a stable solution. Second, the danger of saltwater intrusion had not been solved. Third, both basins needed an integrated regulation of water production, because of water flows between basins. Leaders of water producers in both basins joined efforts to prepare new legislation for solving critical groundwater problems. One of the results of that effort was a draft legislation creating a new type of district empowered to undertake broad replenishment responsibilities financed primarily by a "pump tax".
The new Water Replenishment District Act, approved by the state legislature in 1955, authorized citizens located in southern California to create such a new district. Some conditions applied: at least 10% of the registered voters residing within the boundaries of the proposed district had to sign the proposal, specifying the limits of its taxing power; the Department of Water Resources had to agre on the beneficial effects of the new district; a majority of votes, in a special election held to consider the creation of the new district, had to approve the initiative. The new legislation provided a general "constitution" for new water districts: water producers in any specific area could use that general framework to create a particular "constitution" for their own district. In 1959, the Central and West Basin Water Replenishment District was created, from an initiative of water producers, with the approval of the citizens living in the area, with public powers to tax, to sue, and to engage in the provision of collective goods.

5.1.3 Contingent strategies in sequential, incremental, and self-transforming processes

Which lessons can we learn from the above resumed case study? To understand how agents behave in the real world, we must abandon the idea of economic agents designing a perfect mechanism to give, once and for all, a final solution to a complex problem involving strong uncertainty, divergent interests among participants, and the risk of opportunistic behaviour. The fundamental obstacles to such a move are, in one or other way, related to the radical incompleteness of information agents can collect - about the natural world, and about the social and institu-

tional world. **Economic agents facing an undesirable state of the world, and struggling to change it, do not design ex ante mechanisms to govern all ex post eventualities in a desired future state of the world. Real agents adopt contingent strategies as their contribution to sequential, incremental, and self-transforming processes.**
The groundwater pumpers of the case study, in order to avoid getting stuck in a bad for all and unsustainable pumping race, made a substantial investment to change the situation. However, that investment was not a blind bet. The investment was not made in a single step. The process involved many small steps, most of them of low initial costs, most of them conditional on concrete information about what other participants were doing. (It is worth reading the full report from Ostrom, with much more detail than can be given here). The sequential and incremental nature of the process allowed participants to experience benefits of initial steps, and to collect evidence of others doing their share, before moving to larger investments and more permanent commitments. Appropriators from some basins acted after learning what appropriators from neighbouring basins had already achieved and by which means. The voluntary associations established by the participants changed the structure of the game, from an interdependent situation with individuals acting in an independent manner, to an interdependent situation with coordinated action. An essential element of the coordination action was about information: organizing to obtain information about the natural resource, as well as sharing information about mutual interests and dispositions to behave in some ways and not in other ones. Another important aspect of the case study is the relationship between different levels of institutional action: self-organizing efforts of local participants were not replaced, but well supported, by the judicial system, as well as by the state of California's political and administrative structures. The case study illustrates how, not only changing, but also creating new institutions, can be incremental processes, where each institutional change become the foundation for the next change - step by step [Ostrom, 1990, pp.139-141].
The kind of compromises of the groundwater pumpers of the case study are **contingent strategies**, because **the behaviour of anyone depends on other one's behaviours (no one wants to be the sucker**, no one wants to be exploited by nonconformers to the agreed rules). **Self-commitment, being contingent, needs information about other parties' behaviours to be sustainable** - meaning that appropriate monitoring must be part of the institutional arrangement. Monitoring modalities could make a positive contribution to sustainability, namely involving the parties on monitoring activities while keeping monitoring costs low. Direct participation of parties on monitoring activities facilitates adequate sanctioning decisions: graduated sanctions, taking into account the concrete situation (for example, discriminating continued and unreasonable breach of the

rules from occasional, by error or in an emergency deviation) tend to consolidate the system. Contingent strategies are the good ones for fallible individuals in complex and uncertain environments, if they don't want to get stuck in Pareto-inferior situations because of their individual and independent "dominant strategies" [Ostrom, 1990, pp.185-187].
Giving the right place to the consequences of information incompleteness in the real world of real agents in social settings can help to avoid some wrong assumptions about collective action. The assumption of complete information must be replaced by a consideration of how individuals actually obtain information, who has what information, and whether or not information is biased. The assumption of independent action must be replaced by a consideration of how and why individuals can take into account the effects of their actions on the choices made by others. The assumption of perfect symmetry must be replaced by a consideration of the individual diversity that can be found within any sophisticated population of social agents, partially explainable by the different vantage points resulting from the different opportunities at hand for different individuals. The assumption of no human (agent) error in the assessment of other agents' behaviour, eventually followed by crude sanctioning activities, must be replaced by a consideration of the need to discriminate opportunist behaviour from occasional error in implementing rules. The assumption that agents are "rational idiots" must be replaced by a consideration of the existence of social norms, of the fact that they constrain behaviour, some physically possible actions that would maximize benefits for the individual not being at his disposition because of external and internal enforcement mechanisms. Both the assumptions of zero monitoring and zero enforcement costs must be replaced by a consideration of the existence of different monitoring and enforcement mechanisms and of the usually different costs they imply and the different degrees of effectiveness they tend to guarantee. The assumption of fixed structures of action situations must be replaced by an examination of the capacities agents enjoy to change the rules of the situation itself, and also by a consideration of how the surrounding regime (broader institutions) enhances or inhibits local institutional change.

5.1.4 Incomplete contracts: understanding decentralized aspects of economies

Information incompleteness has fundamental consequences at all levels of an understanding of large systems of multiple agents. The incomplete contracts issue shows some of its ramifications into crucial aspects for designing collective sys-

tems[7].
Economists in the Walrasian tradition based their analyses of the functioning of decentralized economies on the notions of market and price system: supply meets demand around a posted price; the same goods and services (with the same properties, fixed in advance) traded at the same price and under the same rules and virtually at the same time (by a kind of contract of null duration); all the market actors participating in the process. Given the unrealistic assumptions of this view of a decentralized system (in practice, agents exchange goods and services outside of equilibrium and in bilateral contexts, subject to transaction costs and informational asymmetries), there is a need for a more realistic theory, a theory of contracts, where "a contract is an agreement under which two parties make reciprocal commitments in terms of their behavior - a bilateral coordination arrangement" [Brousseau and Glachant, 2002, p.3].

From complete to incomplete contracts: three principal currents

To be complete, a contract should design ex-ante a complete set of behavioural rules that will ex-post solve all coordination problems that can possibly arise during implementation time. Several theories endeavour to explain in which sense contracts usually or always remain incomplete, and why. Within contract economics three principal currents can be distinguished according to the theoretical traditions they belong to: Incentive Theory, Incomplete Contract Theory, and Transaction Costs Economics (the New Institutional Transaction Costs theory). Each has a different view on contract incompleteness.

Incentive Theory (IT)

Incentive Theory (IT) draws on several of the traditional hypothesis of Walrasian economics, namely the substantial rationality of economic agents; that they possess complete information concerning the structure of the issues they confront (while not being able to precisely anticipate the future, they do know the structure of the problems that may occur and, thus, envisioning the future on the basis of probabilities); that they are endowed with unlimited computational abilities; and that they have a complete, ordered and stable preference set.
IT diverges from the Walrasian assumptions in accepting that two contracting parties will usually not have access to the same information on certain variables, because one party should not know, ex ante, the private information of the other party (for example, on his preferences, the quality of his resources, his willingness

[7]This paragraph is based mainly on [Brousseau and Fares, 2000] and [Brousseau and Glachant, 2002].

to pay, or his reservation price). This asymmetry of information is the cause for **adverse selection** (where the asymmetric information is **exogenous**, i.e., not subject to manipulation during the exchange by the party possessing it) or **moral hazard** (where asymmetric information is **endogenous**, i.e., vulnerable to such manipulation). Adverse selection can be exemplified by a potential employer's uncertainty concerning a job seeker's level of competence, while moral hazard can be exemplified by the uncertainty on the level of effort the employee will supply.
Putting the canonical situation on the terms of principal-agent theory, we have: the principal is the under-informed party; the agent is the informed party; the principal needs to put into place an incentive scheme to induce the agent to either disclose information (in adverse selection situations) or to adopt behaviour that is in line with the interest of the principal (in moral hazard situations). The incentive scheme consists of remuneration being conditional on signals that result from the agent's behaviour.
This approach relies on two kinds of assumptions. First, the principal knows the probability function of the hidden variables, and knows the agent's preference structure, so being able to calculate the conceivable possible remuneration schemes and anticipating the agent's reaction to them. These assumptions about the nature of the world and the rationality of the agents are seen as problematic for adherents to bounded rationality hypothesis. Second: there is an institutional framework ensuring that the principal will respect his commitments, so making them credible to the agent. This assumption is problematic for those studying the concrete and diverse forms institutions take at different times and places, and the possibility of agents modifying them in some directions [8].

Incomplete Contract Theory (ICT)

Incomplete Contract Theory (ICT) , while also staying close to neoclassical theory, departs from its assumptions by a key hypothesis: complete contracting is impossible when the values of some of the central variables of the future interaction between the contracting parties (such as the level of effort) are not verifiable, ex post, by a third party (a "judge"). The focus on the issues arising from non-verifiability (taken as a failure of the "judge") amounts to an interest on the institutional framework. So, for ICT, contractual incompleteness originates from the bounded rationality of the judge (the entity that is responsible in the last resort for the enforcement of the contract). There is an information flow about the contract implementation, which is observable but not verifiable by a third party (the judge). There are two problematic aspects of this approach. First, it retains contradictory assumptions about the agents' rationality (perfect rationality) and

[8]For a recent exploration of the potential of IT, see [Malin and Martimort, 2002].

the institutional framework / the judge rationality (bounded rationality). Second, focusing only on the types of contract that can be implemented given the features of pre-existing institutions, the institutional framework remains exogenous in the analysis [9].

Transaction Cost Economics (TCE)

With [Williamson, 1975, 1985], New Institutional Economics (NIE) starts a theory of incomplete contracts, on the basis of the concepts of Transaction Cost Economics (TCE)[10]. A fundamental aspect of this approach is that it proposes to endogeneize the forming of institutions and governance structures into the analysis.
From the perspective of NIE, complete contracts cannot be settled. For three main reasons.
First, because of the bounded rationality [Simon, 1976, 1987] of contracting parties. Since "economic agents do not know all the solutions to the problems they face, are unable to calculate the possible outcomes of these solutions, and cannot perfectly arrange these outcomes in order in their space of preferences", they will not be able to write complete rules to deal with every relevant contingency that can arise when implementing a contract.
Second, because of some characteristics of the (economic) world, pointed out at a general level by transaction costs theories, that also militate against the possibility of complete contracts: decisions are time-consuming and costly, agents make mistakes, and strong information asymmetries among them (because their visions of their present and future economic positioning are not shared) are always at stake.
Third, because of radical uncertainty - in the sense of [Knight, 1921] and [ODriscoll and Rizo, 1985]. While with risk and Bayesian uncertainty the possible characteristics of the future are known by the agents, they being only uncertain about what will actually happen (and this being formalized by a probability function), with radical uncertainty agents do not know the possible characteristics of the future states of the world.
So, for NIE, contractual incompleteness originates from the bounded rationality of each individual involved in the economic system, and from uncertainty. All the participating parties involved in a contractual process are assumed to be bounded rational. Each coordination mechanism, designed and run by agents whose rationality is bounded, must be imperfect. Institutions are the realistic response to the coordination problems so arising: complementary of various coordination devices (like contracts, organizations, and institutions). Agents have to build a diversity

[9] For pioneer work on ICT, see [Grossman and Hart, 1986] [Hart and Moore, 1988].
[10] For developments, see also [Williamson, 1996].

of complementary governance mechanisms to reduce transaction costs and imperfection. Thanks to this incorporation of institutional aspects into the analyses, it is said that the institutional framework is endogeneized.
So, agents are only able to do incomplete contracts: some future states of the world are predicted and a set of rules of mutual behaviour in those cases is designed ex-ante. For unpredicted states of the world, or for situations where ex-ante designed rules are recognized ex-post as inefficient by the parties, *several provisions can be made ex-ante*. For example, *a decision-making device can be set to produce ex-post rules of behaviour* for all parties, all parties agreeing to following them. The decision-making device can be of different kinds, or a combination of them: authority within a hierarchy, where the decision-making device is one of the contracting parties; a negotiation structure; or a third party (a court, for example; or a extra-judicial arbitrator).
Further, *enforcement mechanisms* are needed to give credibility to contractual commitments. The recourse to the judicial system is a possibility, but, exactly because of the incompleteness of the contracts, it is problematic. Incomplete contracts always contain some level of vague commitments, involving the recourse to informal behaviour (verbal instructions, for example), and rather vague behavioural principles (like requiring a cooperative disposition from the parties), which assessment is difficult, also taking into account the problem of access to both the relevant information and the relevant knowledge. All this makes room for contractors to adopt sophisticated forms of opportunism. This kind of problems can be faced by the way of *self-enforcing mechanisms*: incentive and coercion schemes (to incite some behaviours and dissuade others[11]), supervision devices (to verify how parties comply), and arbitration mechanisms (to resolve conflict)[12].

Institutions and private orders

Transaction Costs Theory, replacing substantial rationality by bounded rationality, and risk by radical uncertainty, as key assumptions on the nature and situation of economic agents, renders complete contracts an impossible thing in real world. In addition, institutions that are ultimately responsible for ensuring the performance of contracts ("judges") cannot enforce clauses relating to unverifiable variables, take a long time to decide, make mistakes. So, parties cannot

[11]However, NIE views on incentives depart from Incentive Theory approach. Incentive Theory sees incentive mechanisms as based on marginal remuneration. Agents with bounded rationality in a complex environment are not able to accurately calculate marginal productivities. NIE sees incentives as based on the sharing of the outcome of an efficient co-operation and the logic of deterrent.

