Institutional Environments

Porfírio Silva Rodrigo Ventura Pedro U. Lima
Institute for Systems and Robotics,
Instituto Superior Técnico
Lisbon, Portugal
porfiriosilva@isr.ist.utl.pt yoda@isr.ist.utl.pt pal@isr.ist.utl.pt

ABSTRACT
The concept of environment is of paramount relevance for new strategies to model systems of multiple artificial agents. This paper introduces a set of definitions designed to guide the modelling of institutional environments. This is part of ongoing research on a new strategy to conceptualize multi-robot systems, which takes a network of institutions as the control system for a collective of artificial embodied agents with bounded rationality and bounded autonomy. The definitions, given as structured tuples, attempt to capture a hypothesis on the main constitutive elements of the social order dynamics. That hypothesis is part of the institutional approach, which aims at responding to some difficulties of current perspectives on environment.

Categories and Subject Descriptors
J.4 [**Computer Applications**]: Social and Behavioral Sciences – *Economics, Sociology.*

I.2.9 [**Computing Methodologies**]: Artificial Intelligence – Robotics.

General Terms
Design, Economics, Theory

Keywords
Institutional Environments, Institutional Robotics

1. INTRODUCTION
This paper is part of an ongoing research on a new strategy to conceptualize multi-robot systems, which takes a network of institutions as the control system for a collective of artificial embodied agents with bounded rationality and bounded autonomy [15]. We conceive institutional environments as networked institutions embedded in wider environments. Our aim here is to suggest a set of definitions designed to guide the modelling of institutional environments.

The definitions, given as a tuples structure, try to capture a hypothesis on the main constitutive elements of the social order dynamics. The suggested definitions for "node of the institutional network", "institutional agent", and "institutional network", framed by an explicit presentation of our hypothesis on dynamics of social order, are given in Section 4.

Motivations for our perspective on institutional environments are presented in Sections 2 and 3. In Section 2 we mention some contributions to the emergence of the concept of environment as a tool of paramount relevance for new strategies to model systems of multiple artificial agents. Some difficulties associated with such concepts are also considered. In Section 3 we refer to the Institutional Robotics approach as the framework for our current research on institutional environments.

2. MIND AND ENVIRONMENT: FROM MENTALISM TO INTERACTION
The concept of environment emerges as a tool of paramount relevance for new strategies to model systems of multiple artificial agents. In this Section we mention some contributions to that process, mainly related to a shift from mentalist to interactionist underlying metaphors.

According to [1:49–54], a metaphor has been prevailing over cognitive science research programme on mind. It is the metaphor of an abstract inner space opposed to the outside world, whether the outside world includes the body or not. That same metaphor conceives a boundary between inner and outer spaces being traversed by perceptive stimuli (headed inward) and behavioural responses (headed outward). The unsuitability of this metaphor reveals itself where this dominant approach to mind is driven by its own difficulties to blur the difference between inside and outside, as a consequence of the endeavour to reproduce the entire world inside the head. This diagnosis of what Agre calls "mentalism" helps to understand the three great neglects at the heart of Good Old-Fashioned Artificial Intelligence