[12]Note that divergent assessment of the same situation can arise between partners that are all of good faith (involuntary opportunistic behaviour can be the result of radical uncertainty).

rely on external mechanisms as strong guarantees. So, contracting parties must create a "private order" to ensure cooperation ex post. Public components of the institutional framework must be combined with formal collective "self-regulatory" mechanisms (such as professional codes of conduct enforced by professional associations), and with informal analogs (such as behavioural rules imposed by relational networks based on different kinds of social groups), to make the best of the potential complementary of all these components of the social and economic realm.
Agents can set up "interindividual governance structures" in order to compensate the incompleteness of the ex-ante contractual obligations and ensure self-enforcement. "The collective governance that is exercised by Institutions is incomplete and imperfect" [North, 1990], namely because, in most cases, institutions are not intentionally or specifically designed to govern economic interactions, and they are not tuned to fit a specific kind of transaction (they are at best shaped to the lowest common denominator of a set of transaction).
A very positive aspect of this dynamics: the incompleteness of the institutional framework gives the agents some freedom to shape its evolution, not only directly designing and implementing new collective governance structures, but also for example creating "interindividual governance structures" that will impact the already existing institutional mechanisms (for example, revealing their weaknesses and making some of them obsolete).
The contractual approach is at the root of a renewed analysis of the functioning of a decentralized economy. The problem of incomplete contracts led to the abandonment of the idea of economic agents designing ex ante mechanisms to govern all eventualities in any future state of the world. The NIE view on this problem underlines the role played by institutions in a world of uncertainty. An essential element of the interplay in contractual relationships comes from their institutional environment. Institutions frame decentralized contractual relationships: the institutional environment provides the rules of the game, so determining the modalities and the conditions of the contracts' efficiency. In this sense, property-rights regimes are of particular interest to study the linkage between institutional environments and decentralized contractual relationships.

5.1.5 Incomplete institutions and the invisible hand

As a joint result of bounded rationality, bounded autonomy, and the imperfections found in the real world, all together having its consequences mutually reinforced by information completeness, we live in a world of institutional incompleteness. First of all, for all those reasons, formal institutions are inevitably incomplete. The gaps in formal institutions are covered, at some extent, by informal rules - but these cannot be fully enforced by explicit means. "I do not think it is possible to

elucidate necessary and sufficient principles for enduring institutions, as it takes a fundamental willingness of the individuals involved to make any institution work. No set of logical conditions is sufficient to ensure that all sets of individuals will be willing and able to make an institution characterized by such conditions work." [Ostrom, 1990, note 36 to page 91]
Institutional incompleteness is also the result of contingency in real life. In a world of incomplete institutions, formal enforcement also must have some limits. Formal (or legal) enforcement is supplemented by informal (extralegal) guarantee instruments (from hostages to reputation) against "bad" behaviour - so creating a private order to regulate future conflict, be it a consequence of opportunism or of honest disagreement among parties. Since all the contingencies cannot be anticipated ex ante, a rational institutional designer does not try to regulate everything to the last detail. Leaving gaps in institutional design is wisdom. Some of these gaps will be filled in by jurisprudence and legal practice, others by social norms - and others will be neither immediately nor explicitly filled in, creating an informal area. In this informal sphere the "invisible hand" works, slowly but powerfully - in some cases, stabilizing, in other cases destabilizing a system (in a whole country or in a particular organization). The informal rules are taken for some authors as the space for customs, routines, or habits [Furubotn and Richter, 1997, pp.15-20].
A specific development of the incompleteness topic is "transaction costs". Because of its importance, both in economics and for our current purposes, it will be addressed in the next section.

5.2 Transaction Costs Economics

One prominent consequence of recognizing that information is fundamentally incomplete within the economic world is the need to systematically take into account transaction costs. The following quote from Furubotn and Richter [1997] illustrates the link between what has been previously said about incompleteness and the transaction costs issue:

> "In the neoclassical world of costless transactions and perfect foresight, such institutions (constitutions, laws, individual contracts, and so on), are complete and perfect. Their provisions, which are perfectly enforceable by law, will be observed with absolute precision. Courts work without cost in resources and time. Moreover, in this special environment, it is known in advance how courts will decide in the event of litigation. Lawsuits could be carried out by computers because of perfect information and perfect law and contracts. Strictly speaking,

there will be no need for lawsuits because decision makers understand the conditions of the system and act with perfect rationality. This, then, is the neoclassical vision. It is the ideal world of the public administrator who dreams of perfect social engineering."
[Furubotn and Richter, 1997, p.15]

5.2.1 Why transaction costs make a difference to orthodoxy in economics

According to Ronald H. Coase, 1991 Nobel Prize in Economic Sciences, the concentration of orthodox economics on the issue of determination of prices has led to an exclusive interest only in what happens on the market (the purchase of factors of production, and the sale of the goods produced), and to the neglect of other aspects of the economic system. The internal arrangements of organizations, which are a large share of the actual workings of economy, has been largely ignored. This neglect was still favoured by the growing abstraction of the analysis, the firm being described as a "black box".
Douglass North, 1993 Nobel Prize in Economic Sciences, gave the same explanation and we can consider it here in some more detail: "Consider first the standard neoclassical Walrasian model. In this general equilibrium model, commodities are identical, the market is concentrated at a single point in space, and the exchange is instantaneous. Moreover, individuals are fully informed about the exchange commodity and the terms of trade are known to both parties. As a result, no effort is required to effect exchange other than to dispense with the appropriate amount of cash. Prices, then, become a sufficient allocative device to achieve highest value uses." [North, 1990, p.30]
North further suggests how to remediate such an unrealistic conception of the economic world: "To the Walrasian model (...) I now add costs of information. (...) These include the costs of measuring the valued attributes of goods and services and the varying characteristics of the performance of agents. The net gains from exchange are the gross gains (...) minus the costs of measuring and policing the agreement and minus the losses that result because monitoring is not perfect. On a common sense level, it is easy to see that we devote substantial resources and efforts to the measurement, enforcement, and the policing of agreements." [North, 1990, pp.30-31]
The costs North is integrating into the economic analysis are transaction costs. They are needed to understand the history of economic activity: " (...) one cannot take enforcement for granted. It is (and always has been) the critical obstacle to increasing specialization and division of labor." [North, 1990, p.33] They are also needed to understand omnipresent features of the economic world: " (...) the agency issue is ubiquitous in hierarchical organizations." [North, 1990, p.32] Considering these costs of running the economy makes a big difference in comparison

to more abstract approaches: "If we return to the Walrasian model (...), we assume that there are no costs associated with enforcement of agreements. Indeed, as long as we maintain the fiction of a unidimensional good transacted instantaneously, the problem of policing and enforcement are trivial.(...) It is because we do not know the attributes of a good or service or all the characteristics of the performance of the agents and because we have to devote costly resources to try to measure and monitor them that enforcement issues do arise." [North, 1990, p.32]

5.2.2 What are 'transactions' and 'transaction costs'. Taxonomy

A first approach to the meaning of the term "transaction" is given by Williamson: "A transaction occurs when a good or service is transferred across a technologically separable interface. One stage of activity terminates and another begins." (cited by [Furubotn and Richter, 1997, p.41])
However, restricting the definition to situations in which resources are transferred in the physical sense of "delivery" is no longer generally accepted. Another definition is given by Commons (cited by [Furubotn and Richter, 1997, pp.41-42]): transactions "are the alienation and acquisition between individuals of the rights of future ownership of physical things". The definition now opens to the transfer of resources in the legal sense (transfer of property rights).
Another extension of the notion is a result of taking into account, not only material things or legal rights, but, at a more general level, information. Information can impact economic activities while not involving directly the transfer of physical products. The quote from North (above) is an example of that move.
One further step is to consider, not only economic transactions, but other social actions as well. This way, actions necessary to establish, maintain or change social relationships, can be considered as transactions - economic transactions being a subset of social transactions. Institutional approaches to economics underline the fact that economic activity of human beings takes place in a society with some kind of institutional framework, which can be better understood taking into account different kinds of transactions. Political transactions, for example, are a specific part of these social transactions.
The taxonomy offered by [Furubotn and Richter, 1997, pp.43-48] encompasses a broad range of approaches to transaction costs, from the more narrowly focused on economic activity to those more inclined to strengthen the links between economics and other social and political sciences. It, thus, seems helpful to illuminate an exploration of the concept.
They mention three main categories of transaction costs, from the market to the political operation of the society where economic activity takes place.

(I) Market Transaction Costs

(1) Search and information costs

(a) individuals contemplating particular market transactions must search for suitable parties with whom to deal (by advertising, visiting prospective customers, creating and organizing fairs, weekly markets, stock exchanges, and so on)

(b) there is a need of communication among prospective parties to the exchange (postage, telephone, sales representatives,...)

(c) costs related to the gathering of information about the same good at different prices from different suppliers;

(d) testing and quality control (and credentials of suppliers, in the case of services)

(2) Bargaining and decision costs, related to the writing of contracts, bargaining and negotiating its provisions (including the difficulties arising from informational asymmetry, where bargaining parties possess private information).

(3) Supervision and enforcement costs, arising because of the need to monitor the agreed upon delivery times, measure quality and amounts, and all matters related to protecting rights and enforcing contractual provision (taking into account that, at some extent, violations of contracts are unavoidable).

(II) Managerial Transaction Costs

(1) The costs of setting up, maintaining or changing an organization (personnel management, investment in information technology, defense against takeovers, public relations, lobbying);

(2) The costs of running an organization, including the costs of decision making, monitoring the execution of orders, measuring the performance of workers, agency costs (costs that arise in a principal/agent relationship[13]), costs of information management, and so on.

(3) The costs associated with the physical transfer of goods and services across a separable interface.

(III) Political Transaction Costs

(for a capitalist market to exist, some kind of institutional arrangements must be in place - and the provision of such a framework involves

[13]Agency costs are costs arising in a principal/agent relationship. There is an agency relationship between two (or more) parties where one party, the agent, act for, on behalf of, or as representative of a second part, the principal. For several possible reasons (some reducible to opportunism), the agent not always act fully in the interest of the principal. From the principal side, it is always costly to implement measures to limit those divergences.

costs)
(1) The costs of setting up, maintaining, and changing a system's formal and informal political organization (legal framework; administrative structure; military, educational and judiciary systems; political parties).
(2) The costs of running a polity: "duties of the sovereign"; the costs of measuring, monitoring, creating, and enforcing compliance; the costs of running organizations designed to participate in the political decision-making process (political parties, labour unions, employers' associations); the costs of "the domestication of force" (any kind of force is assumed to be perfectly under control in the neoclassical view of the environment, without any possible disturbance to transfer of property rights).

5.2.3 From a 'Science of Choice' to a 'Science of Contract'. From the firm as a function to the firm as a governance structure

Oliver Williamson (2009 Nobel Prize in Economics for "his analysis of economic governance, especially the boundaries of the firm"), in a text about the main issues of interest for Transaction Cost Economics [Williamson, 2005], defines orthodoxy in Economics, as developed throughout the 20th century, as "a science of choice". This "science of choice" has two sides: the theory of consumer behaviour (consumers maximize utility), and the theory of the firm (firms maximize profit). The dominant paradigm for Economics focuses on how quantities are influenced by changes in relative prices and available resources. As Lionel Robbins wrote in 1932: "Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses". So, within this paradigm, neoclassical economics is predominantly concerned with price and output, describing the firm as a production function (which is a technological construction).
Transaction Cost Economics (TCE), while recognizing the role of markets, emphasises the allocation of economic activity across alternative modes of organization and describes the firm as a governance structure (which is an organizational construction). The orthodox way of conceiving the firm as a production function amounts at seeing it as just a technological construction, because, that way, the transformation of inputs (of land, labour, and capital) into outputs (of goods and services) is seen only as a function of the technology employed. Conceiving the firm also as a governance structure, as an organizational construction, the way TCE does, is to take into account the influence of institutional aspects of that transformation. TCE accepts that market competition also serves governance purposes,

in the context of simple market exchanges, with large number of parties - but is predominantly concerned with complex market exchanges, with small number of parties on each side of the transaction, in contexts of incomplete contracting. And, methodologically, TCE does not accept to confine the analysis to the price-theoretic apparatus, focusing instead on strategic hazards and the cost of deploying governance schemes to mitigate these hazards. Focusing on many kinds of transactions, on diverse exchange contexts, TCE endeavours to bring out their latent contractual features. In this sense, TCE is (part of) a "science of contract".

A conceptual map

Figure 1 in [Williamson, 2005] maps the main distinctions relevant to understand TCE's positioning as contrasted to other approaches.

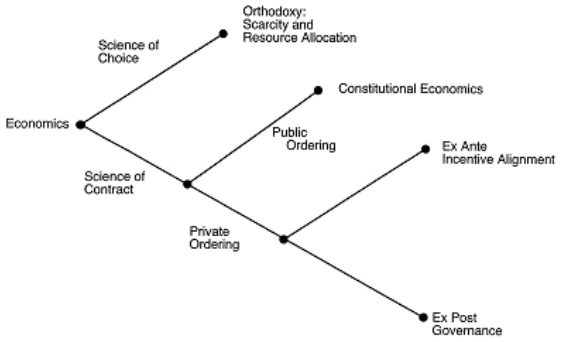


Figure 5: in Williamson, 2005

As above mentioned, TCE is part of a Science of Contract, while neoclassic economics is part of a Science of Choice.
Constitutional Economics deals with "Public Ordering" issues, related to politics as a structure of complex exchanges among individuals, within which people try to secure by collective means their own private objectives that cannot be secured only by simple market exchanges. "Private Ordering" is about efforts deployed by the immediate parties to a transaction to craft governance structures attuned to their needs.
Within Private Ordering, while the Ex Ante Incentive Alignment branch in mainly concerned with mechanism design, agency theory, formal property rights, TCE focus basically on the Ex Post Governance of contract relations (contract implementation, where maladaptation problems appear). The focus on Ex Ante aspects of the governance of a contractual relation tends to see situations as if formal incentive alignment could possibly annihilate the need of Ex Post Governance - a

perspective that is not at all shared by TCE.

5.2.4 Some fundamental assumptions of TCE

About human actors.
The key attribute of human actors (namely as economic agents) is bounded rationality (as Herbert Simon has already put it in 1957: behaviour that is "intendedly rational but only limitedly so"). For TCE, "the chief lesson of bounded rationality for the study of contract [is] that all complex contracts are unavoidable incomplete" (46). But bounded rationality does not imply that human actors are myopic: they "look ahead, uncover possible contractual hazards, and work out the contractual ramifications [Williamson, 2005, p.46].

About the unit of analysis.
The unity of analysis for lens of contract purposes is the transaction. Three dimensions are important to analyse transactions [Williamson, 2005, p.47]:
(i) asset specificity;
(ii) the disturbances to which transactions are subject (and to which potential maladaptations accrue);
(iii) the frequency with which transaction recur (which bears on reputation effects and on the incentive to incur the cost of specialized internal governance).

About the main purpose of Economics' analysis.
The central problem of Economics can be seen as the problem of *how economic actors adapt* to changes in the market, both *individual parties' adaptation* through response to prices, and *cooperative adaptation* through organization (administration, namely within firms). To better understand adaptation, TCE describes the firm not as a production function (which is a technological function) but as a governance structure (which is an organizational construction). And the market, also, is described as a governance structure. In a world where complex contracts are incomplete, and where its implementation faces disturbances, without any possible full perfect anticipation of all contingencies, the only way actors have to face hazards and restore efficiency is to craft governance structures able to dissipate threatening impasses. [Williamson, 2005, p.48]

About governance structures.
Examining economic organization through the lens of contracts places the spotlight on ex post adaptation. The three attributes of principal importance for describing governance structures are
(1) incentive intensity,

(2) administrative controls, and
(3) contract law regime.
Incentive intensity is lower within firms than on markets. Administrative controls are much more important within firms than on the market. Disputes in markets are subject to law and courts, whereas most internal disputes of firms cannot be heard by courts. Actor have to analyse tradeoffs like this one: taking a transaction out of the market, and organizing it internally in the firm, involves weakening incentive intensity and adding administrative controls. [Williamson, 2005, pp.48-51]

5.2.5 A paradigm of TCE analysis: 'vertical integration'

By vertical integration we refer to the kind of situation where a firm, being in need of a given product, for example as a component of its own production, has to decide whether to buy the product from another firm that is selling it, or to take the appropriate provisions to become able to make internally the same product. So, vertical integration is a *make-or-buy decision*, a decision with effects on specialization, and it is interesting because it is about going to the market or, alternatively, using an organization within which administrative restrictions apply. Reasons to sometimes prefer internal supply relate to transaction costs of going to the market.
According to [Williamson, 2005], vertical integration is a paradigm: it embodies the main issues TCE finds relevant to analyse contracting. To make a decision on how to fulfil any need of its economic activity, a firm has to decide either to own the means of doing it or to contract with an adjacent stage: backward into raw materials, laterally with components, forward into distribution. For some activities, ownership is impossible (for example, firms cannot own workers or final costumers) or rare (firms usually don't own their suppliers of finance).

5.2.6 Asset specificity

One dimension that is considered important to analyse any transaction (see above) is "asset specificity". We can now say something more on that concept, while giving an example of how TCE makes predictions (that can be subject to empirical test).
Assume a firm can make or buy a component, and assume further that the component can be supplied by either a general purpose technology or a special purpose technology. Let k be a measure of asset specificity. For transaction of a general purpose technology, $k = 0$. Transactions involving a special purpose technology have $k > 0$. This gives an informal notion of asset specificity.

Figure 2 from [Williamson, 2005] puts to a use the asset specificity concept to compare spot markets[14] (M), organizations, or hierarchies (H), and hybrid modes of contracting with some kind of credible commitments mechanism (X). The transaction cost consequences of organizing transactions in Markets (M) or hierarchies (H) are shown: the bureaucratic burdens of hierarchy place it at a disadvantage with $k = 0$, but the difference narrows as asset specificity builds up and eventually reverse with large asset specificity ($k >> 0$), where the need for cooperative adaptation becomes especially great. Hybrid modes of organization (X), with good combinations of incentive intensity and administrative control, can take intermediate values.

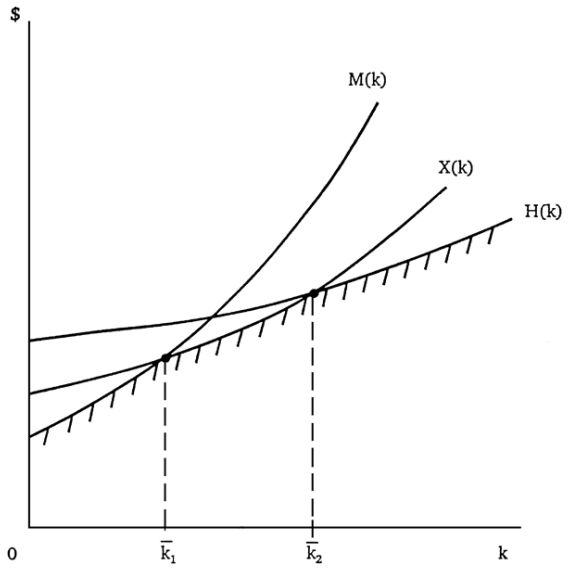


Figure 6: Transaction costs and asset specificity, in (Williamson, 2005)

Let us give just another example of asset specificity and its importance for TCE analysis. Cases where labour is easily redeployed to other uses without loss of production (even if such labour is highly skilled) are cases of low asset specificity with respect to labour. Where workers acquire firm-specific skills, and so will incur in loss of value in case of a premature termination, are cases of higher asset specificity with respect to labour. Different levels of asset specificity can ask for different governance structures. Staying within the same example: "Because continuity has value to both firm and worker, governance features that deter termination (severance pay) and quits (nonvested benefits) and which address and

[14]Spot market: a market where contracts are immediately effective, goods being delivered immediately.

settle disputes in an orderly way (grievance systems) to which the parties ascribe confidence have a lot to recommend them. These can, but need not, take the form of 'unions'." [Williamson, 2005, p.55]

5.2.7 Calculating transaction costs

Some authors consider transaction costs as, essentially, the costs of specialization and division of labour. Even if that vision can be seen as somewhat biased, more appropriate to interpret transactions closer to the market and not so to political transaction, it can incentive a look at the tradeoffs involved in the choice of some organization modes, instead of available alternatives. For example, one can try to calculate how a specific management structure can influence the agents' behaviour, because of the level of transaction costs depending heavily on the behaviour of individuals: " Monitoring and enforcement costs, in particular, will tend to be low if mutual trust predominates in the society." For example, "under favorable conditions, property rights will be respected, and comparatively, uniform ideas will exist about the nature of fair solutions to conflicts." [Furubotn and Richter, 1997, p. 49]
Several authors have already calculated, for specific countries and times, estimates of the transaction cost for an economy as a whole. For example, Wallis and North estimated that the transaction costs for the American economy represented, in 1970, from about 46 percent to about 54 percent of GNP. In a historical perspective, in a century, from 1870 to 1970, the transaction cost percentage more than doubled [Furubotn and Richter, 1997, pp.51-52]. In order to consider a future exploration of the possibility of using transaction costs to compare different organization modes for a collective, it would be useful to give some examples of how economists measure them (For these, and more, and more complex, examples, see [Furubotn and Richter, 1997, pp.49-53]).
Example 1. For a given kind of product in the retail market, prices vary, in some cases importantly, for similar products or even for the same product. A consumer seeking to purchase a piece of that kind of product has to devote time and effort to secure information about the product. However, many consumers avoid expending time and effort on the exercise. We can say that the price differences observed relative to an average price can be interpreted as measures of the costs of the consumer's transaction activities.
Example 2. For the purchase of expensive objects like houses, consumers tend to hire advisers (realtors, lawyers, financial consultants). The fees paid to them can be measured as transaction costs.
Example 3. From the suppliers' side, transaction costs of selling consist of, namely, the transporting and marketing activities.

6 Coordination Artefacts plus Models of the World within Institutional Environments

In this chapter we propose a relatively simple concept to guide the creation of institutions for embodied artificial agents (robots). We will endorse the suggestion that institutions are coordination artefacts, one among a variety of possible types of coordination artefacts. However, having recognized this fundamental fact about the social world that is its incompleteness (incomplete information, incomplete contracts, incomplete Institutions), this must have consequences for our understanding of institutions.
The main consequence is to admit that it will be impossible to design sophisticated systems of multiple robots using only mechanisms for direct interaction. We need indirect mediated interaction mechanisms. And we need a mechanism to guide us designing agents able to behave properly within institutional environments heavily relying on mediated interaction.
Section 6.1. will introduce the suggestion that institutions are coordination artefacts of a specific kind. Section 6.2. argues the need of providing models of the world to agents to let them be able to respond to uncertain and complex environments.

6.1 Institutions as Coordination Artefacts

One important aspect of understanding institutional environments consists in acknowledging institutional diversity. The need to respond to so many different contingencies in so many different action situations lead agents to multiply and diversify institutional arrangements. Any agent in a sophisticated social world faces that diversity all the time: "The same individuals who energetically pursue profit-maximizing strategies from 8 a.m. to 5 p.m. every workday [acting as a rational egoist] may also volunteer several evenings a month on neighborhood projects, contribute substantial funds to diverse charities, regularly vote, and be known to friends and coworkers as kind, considerate individuals who always do more than their share of any team project." [Ostrom, 2005, p.118] We can find other kinds of institutional diversity if we look at different cultures: "We know that when we are shopping in a supermarket that we can take a huge variety of goods off the shelf and put them in a pushcart. Before we put these same goods in our car, however, we need to line up at a counter and arrange to pay for them using cash or a credit card (something else that was not so widely available a few years ago). When we are shopping in an open bazaar in Asia or Africa, however, the do's and don'ts differ. If we go at the end of the market day, we may bargain

over the price of the fruit that is left on the stand-something we could never do in a supermarket where fruit will be refrigerated overnight. If we are in the household goods section of the bazaar, vendors would be astounded if we did not make several counteroffers before we purchased an item. Try that in a furniture store in a commercial district of a Western country, and you would find yourself politely (or not so politely) told to leave the establishment. Thus, there are many subtle (and not so subtle) changes from one situation to another even though many variables are the same." [Ostrom, 2005, pp.4-5]

The modelling of artificial societies reflects in some way this diversity by experimenting with different kinds of artificial institutional devices. A few examples are: norms [Hexmoor et al., 2006], trust and reputation [Sabater and Sierra, 2005, Hahn et al., 2007]; individual rights combined with argumentation mechanisms [Alonso, 2004]. Facing such a variety, how would we choose the most promising concept? Perhaps we need them all. "It does not seem possible to devise a coordination strategy that always works well under all circumstances; if such a strategy existed, our human societies could adopt it and replace the myriad coordination constructs we employ, like corporations, governments, markets, teams, committees, professional societies, mailing groups, etc." [Durfee, 2004, p.14] So, we keep them all, and more - but we need a unifying concept to give the whole some consistence.

"Environment" is such a concept. [Weyns et al., 2005a] suggests the need to go deeper than the subjective view of the environment (in wide use within MAS), where it is somehow just the sum of some data structures within agents. What we need to take into account is the active character of the environment: some of its processes can change its own state independently of the activity of any agent (a rolling ball that moves on); multiple agents acting in parallel can have effects any agent will find difficult to monitor (a river can be poisoned by a thousand people depositing a small portion of a toxic substance in the water, even if each individual portion is itself innocuous). Because there are lots of things in the world that are not inside the minds of the agents, an objective view of environment must deal with the system from an external point of view of the agents [Weyns et al., 2005b, p.128].

One can wonder if this can be relevant to robotics, where agents already behave sensing and acting in real (not just software) environments. We suggest the answer is affirmative. Dynamic environmental processes independent of agents' purposes and almost unpredictable aggregate effects of multiple simultaneous actions are not phenomena restricted to physical environments. Similar phenomena can occur in organizational environments: if nine out of ten of the clients of a bank decide to draw all their money at the same date, bankruptcy could be the unintended effect. And, most of the time, social environments in robotics are poorly modelled. So,

the objective view of the environment could apply not only to physical features, but also to the social environment of the agents. We further suggest that both physical and social environments are populated with strange artefacts: artefacts with material and mental aspects. Let us see, following [Tummolini and Castelfranchi, 2006].

An artefact is the result of some action, something done by an agent to be used by another (or the same) agent. An artefact may not be an object: footprints left on a mined field for the followers are artefacts. An artefact may not be designed explicitly by anyone. The unintended effect of some people walking in the grass to cross the park could be, after some time and more people following the traces, a path recognised by most people as the "official" one. At some point, it becomes an artefact. If an artefact is shaped for coordinating the agents' actions, it is a coordination artefact [Tummolini and Castelfranchi, 2006, pp.318-319]. Institutions are a special kind of coordination artefacts.

Tummolini and Castelfranchi put institutions in perspective against other artefacts. Consider first single-agent actions. Single-agent actions can be coordinated actions if they contribute to solve an interference problem with other agents. Some artefacts have physical characteristics that represent opportunities (they enable or facilitate the execution of some action) and constraints (they create obstacles or impediments to the execution of some action). The physical opportunities and constraints of some artefacts are sufficient conditions to enable a single-agent coordinated action, even if the agent doesn't recognize them (the wall of a house keeps people inside and outside separated).

Sometimes, the agent must additionally recognize the opportunities and constraints of the artefact. Cutting a piece of meat in slices with some sort of knife is such a case (the agent must know how to use the knife), but it is not a coordinated action. Coordinated actions by a single agent can also be enabled by physical opportunities and constraints of an artefact and their recognition by the agent: sitting at a table with other people needs some knowledge (not try to seat at a place already occupied).

To consider more interesting artefacts isn't enough to focus on physical opportunities and constraints. Some artefacts are associated in the agent's mind with cognitive opportunities and constraints (deontic mediators, such as permissions and obligations). In such cases, to enable a single-agent coordinated action, both physical and cognitive opportunities and constraints must be recognized by the agent. A driver approaching a roundabout is obliged, only by physical properties of the artefact, to slow down and go right or left to proceed on. However, appropriate regulation of the traffic needs something more. Traffic regulations indicate which direction all drivers have to choose not to crash with others (for example, in Mozambique, drivers must by default go left at roundabouts).

Furthermore, artefacts can be completely dematerialized. Such artefacts enable single-agent coordinated actions only by means of cognitive opportunities and constraints recognized by the acting agent. Social conventions and norms are relevant examples of the kind. A traffic convention to drive on the right works independently of any material device. It's interesting to note that conventions work that way despite their arbitrariness. A traffic convention to drive on the left is an equally good artefact to achieve the same goal. Both, when in place, make us expect a specific behaviour from others (and from ourselves) in such and such situations. The recognition of the cognitive characteristics of the norms and conventions are not only sufficient, but also necessary conditions to enable the agent action.
Consider now multi-agent coordinated actions. "There exist some artefacts 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" [Tummolini and Castelfranchi, 2006, p.320]. Institutions belong to this kind of artefacts.

"**different ways in which artifacts support the agents in achieving coordination**" [Tummolini and Castelfranchi, 2006, pp.319-320]

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.

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

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.

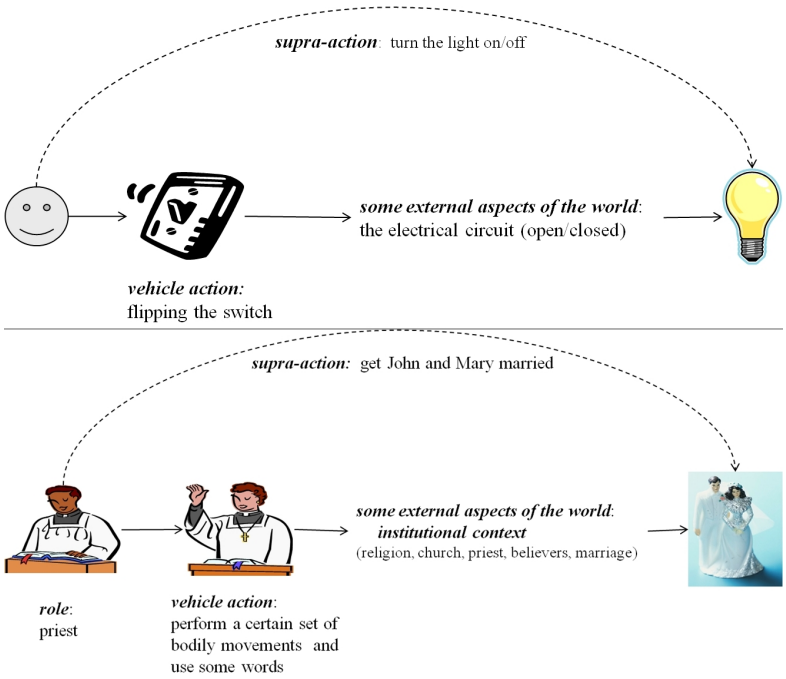
Proposition 4. There exist some artifacts such that their physical opportunities and constraints and the recognition of their use by an agent and the set of 'cognitive opportunities and constraints' (deontic mediators) are necessary and sufficient conditions to enable a single-agent coordinated action.

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

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. These are institutions.

The definition takes institutional actions as multi-agent coordinated actions per-

formed by a single-agent. How this could be? Because of a cognitive mediation intertwined with the agents' behaviours. While traditional views on institutions take them as structured sets of rules and conventions, for Tummolini and Castelfranchi the basic coordination artefact at work is the institutional role played by an agent with the permission of others. The involved participants in an institutional action believe that an agent (Paul) plays a role (priest) and so he has the artificial power of doing a multi-agent coordinated action (the marriage of John and Mary). The participants' recognition of Paul as a priest leads to the belief that he has the power of marrying John and Mary. And both recognition and belief are intertwined with the behaviour of treating Paul as a priest - and treating John and Mary, from some point in time on, as a married couple.



The single-agent action of an agent playing a role is the vehicle action for a collective action. This feature of institutional action parallels some features of physical action. If an agent intends to turn on the light in the room, he must flip the switch. Flipping the switch is the vehicle action for the supra-action of turning the light on. In this context, the agent relies on some external aspects of the world (the functioning of the electrical circuit). When a collective of agents intend to get John and Mary married, the priest must perform a certain set of bodily movements, counting as marrying. That set of movements is the vehicle action for the supra-action of marrying John and Mary. The collective of agents rely on some external aspects of the world - namely, the institutional context that

makes a person a priest with some powers [Tummolini and Castelfranchi, 2006, pp.320-321].
The Tummolini and Castelfranchi's proposal seems to get institutions based only on direct interaction, when it suggests that institutional actions are multi-agent coordinated actions performed by a single-agent. But this is a deceptive appearance. The parallel with flipping the switch to turn the light on clarifies the error: the electrical circuit is not in place by chance, nor by the natural order of the world. It is in place because other agents previously prepared the environment; they were able to so prepare the environment because they had been learned about how electrical circuits work; it had been possible to teach them about electrical circuits because others have previously discovered how electricity works; and so on and so on; and I can turn the light on because I was taught to do so. This is just another instance of the illusion of direct interaction.
Taking the given examples as single-agent actions is somehow misleading, because they actually are examples of historical accumulation of regulation and practice, involving many agents and their continuing interaction. But this is just an example of usual blindness towards mediated interaction, something that many writers seem to consider a hallmark of scientific style of reasoning[15]. What we need, now, is some way to model how mediated interaction combines with direct interaction to make it work in such a complex world. This is why we now turn to the models of the world issue.

6.2 Representation, Mental Models, and Ideologies

6.2.1 Institutional Environments are about Mediated Interaction

Within institutional environments, interactions among participants are directly influenced by operational rules, some of them with a purely local character. Notwithstanding, and given the multilevel nature of the institutional realm (see 4.1.2.), operational situations are affected by higher level rules: institutional-choice rules. Institutional-choice rules are collective-choice rules (determining who is eligible to be a participant and which procedures are to be used to change operational rules), and constitutional-choice rules (determining how, who, and within which limits can collective-choice rules be changed). Institutional-choice rules, being changed at different space and time scales, and possibly by partially different sets of agents, impact operational situations in a indirect manner. This is why institutional settings have, beyond direct interaction, several (possibly many) levels of indirect interaction. Since self-organizing capabilities are an important vector of agents' competence to behave in institutional environments (see 4.1.3.), agents that are unable to understand and to act at institutional-choice situations have limited

[15]We can have mediated interaction both with and without representation; see 6.2.5. below.

capabilities to pursue their own goals in complex social settings. So, socially intelligent agents need capabilities to engage in indirect interaction.
Indirect or mediated interaction is characterized by properties such as name uncoupling (interacting entities do not have to know each other explicitly), **space uncoupling** (interacting entities do not have to be at the same place), **and time uncoupling** (interacting entities do not have to co-exist at the same time). Communication is an example of such an indirect (not local) interaction. "Especially in open, highly dynamic, distributed systems, these properties enable flexible and robust interaction among the cooperating entities." [Weyns et al., 2005a, p.14]
One crucial point (already mentioned, see 3.2.2.) of the institutional approach suggested by John Searle, and confirmed by Institutional Economics, is that institutions allow direct and immediate interaction being replaced by indirect and mediated interaction of a much more sophisticated kind. With the deontic apparatus associated, for example, to property or marriage, we no more have to rely on purely direct interaction with things or other people in order to sustain social arrangements, and we can maintain them in the absence of the original physical setup. People can remain married even though marriage is originally about cohabitation and they now have not lived with each other for years. People can own property even though property is originally about physical possession and now the property is a long way away from them.
Money is a classical example of the power of institutions in providing the means for mediated interaction: "Adam Smith pointed out the hindrances to commerce that would arise in an economic system in which there was a division of labor but in which all exchange had to take the form of barter. No one would be able to buy anything unless he possessed something that the producer wanted. This difficulty, he explained, could be overcome by the use of money. A person wishing to buy something in a barter system has to find someone who has this product for sale but who also wants some of the goods possessed by the potential buyer. Similarly, a person wishing to sell something has to find someone who both wants what he has to offer and also possesses something that the potential seller wants. Exchange in a barter system requires what W. S. Jevons called 'this double coincidence'." [Coase, 2002, p.35]
The study of property is an interesting example of how Institutional Economics has contributed to unveil deep levels of mediated interaction entrenched in institutional settings. Property, once seen as the mere owning of a physical thing, can be better understood as a social relationship shaped by the institutional environment of duties and rights. A Ronald Coase's 1960 paper is a milestone of these changing perspectives. His "The Problem of Social Cost" endeavors to examine the problem of externalities: how to deal with the (sometimes non fully computable)

effects of some actions of an economic agent on others (in general, the concern is with harmful effects). The classical example is that of a factory the smoke from which has harmful effects on neighbors. The usual way of analyzing this problem within economics were in terms of sanctioning the agent (making the owner of the factory liable for the damage caused), or alternatively, placing a tax proportional to the damage. Now, Coase suggests that this way of dealing with the problem is inappropriate, because it does not correspond to an understanding of this kind of situation, which is not unilateral but reciprocal in nature: "The question is commonly thought of as one in which A inflicts harm on B and what has to be decided is: how should we restrain A? But this is wrong. We are dealing with a problem of a reciprocal nature. To avoid the harm to B would inflict harm on A." And Coase gives an example: "the case of a confectioner the noise and vibrations from whose machinery disturbed a doctor in his work. To avoid harming the doctor would inflict harm on the confectioner. The problem posed by this case was essentially whether it was worthwhile, as a result of restricting the methods of production which could be used by the confectioner, to secure more doctoring at the cost of a reduced supply of confectionery products."

By the end of the same article, another notion Coase puts forward makes it clearer why, in a precise sense, we must talk of property as a social relationship: property in not the owning of a physical thing, but rather the possession of a specific right, the right to carry out a specific list of actions. Coase is talking of factors of production and says that it is a faulty concept of them thinking they are physical entities that can be acquired and used: "what the owner in fact possesses is the right to carry out a circumscribed list of action". And that list can change, evolve, and be the subject of dispute - depending on the complex social relationships it relates to. In Coase words: "The rights of a landowner are not unlimited. It is not even always possible for him to remove the land to another place, for instance, by quarrying it. And although it may be possible for him to exclude some people from using 'his' land, this may not be true of others. For example, some people may have the right to cross the land. Furthermore, it may or may not be possible to erect certain types of buildings or to grow certain crops or to use particular drainage systems on the land. This does not come about simply because of government regulation. It would be equally true under the common law. In fact it would be true under any system of law." [Coase, 1960]

Some years later, Harold Demsetz insisted on, and developed the concept of property as a social relationship, as a bundle of rights possibly attached to a physical thing: "In the world of Robinson Crusoe property rights play no role. Property rights are an instrument of society (...). An owner of property rights possesses the consent of fellowmen to allow him to act in particular ways (...) [and] expects the community to prevent others from interfering with his actions, provided that

these actions are not prohibited in the specifications of his rights." [Demsetz, 1967, p.347] Transactions in the market place are not just exchanges of physical commodities or services; they are also, and much more, exchanges of rights. Questions about the bundle of rights attached to a physical commodity or service are prior to those about the physical commodities or services themselves. Prices and quantities in exchanges are not brute facts of nature; they depend on rights attached to them. Thus, it is not surprising that property rights mediate complex relations holding between people and things: "property rights convey the right to benefit or harm oneself or others. Harming a competitor by producing superior products may be permitted, while shooting him may not. A man may be permitted to benefit himself by shooting an intruder but be prohibited from selling below a price floor." [Demsetz, 1967, p.347] Different forms of ownership can be distinguished pointing out who can exercise some rights and who can exclude others from the exercising of some rights [Demsetz, 1967, p.354].

Property rights exemplify one important aspect of the embedding of institutions into the wider world. As already mentioned (see 4.2.2.), the structure of an action arena suffers several influences from factor beyond the reach of immediate action of agents. The biophysical world is one of such factors; the culture and the more general structure of the community is another one. Institutions cannot always be thought in abstraction of these exogenous variables. "We have only to compare property rights in Beirut in the 1980s with those of a modern small-town U.S. community to cover the spectrum. In the former, most valuable rights are in the public domain, to be seized by those with the violence potential to be successful; in the latter the legal structure defines and enforces a large share of rights and those valuable rights in the public domain tend to be allocated by traditional norms of behaviour." [North, 1990, pp. 33-34]

Understanding that institutional environments have so powerful properties because of they allowing sophisticated indirect or mediated interaction is crucial either to deal with all the sophistication of human institutions or to synthesize artificial institutional environments for robots. Since the intellectual bias favouring the prominence of local interaction is so strong, and takes so many different forms, there is a need to insist on some arguments in favour of taking into consideration the action possibilities open by indirect and mediated interaction. The next three paragraphs in a row serve such an endeavour.

First, the spontaneous order hypothesis will be putted to a test within Multi Agent Systems. We give specific attention to this issue, because of its close links to emergentist views of collective systems. The "design for emergence principle" [Pfeifer and Bongard, 2007] states that a desired functionality should not be programmed into a group of agents, but emerge from a set of simple rules of local interaction. Experiments within MAS show that, at least in some situations, this purely direct

and local interaction leads to inefficient solutions to collective problems.
Second, some examples, taken from the theory of judgment aggregation, will illustrate the more general problem of aggregation: how can several individuals within a group make their own individual choices and, at the same time, guarantee that their group will make consistent collective choices. The aggregation problem show at what extent the individual can be asked to understand the level of mediated interaction in order to make his actions compatible with the success, and even the preservation of the group.
Third, the Principal/Agent problem, one of the research topics of the New Institutional Economics, shows that it is not feasible to think only in terms of direct interactions if we need to deal with sophisticated collective systems where some agents act for, or on behalf of, or as representatives of others.

6.2.2 Putting the Spontaneous Order Hypothesis to a test within Multi Agent Systems

Many researchers, dealing with social phenomena within different disciplines, work to show that a social order might spontaneously come into existence and be reproduced without any coordination devices deliberately set up by agents. This "spontaneous order hypothesis", associated to self-organization and emergence, influences research programmes in "mainstream scientific disciplines", like Economics, as well as in quite new disciplines belonging to the cloud of the Sciences of the Artificial (e.g., Multi Agent Systems, Collective Robotics). The spontaneous order hypothesis, with its links to emergentist views, stems from the prominent role conferred to local interaction. In a set of experiences within Multi Agent Systems, Caldas [2001] advances our understanding of the hypothesis researching a series of related aspects of a question that can be so formulated: "Could we show that, at least in some situations, merely emergent processes may lead to inefficient solutions to collective problems?". To that effect, Caldas take situations previously identified in experimental economics and simulate them with a version of the Genetic Algorithm (GA)[16]. Short presentations of some of these simulations will be given, without computational details. The GA population of these simulations represent collections of sets of rules associated with the set of actions available to

[16]Genetic algorithms (GA) were invented and developed by John Holland in the 1960s and the 1970s as an abstraction of biological evolution, and, as such, making us capable of importing the natural phenomenon of adaptation into computer systems [Holland, 1975]. GA is a method for moving from one population of "chromosomes" (each chromosome consisting of "genes ", e.g. bits, each "gene" being an instance of a particular "allele", e.g. 0 or 1) to a new population by using a kind of "natural selection" together with the genetics-inspired operators of crossover, mutation, and inversion . For an introduction, see [Mitchell, 1998].

agents; the fitness function for each agent maximizes his payments.
Co-ordination problem 1. A set of individuals, kept in isolation from each other, must choose one of 16 colours. Each participant choice will be rewarded in accordance with the rule: multiply a fixed amount of money by the number of players that have chosen the same colour (absolute frequency). The experiment repeats a number of times with the same players. After each repetition, players are informed of frequencies and payoffs by colour, so participants can change their choices next time, what they indeed do to maximize payments. From the starting point, where the participants had no motive to choose colour 1 rather colour 2 or any other one, the behaviour rapidly converges to choosing the colour that, at the origin by chance, turned out to be the most often selected. The rule "choose colour x" emerges as a shared norm (convention). It seems that the "spontaneous order hypothesis" works.
Co-ordination problem 2. A new experimental situation departs from the previous one in just one detail. The payoff to each individual now depends, not only on the absolute frequency of the chosen colour, but also on a characteristic made "intrinsic" to each colour by the experimenter. For example, all other factors remaining equal, the choice of the colour number 16 pays 16 times more than colour number 1. The precise design of this artefactual intrinsic characteristic remains unknown to the players. The convergent choices of all participants to colour 16 is the most valuable situation to every participant, but that convergence is highly unlikely to occur in the absence of any opportunity to agree on a joint strategy. An initial accidental convergence to any colour creates an attractor capable of strengthen itself from repetition to repetition. The weight of a very frequent choice increases its atractivity and discourages any participant from moving to a less frequent choice. Even if a participant knows the exact function that determines the payoff, any isolated option for the best theoretical option will neither improve the individual payoff nor move the collective dynamics towards a path conducive to a higher collective payoff. The "spontaneous order hypothesis" is in trouble, even with mere co-ordination problems, when the best for each individual is also the best for the collective (for other individuals).
The situation gets worse with a "co-operation problem", when a moral dilemma is at stake because the best outcome to the collective and the best outcome to an individual are not necessarily coincident.
Co-operation problem. Now, the individuals must post a monetary contribution (from 0 to a predefined maximum) in an envelope and announce the amount contained in it. The sum of all the contributions is multiplied by a positive factor ('invested') and the resultant collective payoff is apportioned among the individuals. For each participant, the share of the collective payoff is proportional to the announced contribution, not to the posted contribution. As all participants

know these rules, they realize that to maximize payoff an individual must contribute nothing and announce the maximum. So, it is with no surprise that, after some initial rounds, free-riding behaviour emerges: the posted contributions tend to zero while the announced contributions are kept close to the maximum. The group follows collectively a path that all of his members consider undesirable: the time will soon arrive when there is no more money to distribute.
This set of experiences suggests collective order does not always emerge from individual decisions alone. There are many types of situations where things are prone to get wrong. However, it can be asked why, if things are really so difficult, so many instances of collective action sustained over time are actually known, and not only collective disaster emerging from mere spontaneous interaction of selfish individuals. Coordination devices deliberately set up by agents, and powerful enough to free agents from mere local and immediate interaction, had proven its usefulness (see Ostrom, 1990, for several case studies, both of robust institutions and failure cases). Now, these coordination devices, of an institutional nature, deeply involve indirect mediated interaction. This strongly suggest that pure direct local interaction can, at least in some situations, prove insufficient to the attainment of efficient solutions to collective problems.

6.2.3 The Aggregation Problem: from individual choice to collective choice

No single individual makes institutional choices. In a institutional-choice situation, the basic alternatives available to the individual are (1) to support the continuance of the status quo rules or (2) to support an alternative set of rules: "Whether or not a change in rules will be accomplished will depend on the level of support for the change and the aggregation rule used in the institutional-choice situation" [Ostrom, 1990, p.194]. This calls our attention to the aggregation problem: it is not an easy thing to reach consistent collective choices from a bundle of individual choices. From the same set of individual choices, alternative aggregation methods can lead to radically different collective choices. Agents focusing exclusively on the individual decisions will be unable to measure the actual meaning of their own option in terms of their impact on collective choice. Aggregation is about the linkage from local interaction to mediated interaction, and back (aggregated outcomes will modify the exogenous variables affecting the internal world of individual choice) (see 4.2.2.). We will use the theory of judgment aggregation to illustrate the more general problem of aggregation.
The theory of judgment aggregation addresses the following question: How can several individuals within a group make consistent collective judgments on a given set of connected propositions on the basis of the group members' individual judgments

on these propositions? How the individuals' judgments can be aggregated into consistent collective judgments? This problem arises in many different settings, ranging from legislative committees to expert panels, from juries and multi-member courts to large social organizations. The recent interest in judgment aggregation was sparked by the observation that majority voting, perhaps the most common democratic procedure, fails to guarantee consistent collective judgments whenever the decision problem in question exceeds a certain level of complexity. This observation was shown to illustrate an impossibility result: roughly speaking, there does not exist any method of aggregation which (i) guarantees consistent collective judgments and (ii) satisfies some other desirable features of collective decision, such as determining the collective judgment on each proposition as a function of individual judgments on that proposition and giving all individuals equal weight in the aggregation [List, 2008, List and Puppe, 2009].
The aggregation problem is relevant to show at what extent the meaning of local direct interaction can be a function of mediated interaction (institutions being the mediator device). To make our point, we will present some aspects of the work of Philip Pettit on the discursive dilemma, previously discussed in a legal context under the name "doctrinal paradox".
An example of the doctrinal paradox is as follows [Pettit, 2003]. A three-judge court has to decide a tort case and consider the defendant liable if and only if the defendant's negligence was causally responsible for the injury to the plaintiff and if the defendant has a duty of care towards the plaintiff. Now, which decision has been taken when judges voted as follows?

	Cause of harm? (Premise 1)	Duty of care? (Premise 2)	Liable? (Conclusion)
Judge A	Yes	No	No
Judge B	No	Yes	No
Judge C	Yes	Yes	Yes

Looking at the table it is easy to see that, in order to know the outcome, the verdict, we need to know what decision procedure is at work. We have two options.
First, each judge individually considers all the available evidence and relevant pieces of law and vote for a given verdict. The verdict flows directly from a vote on the conclusion. This is a conclusion-based decision procedure. In this example, a majority of judges (A and B) vote against the liability of the defendant. The defendant goes free.
Second option, all judges collectively assess the available evidence on each of the

premises and defines the court's opinion about each of the premises. The conclusion follows logically the premises. This is a premises-based decision procedure. In this example, a majority (A and C) considers the defendant's negligence as cause of the injury, and other majority (B and C) considers that the defendant had a duty of care towards the plaintiff. Under these circumstances, the conjunction of the premises implies that the defendant would be found liable.
The doctrinal paradox consists in having different outcomes to the same case, with the same set of votes from all individual judges, just because of the adoption of different procedures. The votes on the case are instances of direct interaction between the judges. The decision procedure is an instance of a mediation mechanism. The outcome doesn't depend only on the current behaviour of the judges, but it suffers a decisive influence from previous decisions on procedural matters. The time and place of voting the case is neither the time nor the space of defining the decision procedure: there is space uncoupling and time uncoupling at work in this scenario. And there is also name uncoupling: most probably, the agents having defined the procedure are not the same as the judges of the current case and they even unknown to them. So, the fundamental properties of mediated interaction are at work in this scenario.
In this example the paradox arises in a case where the conclusion depends on a conjunction of premises. But the paradox can also arise in cases where the conclusion flows from a disjunction of premises. One example is as follows.
Again at a court, a defendant claims for a declaration of nullity of a trial where he had confessed and had been found guilty. The defendant argues on two grounds for claiming nullity. First, illegal procedures had been used to obtain evidence. Second, the confession had been obtained through coercion. In a given legal framework, anyone of these two allegations, if accepted, justifies the nullity of the trial and so the appellant should be given a retrial. This time the situation is as follows:

	Inadmissible evidence? (Premise 1)	Forced confession? (Premise 2)	Retrial? (Conclusion)
Judge A	Yes	No	Yes
Judge B	No	Yes	Yes
Judge C	No	No	No

Again, a three-judge court has to decide whether the appellant should be given a retrial either if (1) inadmissible evidence has been used or (2) a forced confession has taken place. It is not difficult to realize that, again in this situation, a conclusion-based decision procedure yields a given outcome (a retrial would be

given to the appellant) whereas a premises-based decision procedure yields the opposite verdict. It is another instance of the doctrinal paradox.
These examples may seem too formal to be of interest outside courts. But such kind of situation can occur in many decision occasions within organizational structures of some complexity, like companies or schools, or even within more episodic entities, like an appointments committee, a jury or a commission of inquiry. Let's see the tenure example [List, 2006]. A university committee has to decide whether or not to give tenure to a junior academic. The requirement for tenure is excellence in both teaching and research. One of the three committee members finds the candidate excellent in teaching but not in research. The second thinks she is excellent in research but not in teaching. The third thinks she is excellent both in teaching and in research. So a majority considers the candidate excellent in teaching, a majority considers her excellent in research, but only a minority - the third committee member - thinks the candidate should be given tenure. In this situation again, a choice on the decision procedure will make all the difference to the decision itself on the practical question on hand.
These examples show that manipulating some aspects of our institutional environment can radically alters the constraints imposed on individual agents, so importantly changing their ways of life - and, so, they show how important can be mediated interaction to understand the workings of collectives.
Philip Pettit works on several ways to generalize the doctrinal paradox, calling these generalized forms "discursive dilemmas". One kind of these discursive dilemmas takes a diachronic form. It arises in scenarios where individuals belonging to a group contribute to a series of collective decisions taken over a period of time, where some constraints apply to those decisions and where consistency across time is a value to preserve.
Imagine, as an instance of the diachronic sort of the discursive dilemma, that a political party running for general elections must announce a series of budgetary options over time. Taking into account that budgetary resources have limits, and that those limits relate to taxes, this political party cannot simply promise increase spending in every policy while promising not to increase taxes.

	March Increase taxes?	June Increase defence spending?	September Increase health spending?
A	No	No	No (reduce)
B	No	No (reduce)	Yes
C	Yes	Yes	Yes

At some point in a series of decisions of this kind, the party is no longer free to take whatever decision it wants on a new related issue, because of the risk of discredit for adopting a position that is inconsistent with the views previously espoused.
This can happen if the different collective decisions are taken by majority votes, if decisions are taken one by one for successive issues despite their connectedness, and if different occasional majorities form and prevail in each moment of internal decision. One decisive point here is this qualification "occasional", for these occasional majorities. We call them "occasional" because they do not represent any shared global view on the whole set of related issues at stake across time. Different individuals on the same occasional majority have different reasons to support that line at that time. And they will probably vote in disparate ways the next time a related issue is submitted to their consideration.
The dilemma at stake here is the following: will the collective, the group, still allow each individual member to vote exclusively for reasons of their own, without any consideration on the consistency of the resulting series of decisions? Or will the group impose, as a first rank criterion, the rule that no collective decision could in any circumstance be inconsistent with the series of previous decisions, thus imposing a restriction on individual member's choices?
Now, this is a dilemma that any purposive group will frequently experience. They will not be effective promoters of their purposes if they tolerate inconsistency or incoherence in their judgments across time. To guard against inconsistency, and against consequences like defeat or disintegration of the group, the group would need some mechanism or procedure which would allow individual members, or a subset of them, to monitor collective commitments resulting from prior resolutions, and injecting such knowledge into the collective decision procedures, functioning as a collective memory related to the collective purposes, so the group becoming the true agent of deliberation [Pettit, 2007].
Pettit says that these groups are groups with their own mind. But this is not our point here. Our point here is that the understanding of some collective phenomena may require considering the multiple levels at which they occur - and that those multiple levels work by mediated interaction: interacting agents do not have to know each other explicitly; they do not have to be at the same place to be part of the same process; they do not have to act at the same time.

6.2.4 The Principal-Agent Problem

There is an agency relationship, or a Principal/Agent relationship, between two (or more) parties, where one party, the agent, act for, on behalf of, or as representative of a second party, the principal. The Principal/Agent problem stems from the fact that, for several possible reasons, the agent not always acts fully in the interest of

the principal, and the principal faces difficulties in trying to monitor the actions of the agent. It is always costly to the principal to implement measures to limit divergences between his/her own assessment of the situation and the agent's view of the same situation. In general, many exogenous disturbances (e.g., variations in the weather) give the agent valid excuses - or reasons - for bad results [Furubotn and Richter, 1997, pp.148-156,250-258].

At the root of the Principal/Agent problem is information. More precisely, **asymmetry of information**, something that is ubiquitous in real life: "(...), the seller of oranges [know] much more about the valuable attributes of the oranges than the buyer, the used car dealer [know] much more about the valued attributes of the car than the buyer [Akerlof, 1970], and the doctor [know] much more about the quality of services and skill than the patient. Likewise, prospective assistant professors know much more about their work habits than does the department chairman or, to take another example, the purchaser of life insurance from an insurance company knows much more about his or her health than the insurer does. Not only does one party know more about some valued attribute than the other party, he or she may stand to gain by concealing that information." [North, 1990, p.30]

Economics' writers distinguish between two types of consequences of asymmetry of information on a principal/agent relationship: moral hazard, adverse selection. Sometimes the distinction is made coincident to the relative time of events and a contract: we talk of **adverse selection** if the events occur in the period before the conclusion of the contract; we talk of **moral hazard** if events occur during the execution of the contract. Moral hazard can involve an **hidden action**, where the agent's action is not directly observable by the principal (e.g., the effort level put forth by a worker; in a physician-patient relation, and due to the superior knowledge of the physician, the patient cannot know if the physician's action is as diligent as it could be), or **hidden information** (the agent has made some observation that the principal has not made (e.g., the agent knows the output of a department, but the principal does not). In a centrally planned economy, the agent, the manager of a productive unit, knows the output of that unit, but the principal, the central planning unit, does not - and the productive unit may have incentives not to reveal their potentiality, to avoid harder requirements from the central planning unit. The same can be said about units within a single firm. The economic response to a principal/agent problem caused by asymmetry of information can be, on the part of the principal, to design an **incentives plan** that could induce agent to act advantageously for the principal.

Basic models of the principal/agent relationship have been extended into a range of variations, for instance: many agents and relative performance evaluation (there is one principal and problem can enter into scene, where agents who cheat can-

not be identified because joint output is the only observable indicator of inputs); common agency (situations with one agent and several principals); hierarchies; dynamic models (how the repetition of the same situation helps the principal to determine the effort level of the agent; how long-term efficiency can be achieved through a succession of short-term contracts, if short-term contracts cannot guarantee better than limited commitment).
Now what is interesting is that the "principal/agent problem", as a problem of asymmetric information between agents involved in a given relationship, reveals the extent to which the appearance of a direct relationship can be misleading, actually hiding a complex network of indirect relationships linked to the (supposed) direct link. For instance, in the example of the used car transaction, which is in appearance a direct relationship between two agents, if the buyer wants to protect their own interests cannot rely solely on the information exchanged in a direct way in this relationship. The buyer, who is at a disadvantage in what concerns the information supplied under the direct relationship, has to develop some kind of accessory indirect action to offset this disadvantage. The main interaction is the purchase of the car, and there seems to be a direct interaction between seller and buyer. However, one party (the buyer) needs further action to protect their interests (for example, collecting information from alternative sources). The buyer, not wanting to be a sucker, must engage in indirect interaction to position itself in the apparent direct interaction. Incentive plans, above mentioned as approaches to principal/agent problems, are instances of indirect dimensions of apparently direct relationships, because the incentives depend on extra sources of information not given to all parties within the basic relationship.
Understanding the importance of these problems in any collective system where some agents act for, or on behalf of, or as representatives of others, we realize how it would be restrictive to think only in terms of direct relationships.

6.2.5 Mediated interaction with representation. World models

The lessons learned from the testing of the spontaneous order hypothesis, the aggregation problem, and the principal/agent problem, have shown an important feature of collective action: not always the (rational) behaviour of the individual agents can be taken as the only responsible for the outcomes they get from a given situation. Or, if it is so, an alternative can be needed to improve efficiency: mediated interaction must be added. **Sometimes, the structure of the situation itself is the main factor causing the observed results.** Where the structure of the situation is mainly represented by the institutional mechanisms at work, we need to look at these institutions to understand what is going on.
It has been shown with tools from computational economics that the explaining

factors of a given situation are not necessarily linked to the individual agents' capabilities, but can be due to the structure of the situation itself. In a series of experiments, [Gode and Sunder, 1993, 1997, Bosch and Sunder, 2000, Gode, Spear, and Sunder, 2004] have shown that situations previously explained, under the substantive rationality paradigm, as a result of individual rational behaviour, can be explained by the institutional setup itself. Specifically, they have shown that Pareto efficient outcomes are achieved within double auction contexts by "zero-intelligence" (ZI) traders.

On the agents' capabilities side, ZI traders are software agents whose decision rules fall far short of utility maximization. ZI traders are not endowed with any kind of high level intelligence, motivation or learning of the kind human individuals are supposed to enjoy. They just submit random bids and offers, under some imposed simple constraints (like not permitting traders to sell below their costs or buy above their values).

On the institutional side, double auction markets are very specific contexts. Friedman [1993] defines an auction as a market institution in which messages from traders include some price information - this information may be an order to buy at a given price, in the case of a bid, or an order to sell at a given price, in the case of an ask - and which gives priority to higher bids and lower asks. We can allow only buyers or only sellers to make orders, in which case the market is one-sided, or we can allow both, in which case it is two-sided. A double auction is a two-sided auction.

Now, those experiments show that rules of specific market institutions, and not necessarily the individuals' maximizing capabilities, can be responsible for efficient aggregate outcomes. [Gode and Sunder, 1993, p.120]: "Performance of an economy is the joint result of its institutional structure, market environment, and agent behavior." Interpreting some of the earlier experiments of Gode and Sunder, [Denzau and North, 1994, p.7] say that, **in such kind of context, the difference in institutions alone explains the main differences between diverse aggregate outcomes**.

We have been insisting on the need of taking into consideration mediated interaction and on the fact that mediated interaction is an important feature of institutions. In this context, a precision must be made: we can have mediated interaction both with and without representation; and agents involved in mediated interaction can or not be able to change the representation. Human beings engaged in mediated interaction within institutional environments are endowed with representational powers and are able to modify (at some extent) the content of working representations. But this is not always the case. Consider stigmergy.

The term stigmergy was coined by Pierre-Paul Grassé to indicate that individual entities interact indirectly through a shared environment: one individual mod-

ifies the environment and others respond to the modification, and modify it in turn. Grassé used the concept to explain nest construction in termite colonies. A popular means for stigmergic indirect interaction is through pheromones. [Weyns et al., 2005a, p.17] Many experiments within the Swarm Robotics paradigm are inspired by stigmergic mechanisms, where the individual robots are steered by simple rules, have access to local information only, and have no self-organizing capabilities of their own, the coordination of the swarm relying on built-in so called self-organizing properties of the system itself [Bayindir and Sahin, 2007].
As a matter of fact, within stigmergic systems, both natural and artificial, agents are endowed neither with sophisticated representational systems nor with the ability to modify their representations. This can be a severe restriction to the capability of these agents to respond appropriately to complex and demanding environments. Even if swarm systems can perfectly fit the needs of some specific situations, they would necessarily prove inadequate for some other contexts. Consider, for example, the problem of indistinguishable states. Agents can not react to aspects of the world they cannot sense or understand. For a thermostat, the state "the temperature in the room is too cold and there is somebody inside" and the state "the temperature in the room is too cold and there is nobody inside" are indistinguishable. For every pair physical environment/sensorial apparatus agents are equipped with, there are always some pairs of environmental states that are different and indistinguishable [Wooldridge, 2000, pp.39-40]. Even direct interaction is affected by the problem of indistinguishable states, institutional facts make the problem of indistinguishable states harder. Institutional facts are created by constitutive rules with the form "X counts as Y in context C", as in the case of "this piece of paper counts as money in USA". Thus, constitutive rules create new facts, new levels of activity, when the Y term assigns to objects satisfying the term X a new status they do not already had (see 3. 2.2.). To recognize institutional facts agents need to go beyond immediate perception of physical world and know the history of constitutive rules accumulated layer by layer. Institutional facts are distinguishable only with hindsight, not locally.
To engage in sophisticated mediated interaction, agents need sophisticated representational capabilities. Douglass North is one of the prominent economics' writers having put strong emphasis on the power of representation systems in economic contexts. To understand the point we need to recognize how heterogeneous are the social situations agents can find in the real world of societies with institutions. On the one hand, not all social situations are like the double auction context faced by the "zero intelligence" traders above mentioned. On the other hand, much less structured, and much more complex and uncertain situations, require higher levels of intelligence on the part of the agents. But we cannot simply assume that the agents' intelligence will rise without limits in response to

the uncertainty and complexity of real world situations. That move would lead us back to the unrealistic neoclassical assumptions of perfect rationality.
A substantively rational behaviour of the agents in the world would imply, on their part, full knowledge of all possible contingencies, exhaustive exploration of the decision tree, and a correct mapping between actions, events and outcomes [North, 2005, p.7]. The point is not that this kind of behaviour does not exist. It exists in some economic situations, like competitive posted-price markets, where a price is announced indicating what a firm will pay for a commodity or the price at which firm will sell it, this being done without a link to a actual particular exchange of that commodity. In some situations, posted prices of the major companies are aggregated to form postings, serving as benchmarks for a market. For instance, in the Western Canadian oil market, the major companies post prices as a differential to the West Texas Intermediate posted price. In such a context, the environment makes the situation relatively simple to the agents: the price is viewed as a parameter; only the quantity needs to be chosen. That is the kind of situation that favours the behaviour corresponding to the assumptions of the substantive rationality paradigm.
What we must recognize is that the domain of application of the substantive rationality paradigm is not universal. **At least three features of the environment need to be set at favourable values in order to make the situation optimizers-friendly** [Denzau and North, 1994, pp.7-9]:

(i) Complexity: agents are able to maximize in less complex situations, where the environment reduces the choice behaviour to a parametrical behaviour; complexity can be reduced by familiarity: the frequency of similar choice points has a training effect on agents, enabling them to progressively deal easier and easier with recognizable events;

(ii) Motivation: learning how to behave is easier when involving issues central to the agents assessment of themselves and their world; agents will devote some effort to find a solution when receiving some credible indications that their action will actually impact the outcomes;

(iii) Information: being the situation structured in such a way that pertinent, accurate, cheap and timely information is provided to the agents, allowing them to correct bad models, they can more easily engage in a maximizing exercise.

Now, not all economic situation can (most economic situations cannot) be characterized by low complexity, strong motivation, and accessible and cheap information. "If all choices were simple, were made frequently with

substantial and rapid feedback, and involved substantial motivation, then substantive rationality would suffice for all purposes. (...) But not all choices have all these characteristics. (...) there are hard choices, made in institutional settings that are not as conductive to efficiency as the double auction. (...) It is now time to re-focus on the wide range of problems we have so far ignored that involve strong [or Knightian] uncertainty." [Denzau and North, 1994, p.10]
The point is a philosophical one: we are not possibly able to gather exhaustive knowledge of all those aspects of the world that can, in any circumstance, impinge on our actions. "The 'reality' of a political-economic system is never known to anyone" [North, 2005, p.2]. Of course, we collect data about the world and in some ways we are able to correct some misunderstandings about the workings of the world. However, models are underdetermined by the data: "many models fit any finite data sequence, and data alone cannot judge among this multiplicity of 'generalizations'. Instead, one needs theory to generate hypotheses that can be tested, and impose constraints across sets of hypotheses involving different data in order to usefully perform inductions" [Denzau and North, 1994, p.12, footnote 4]. Even this procedure will not eliminate uncertainty, at least because we cannot always expect encountering events to test our theories.
Even if we don't really "know" the "reality", we "do construct elaborate beliefs about the nature of that 'reality' - beliefs that are both a positive model of the way the system works and a normative model of how it should work" [North, 2005, p.2]. This way, within institutional environments, reality is made both of "brute facts" and of "observer dependent facts": "The structure we impose on our lives to reduce uncertainty is an accumulation of prescriptions and proscriptions together with the artifacts that have evolved as a part of this accumulation. The result is a complex mix of formal and informal constraints. These constraints are imbedded in language, physical artifacts, and beliefs that together define the patterns of human interaction." [North, 2005, p.1]
One aspect of this complexity is that "The belief system may be broadly held within the society, reflecting a consensus of beliefs; or widely disparate beliefs may be held, reflecting fundamental divisions in perception about the society" [North, 2005, p.2]. This amounts to an important role played by representations within society. And this is endlessly renewed: "the very efforts of humans to render their environment intelligible result in continual alterations in that environment and therefore new challenges to understanding that environment" [North, 2005, pp.4-5].
One of the hard choice situations [Denzau and North, 1994] consider are **situations requiring the building of internal representations of the agents with whom one interacts**. With a multiplicity of other agents there will be increased opportunities to flaws in one's mental models of other agents. And, of

course, this adds to the physical world itself as a source of complexity. Institutions help human choosers face strong uncertainty by channeling choices into a relatively small set of actions. Historically, human societies have done so using myths, dogmas, taboos, religion, superstition, and all sorts of ideas that one generation passes to the following.
Notwithstanding the fact that "For the most part, economists (...) have ignored the role of ideas in making choices" [North, 2005, p.5], Denzau and North **use the concepts of "mental model, "ideology" and "institution" to try an understanding of those beliefs about the nature of the reality**. Their definitions of those concepts are as following [Denzau and North, 1994, p.4]:

- Mental models are "the internal representations that individual cognit'ive systems create to interpret the environment".
- Ideologies are "the shared framework of mental models that groups of individuals possess that provide both an interpretation of the environment and a prescription as to how that environment should be structured".
- Institutions are "the rules of the game of a society and consist of formal and informal constraints constructed to order interpersonal relationships".

Mental models, as representations, can be purely subjective and internal to the agent. Ideologies and institutions are shared mental models (plus material external devices). **One important aspect of these shared mental models is that the sharing is itself influenced by the institutional environment**: "Humans attempt to use their perceptions about the world to structure their environment in order to reduce uncertainty in human interaction. But whose perceptions matter and how they get translated into transforming the human environment are consequences of the institutional structure, which is a combination of formal rules, informal constraints, and their enforcement characteristics." [North, 2005, p.6]
As mentioned above (4.3.1.), Ostrom [1990] provides a model of the internal world of individual choice. Incorporating mental models, Ostrom [2005] provides a revised model of the individual (Figure 7).

Taking into account the challenge of the incorrigible incompleteness of knowledge and information, either about the natural world or about the network of social relationships, **mental models become a crucial part of the whole picture**. Individuals attempt to create a mental model or a representation of diverse situations so as to be able to make decisions in these settings. Feedback from the world and the shared culture (**"an intergenerational transfer of past experience"**) can affect the mental models. With a large number of participants,

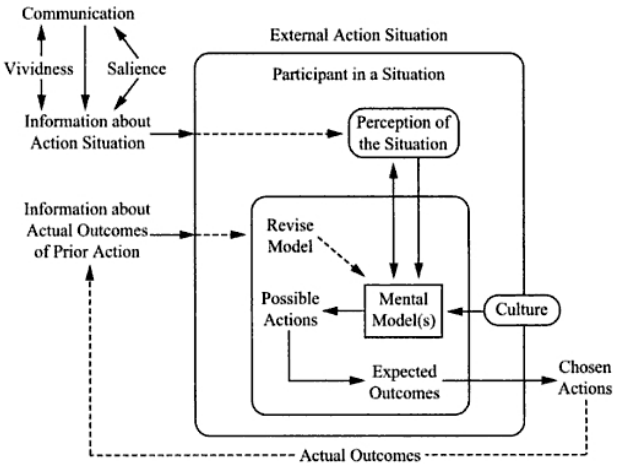


Figure 7: The relationship between information, action-outcome linkages, and internal mental models - and the impact of communication, vividness, and salience on that relationship. In (Ostrom, 2005).

a complex situation, frequent changes, or irregular participation of the agent in the interaction, the convergence of all (or even most of) the participants to a unique picture of the situation is not likely to happen. Information is costly; agents have just imperfect perception and information-processing capabilities. Probably, not all agents will have the same mental models in use. Not all pieces of information have the same salience (the degree to which an element is linked to possible changes in the welfare of the decision maker) and vividness (the amount and quality of the sensory details of the objects encountered), so not gaining the same attention from the agents. If we try to forget about the diversity of mental models, dealing with some particular vision of the world as the unique possible vantage point to that world, we will almost unavoidably get things wrong. The same if we try to synthesize artificial agents as if they could be able to get the big picture from God's point of view. For real agents, **interaction must replace the unreachable goal of perfect information**.

In trying to use mental models and ideologies with systems of multiple robots, we can take these notions in a simplified version inspired by our [Silva and Lima, 2007]. An "ideology" is a set of mental models (not necessarily fully consistent) shared by a subset of all agents. Its spreading among agents largely overlaps with sets of agents linked to some subset of the institutional network. An "ideology" can be "offered" by an institution to any agent prone to adhere or be a condition for adhesion. An "ideology" can result from a modification of the sensor fusion process (modification of the criteria to weight different individual contributions,

for example). "Ideologies" can be about the physical or the social world. Modifying the perception of the agents and their behaviours, "ideologies" can affect the functioning of institutions in many ways: for example providing alternative stereotyped ways of sensing certain situations ("ignore such and such data streams") or undermining mechanisms of social control ("break that rule and we will pay the fine for you with a prize").
If, as we see, authors writing within the tradition of Institutional Economics are interested in the role of world models in social systems populated by natural intelligent agents, authors from the Sciences of the Artificial also have been interested in the role of world models in artificial intelligent systems. James Albus gives us a classic, and a useful example, because it integrates the world models in a proposed global architecture of an intelligent system. We take directly from [Albus, 1991, p.477] the summary of his proposal:

> "The proposed system architecture organizes the elements of intelligence so as to create the functional relationships and information flow shown in Fig. 1. In all intelligent systems, a sensory processing system processes sensory information to acquire and maintain an internal model of the external world. In all systems, a behavior generating system controls actuators so as to pursue behavioral goals in the context of the perceived world model. In systems of higher intelligence, the behavior generating system element may interact with the world model and value judgment system to reason about space and time, geometry and dynamics, and to formulate or select plans based on values such as cost, risk, utility, and goal priorities. The sensory processing system element may interact with the world model and value judgment system to assign values to perceived entities, events, and situations."

Section XII of [Albus, 1991, pp.485-486], is dedicated to world models. The following definition is given: "The world model is an intelligent system's internal representation of the external world. It is the system's best estimate of objective reality."
The basic functions of world models within an intelligent system are enumerated and described this way: (1) update knowledge database with recognized entities and prediction errors (based on correlations and differences between world model predictions and sensory observations at different levels); (2) predict sensory data (that can later be compared with actual sensory data); (3) answer "what is?" queries from task executor and return current state of the world, to be used by behavior generating modules; (4) answer "what if?" queries from task planner and predict results for evaluation (so performing the function of a simulator that helps evaluating plans before trying to actually running them).

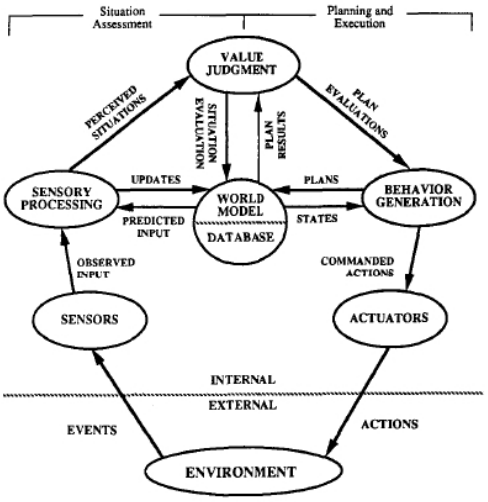


Figure 8: Elements of intelligence and the functional relationships between them. Fig. 1 in (Albus, 1991)

The basic link between what we have said above about mediated interaction and this notion of world models is the following: "The world model contains knowledge of things that are not directly and immediately observable. It enables the system to integrate noisy and intermittent sensory input from many different sources into a single reliable representation of spatiotemporal reality." (p. 485) It should be mentioned, however, that this conception sees the world and its models as if they, ideally, would coincide.The distinction between the internal representation of the world (in the mind), and the external world of reality - is recognized. But it is like the only problem is about the noisy and intermittent input, and the need to integrate data from different sources. The fundamental problems of real agents dealing with information in the real world are beyond this approach. We still are far from the institutional complexity mentioned by Economics' writers. In any case, an approach like that of Albus has the advantage of helping to recognize the importance of representations, of models of the world, in the workings of intelligent systems within environments where mediated interaction is at stake.

7 Conclusions and Future Work

This report sought to provide a structured set of concepts from Institutional Economics to be applied in developing an approach to control systems of multiple robots. We now suggest some links between these concepts, previous work done at ISR / IST, and subsequent phases of the project "From Bio-Inspired to Institutional-Inspired Collective Robotics".

Some bridge-concepts

One way to get to work in robotics within an institutional framework is to use certain concepts that form a bridge between the social sciences' approach and the roboticists' approach. We propose four concepts that can help make that bridge.

Institutions. "Institutions are cumulative sets of persistent artificial modifications made to the environment or to the internal mechanisms of a subset of agents, thought to be functional to the collective order." [Silva, Ventura, and Lima, 2008]
where

- "cumulative" excludes ad hoc interventions (e.g., "breaking legs to implement prohibition") as institutional devices; (broken legs can persist for a while, but cannot "accumulate");
- "persistent" excludes occasional, fortuitous;
- "artificial" means feasible by the agents or their ancestors; excludes "natural";
- "thought to be functional to the collective order": it is not about specific collective goals or particular circumstances; it is about "constitutional" aspects of how many agents interact; e.g., at a nation level, we can change policies and even the government without changing the constitution.

Institutions as "packing information" and "packing decision" devices. Essential features of institutional agents are bounded rationality and bounded autonomy. Institutions are a means to face uncertainty in natural and social world making the best of the resources agents have at hand, because they are "packing information" and "packing decision" devices.
Packing information devices. If a device processes useful information from physical and social environment and provides a result, easy to "swallow" and fitting the needs of individual robots and of the system, it reduces the costs of dealing with the world. ("Traffic controllers" for robots driving in a scenario are of such a kind

of device: they spare the other robots the burden of collecting and analyzing a lot of information about the state of the world, providing a representative small piece of information that is all robots need: "go now" or "wait".) If the institution is well designed, the picture it draws of the world is more inclusive and more accurate than the picture any individual alone could draw.
Packing decision devices. If a large series of individual decisions taking place at similar situations at different times on the same subject can be replaced by a device replicating the core steps of the decisions process, that device spares the individual agent the burden of many repeated decision-making processes eventually leading to the same result. If a large number of agents, sharing the same goal, have to manage a difficult and time consuming decision-making process to achieve coordination, a device taking care of that, while preserving a fair distribution of efforts and rewards, reduces the (individual and collective) effort needed to pursue the goal. (The traffic controllers from the corridor case study are also "packing decision" device: their decision replaces many decisions by all other robots.)
Institutional roles/positions, as well as habits and routines (routines are habits within an organization, linking institutional roles/positions), exemplify "packing information" and "packing decision" devices.

Representations / models of the world. A model of the world is an internal representation of the (physical and/or institutional) world, namely of those parts/aspects of the external world that are not directly observable, not currently observable, or not observable at all.
Instead of having one global world model to serve in all circumstances, agents can have a set of (partially consistent) mental models for specific purposes.
Modifying how sensor data impact the control of actuators, changing models of the world contribute to changing behaviours. Sensor data can even be superseded by internal models of the world, because agents are not always checking the current state of the world and sometimes rely primarily on their internal models of the external world. Changing models of the world in use by a set of agents does not necessarily result from actual changes in the objective external world; they can result from other agents' influencing (agents persuading agents). So, some changes in the world can be the result of sharing models of the world (I change my mind, I persuade others so changing their minds, many individual behaviours change this way, the aggregated result possibly is a new collective behaviour).

Satisficing for bounded-rational individuals. Institutional approaches do not accept hyper-rationalist assumptions about individuals, namely those related to conscious efforts from individuals to collect all the relevant information or to maximize utility. Individuals do have concerns about costs and benefits, but, ex-

cept for specific contexts (e.g., stock market traders), they do not go further than coarse comparisons to what they can see others obtaining. And, perhaps more important, real social agents in sophisticated environments have to deal with a variety of incommensurable goals. For instance, immediate economic advantages, long term coalitions on economic behaviour, and also loyalty to family and neighbours, health concerns, and many other aspects can imply different metrics to the same problem to be solved. General ("moral") reasoning will eventually arbitrate. Individual heterogeneity within a population (due to different time horizons, different opportunities, different level of internal moral pressure) adds still more complexity to the picture (because you can adjust your expectations by misplaced imitation: imitating someone that is not in the some situations you are).
We could try to deal with this problem (modelling a bounded-rational individual) with a modified version of "satisficing". The word "satisfice" (combining satisfy with suffice) was coined by Herbert Simon in 1956, as an alternative to "maximize". Satisficing is to do "well enough". Some approaches to "satisficing" try to make this concept close to the "maximizing" one, including all the efforts to collect information and to compute alternatives into the equation. But this will lead us back to the same kind of difficulties. Sometimes, "satisficing" for a goal treat that goal just as a constraint associated with a diversity of other goals. For example, within the behavioral theory of the firm, profit is not a goal to be maximized, but a constraint: a critical level of profit must be achieved by firms; thereafter, priority is attached to the attainment of other goals. We could try to construct a modified concept of satisficing making it closer to "imitation to what fellows are getting in the neighborhood". Satisficing levels can be dynamically adjusted according to the experience: an agent seeing that others do much better than he has being doing can set higher levels of aspiration (much worse / lower levels).

The four fundamental properties of collectives

The project "From Bio-Inspired to Institutional-Inspired Collective Robotics" identifies four fundamental properties of the collectives, taken as relevant in trying to manage the micro-macro link. The provisional definitions given for each of these properties are taken, and for each one of them, we make some initial suggestions about the contribution institutional concepts can represent to its implementation.

Stability concerns the response of a collective to a perturbation on the coupling between the agents (eliminating communication links among collective members). Models of the world can help here. During periods where no communication at all is possible among the population of agents, models of the world can replace, at least for a while, actual data from natural and social world. Where some commu-

nication is still available, persisting links can be used to update specific aspects of the most needed models of the world.

Robustness will be assessed by removing or adding individuals with specific roles in the collective, and appraising the consequences.
Within an institutional framework, institutional roles (or positions) must be distinguished from particular individual agents. Agents are heterogeneous with respect to some features, but fully interchangeable with respect to some other (basic) features. This makes any agent in principle able to play any role (to occupy any position), even if some learning can be required to attain full mastery. Agents are redundant in relation to positions (roles). To this effect, different roles must not be allocated by fixed, once for all, mechanisms (genetic mechanisms?) - but, instead, by institutional assignment of status functions. If this can be implemented, removing specific individuals from the collective does not amount to renounce to specific roles. On the other side, the adding of individuals with malevolent roles can be countered by a specific feature of institutional roles: to an individual to play a role, other participants must recognize that role as part on the institutional setting, and accept to behave accordingly. The refusal to accept an individual playing a role (because the role is not part of the institutional setting) can be a mechanism to prevent the intrusion of malevolent roles. To make this mechanism work we need individuals having a representation of the whole institutional setting.

Adaptation will be evaluated by performing changes on the environment and evaluating the collective response to those changes.
Norms (social norms), combined with some degree of conformism to "social order" on the part of individual agents, can help individuals to adapt their behaviour to new conditions. Of course, this depends on someone monitoring the environment and taking appropriate decisions to change the norms. However, this does not requires a central authority: different agents can specialize on monitoring different aspects of the environment, and on consulting (a sample of) others on how to change the relevant norms. This amounts to combine norms with social roles/positions.

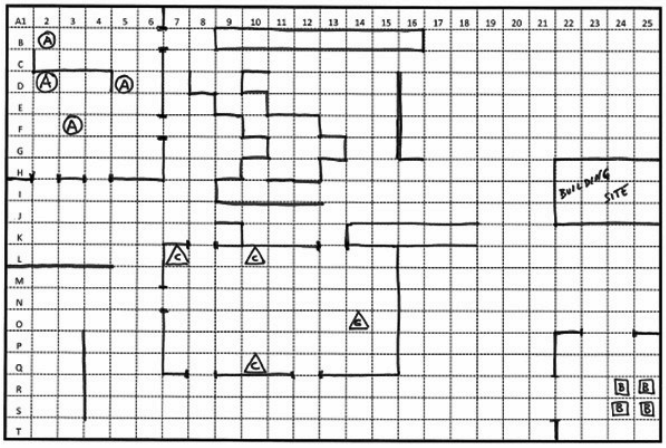
Innovation requires radical changes on the environment to be tested, such that rules of interaction between individuals are no longer appropriate, and new rules have to be set.
To face this challenge we need to exploit self-organizing and self-governing capabilities of the agents. Those capabilities of the agents depend on them being able to act in multilevel environments, so being able to change rules that impact (in an indirect way) the operational level of immediate action. However, I see no imme-

diate suggestion to implement this.

Experiments.

To raise a discussion on how to plan experiments for institutional concepts we make here some initial suggestions.

To experiment institutional concepts as means to deal with the micro-macro problem I suggest using social dilemmas' scenarios. A "social dilemma" exists where there are no prima facie way to make actions guided by the pursuing of (perceived) individual utility easily compatible to (perceived) collective utility. A simplified model of common-pool resources could be the basic setup, because it allows taking into account two important features of most complex social situations at human level: the problematic sustainability of the resources and the temptation to free-ride.



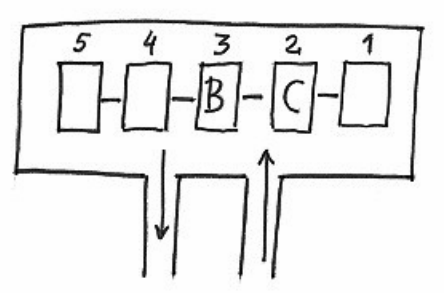
Basic elements of such a scenario could be like the following:

- the basic task is to construct as many specimens of a "virtual object" as possible by assembling, in a specified way, tokens of different resources that can be found in the environment (components A, B, and C to be assembled as virtual objects A+B+C); the variation of the basic task should be easily implementable;
- the experiment takes place in a 2D space; within this virtual environment, there is the "building site" (where the assembling takes place), and fields of sources of the different components needed to build the virtual object;
- individual robots try to maximize private utility functions (delivering components to the building site); the system has a collective utility function,

mainly directed to the global task of building as many virtual objects as possible; the basic social dilemma springs from these two kinds of utility functions;

- the resources system is a renewable resource system; it has a specific replenishment rate; the rate of withdrawal must be balanced with the replenishment rate to avoid (reversible) damage or (irreversible) destruction of the resource system; initially, robots don't have information neither about the localization of the spots where components can be collected nor about replenishment rates of components' sources.

Where the clash between individual and collective utility is more directly seen is at the building site:



- Delivering a component to slot 5 rewards 5 times more than delivering the same component to slot 1. At the example situation (figure), an A component should be delivered to slot 4 to immediate completion of an object, but delivering that component to slot 5 is more rewarding to the individual agent carrying it.

To start a discussion on this, I suggest two possible experiments:

(1) To experiment with institutional roles/positions as "packing information" and "packing decision" devices, concentrate first on the functioning of the building site and introduce an "assembler" there. The assembler works in the antechamber of the building site as follows. He accepts from any three robots the delivery of three components (A, B, C) allowing immediate construction of an object. He calculates the most profitable way to put these components to the slots, giving priority to collective utility, then calculates the sum reward corresponding to this delivery, takes to itself a certain percentage, divide the remainder by the three robots and immediately attributes the reward to each individual robot, delivers

the components to the building site and collect their own reward. After a while, the system calculates individual and collective rewards collected during a given period. Another run of the experiment goes without the assembler and letting each individual robot pursuing individual utility. Individual and collective rewards with both "regimes" are compared. The idea is to experiment if (and at what extent) roles/positions can improve both individual and collective utility and parameters influencing their performance (for example, the percentage retained by the assembler may vary, and it may impact the result: a too expensive assembler can make the device fail when compared to purely individual behaviour).

(2) To experiment with collective construction of models of the world, a shared global map of the localization of components' sources should be built on the basis of information found by individual robots while doing their individual jobs. The performance of the whole system (and of individuals) with and without that shared map should be compared with different runs of the experiment.

References

George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, August 1970.

James S. Albus. Outline for a theory of intelligence. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3):473–509, 1991.

E. Alonso. Rights and argumentation in open multi-agent systems. *Artificial Intelligence Review*, 21(1):3–24, 2004.

L. Bayindir and E. Sahin. A review of studies in swarm robotics. *Turkish Journal of Electrical Engineering*, 15(2):115–147, 2007.

Antoni Bosch and Shyam Sunder. Tracking the invisible hand: Convergence of double auctions to competitive equilibrium. *Computational Economics*, 16(3): 257–284, December 2000.

Eric Brousseau and Mhand Fares. The incomplete contract theory and the new-institutional economics approaches to contracts: Substitutes or complements? In Claude Mnard, editor, *Institutions, Contracts and Organizations - Perspectives from New-Institutional Economics*, chapter The Incomplete Contract Theory and the New-Institutional Economics Approaches to Contracts: Substitutes or Complements?, pages 399–421. Northampton:MA, Edward Elgar Publishing, 2000.

Eric Brousseau and Jean-Michel Glachant. The economics of contracts and the renewal of economics. In Eric Brousseau and Jean-Michel Glachant, editors, *The Economics of Contracts: Theories and Applications*, pages 3–30. Cambridge, Cambridge University Press, 2002.

José M. Castro Caldas. *Escolha e Instituições. Análise Económica e Simulação Multiagentes*. Oeiras, Celta, 2001.

Ronald H. Coase. The problem of social cost. *Journal of Law and Economics*, III: 1–44, 1960.

Ronald H. Coase. The institutional structure of production (prize lecture - 1991 nobel prize in economic sciences). In C. Ménard and M. M. Shirley, editors, *Handbook of New Institutional Economics*, pages 31–39. Springer, 2002.

Rosaria Conte and Cristiano Castelfranchi. *Cognitive and Social Action*. London, The University College London Press, 1995.

Sue Crawford and Elinor Ostrom. A grammar of institutions. In *Understanding Institutional Diversity*, pages 137–174. Princeton, Princeton University Press, 2005.

Francisco Teixeira da Mota. *Alves Reis Uma História Portuguesa*. Lisboa, Contexto Editora and Público, 1996. 4 volumes.

U. N. Danner. *By Force of Habit. On the Formation and Maintenance of Goal-Directed Habits. Doctoral Thesis in Social Psychology.* PhD thesis, Utrecht University, November 2007. Available at http://igitur-archive.library.uu.nl/dissertations/2007-1102-201327/index.htm.

U. N. Danner, H. Aarts, and N. K. De Vries. Habit vs. intention in the prediction of future behaviour: The role of frequency, context stability and mental accessibility of past behaviour. *British Journal of Social Psychology*, 47:245–265, 2008.

Harold Demsetz. Toward a theory of property rights. *The American Economic Review*, 57(2):347–359, 1967.

Arthur Denzau and Douglass C. North. Shared mental models: Ideologies and institutions. *Kyklos*, 47(1):3–31, 1994.

M. Bernardine Dias, Robert Zlot, M. Zinck, J. P. Gonzalez, and Anthony Stentz. A versatile implementation of the traderbots approach to multirobot coordination. In *Proceedings of the International Conference on Intelligent Autonomous Systems (IAS)*, page 325334, 2004.

M. Bernardine Dias, Robert Zlot, Nidhi Kalra, and Anthony Stentz. Market-based multirobot coordination: A survey and analysis. *Proceedings of the IEEE*, 94(7 (Special Issue on Multirobot Coordination)):1257–1270, 2006.

E. H. Durfee. Challenges to scaling up agent coordination strategies. In T. A. Wagner, editor, *An Application Science for Multi-Agent Systems*, pages 113–132. Dordrecht, Kluwer Academic Publishers, 2004.

J. M. Epstein and R. Axtell. *Growing Artificial Societies: Social Science from the Bottom Up.* Washington D.C., The Brookings Institution and The MIT Press, 1996.

Daniel Friedman. The double auction institution: A survey. In Daniel Friedman and John Rust, editors, *The Double Auction Market: Institutions, Theories and Evidence*, pages 3–25. Cambridge, MA, Perseus Publishing, 1993.

Eirik G. Furubotn and Rudolf Richter. *Institutions and Economic Theory. The Contribution of the New Institutional Economics*. Ann Arbor, University of Michigan Press, 1997.

B. P. Gerkey and M. J. Mataric. Sold!: auction methods for multirobot coordination. *IEEE Transactions on Robotics and Automation*, 18(5):758–768, 2002.

Dhananjay K Gode and Shyam Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *The Journal of Political Economy*, 101(1):119–137, February 1993.

Dhananjay K Gode and Shyam Sunder. What makes markets allocationally efficient? *The Quarterly Journal of Economics*, 112(2):603–630, 1997.

Dhananjay K Gode, Stephen Spear, and Shyam Sunder. Convergence of double auctions to pareto optimal allocations in the edgeworth box. Working Paper No. 04-30, Yale International Center for Finance, May 2004. Available at http://ssrn.com/abstract=1280707.

S. Grossman and O. Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719, 1986.

C. Hahn, B. Fley, M. Florian, D. Spresny, and K. Fischer. Social reputation: a mechanism for flexible self-regulation of multiagent systems. *Journal of Artificial Societies and Social Simulation*, 10(1), 2007. Available at http://jasss.soc.surrey.ac.uk/10/1/2.html.

G. Hardin. The tragedy of the commons. *Science*, 162:1243–1248, 1968.

O. Hart and J. Moore. Incomplete contracts and renegotiation. *Econometrica*, 56: 755–786, 1988.

H. Hexmoor, S. G. Venkata, and R. Hayes. Modelling social norms in multiagent systems. *Journal of Experimental and Theoretical Artificial Intelligence*, 18(1): 4971, 2006.

Geoffrey M. Hodgson. *Economics and Institutions A Manifesto for a Modern Institutional Economics*. Cambridge, Polity Press, 1988.

Geoffrey M. Hodgson. What is the essence of institutional economics? *Journal of Economic Issues*, 34(2):317–329, 2000.

Geoffrey M. Hodgson. Reclaiming habit for institutional economics. *Journal of Economic Psychology*, 40(1):1–25, 2004.

Geoffrey M. Hodgson. What are institutions? *Journal of Economic Issues*, XL (1):1–25, March 2006.

Geoffrey M. Hodgson. Institutions and individuals: Interaction and evolution. *Organization Studies*, 28(1):95–116, 2007.

Geoffrey M. Hodgson and Thorbjrn Knudsen. The complex evolution of a simple traffic convention: The functions and implications of habit. *Journal of Economic Behavior and Organization*, 54(1):19–47, 2004.

John Holland. *Adaptation in Natural and Artificial Systems*. Cambridge:MA, The MIT Press, 1975.

F. H. Knight. *Risk, uncertainty and profit*. New York, Harper & Row, 1921.

Christian List. The discursive dilemma and public reason. *Ethics*, 116:362–402, 2006.

Christian List. Judgment aggregation: a short introduction. In U. Maki, editor, *Handbook of the Philosophy of Economics*. Amsterdam, Elsevier, 2008.

Christian List and Clemens Puppe. Judgment aggregation: a survey. In P. Anand, C. Puppe, and P. Pattanaik, editors, *Oxford Handbook of Rational and Social Choice*. Oxford, Oxford University Press, 2009.

Eric Malin and David Martimort. Transaction costs and incentive theory. In Eric Brousseau and Jean-Michel Glachant, editors, *The Economics of Contracts: Theories and Applications*, pages 159–178. Cambridge, Cambridge University Press, 2002.

Claude Ménard and Mary M. Shirley. Introduction. In Claude Ménard and Mary M. Shirley, editors, *Handbook of New Institutional Economics*, pages 1–18. Springer, 2005.

Melanie Mitchell. *An Introduction to Genetic Algorithms*. Cambridge: MA, The MIT Press, 1998.

Douglass C. North. *Institutions, Institutional Change and Economic Performance*. Cambridge, Cambridge University Press, 1990.

Douglass C. North. Institutions. *Journal of Economic Perspectives*, 5(1):97–112, Winter 1991 1991.

Douglass C. North. *Understanding the Process of Economic Change*. Princeton, Princeton University Press, 2005.

G. P. ODriscoll and M. J. Rizo. *The Economics of Time and Ignorance*. Oxford, Basil Blackwell, 1985.

Mancur Olson. *The Logic of Collective Action. Public Goods and the Theory of Groups*. Cambridge (Massachusetts), Harvard University Press, 1965.

Elinor Ostrom. *Governing the Commons. The Evolution of Institutions for Collective Action*. Cambridge, Cambridge University Press, 1990.

Elinor Ostrom. *Understanding Institutional Diversity*. Princeton, Princeton University Press, 2005.

Elinor Ostrom and Sue Crawford. Classifying rules. In *Understanding Institutional Diversity*, pages 186–216. Princeton, Princeton University Press, 2005.

Philip Pettit. Groups with minds of their own. In Frederick Schmitt, editor, *Socializing Metaphysics*, pages 167–193. London, Rowman & Littlefield, 2003.

Philip Pettit. Rationality, reasoning and group agency. *Dialectica*, 61(4):495–519, 2007.

R. Pfeifer and J. Bongard. *How the Body Shapes the Way We Think*. Cambridge: MA, The MIT Press, 2007.

Malcolm Rutherford. Institutional economics: Then and now. *The Journal of Economic Perspectives*, 15(3):173–194, Summer 2001.

J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):3360, 2005.

G. von Schmoller. *Grundriss der Allgemeinen Volkswirtschaftslere*. Munich and Leipzg, Duncker and Humblot, 1900.

John R. Searle. *The Construction of Social Reality*. New York, The Penguin Press, 1996 edition, 1995.

John R. Searle. Social ontology: some basic principles. *Anthropological Theory*, 6 (1):12–29, 2006.

Porfírio Silva and Pedro Lima. Institutional robotics. In Fernando Almeida e Costa, Luis Mateus Rocha, Ernesto Costa, Inman Harvey, and António Coutinho, editors, *Advances in Artificial Life. Proceedings of the 9th European Conference, ECAL 2007*, volume 4648 of *Lecture Notes in Computer Science*, pages 595–604. Springer-Verlag, 2007.

Porfírio Silva, Rodrigo Ventura, and Pedro Lima. Institutional environments. In B. Jung, F. Michel, A. Ricci, and P. Petta, editors, *From Agent Theory to Agent Implementation, Proceedings of Workshop AT2AI-6, AAMAS 2008 - 7th International Conference on Autonomous Agents and Multiagent Systems*, pages 157–164, 2008.

Herbert A. Simon. From substantive to procedural rationality. In S. Latsis, editor, *Methods and Appraisals In Economics*, pages 129–148. Cambridge University Press, 1976.

Herbert A. Simon. Bounded rationality. In Eatwell, Milgate, and Newman, editors, *The New Palgrave: A Dictionary of Economics*, pages 266–267. London, MacMillan, 1987.

Luca Tummolini and Cristiano Castelfranchi. The cognitive and behavioral mediation of institutions: Towards an account of institutional actions. *Cognitive Systems Research*, 7(2-3):307–323, 2006.

Danny Weyns, H. Van Dyke Parunak, Fabien Michel, Tom Holvoet, and Jacques Ferber. Environments for multiagent systems, state-of-the art and research challenges. In Danny Weyns, H. Van Dyke Parunak, and Fabien Michel, editors, *Proceedings of the 1st International Workshop on Environments for Multi-Agent Systems*, pages 1–47. Berlin and Heidelberg, Springer-Verlag, 2005a.

Danny Weyns, Michael Schumacher, Alessandro Ricci, Mirko Viroli, and Tom Holvoet. Environments in multiagent systems. *The Knowledge Engineer Review*, 20(2):127–141, 2005b.

Henry Wigan. The effects of the 1925 portuguese bank note crisis. Working Paper No. 82/04, Department of Economic History, London School of Economics, February 2004. Available at http://www2.lse.ac.uk/economicHistory/pdf/WP8204.

Oliver E. Williamson. *Markets and Hierarchies: Analysis and antitrust implications*. New York, The Free Press, 1975.

Oliver E. Williamson. *The Economic Institutions Of Capitalism*. New York, The Free Press, 1985.

Oliver E. Williamson. *The Mechanism of Governance*. Oxford, Oxford University Press, 1996.

Oliver E. Williamson. Transaction cost economics. In C Mnard and M. M. Shirley, editors, *Handbook of New Institutional Economics*, pages 41–65. Springer, 2005.

M. Wooldridge. Intelligent agents. In G. Weiss, editor, *Multiagent Systems A Modern Approach to Distributed Artificial Intelligence*, pages 27–77. Cambridge:MA, The MIT Press, 2000.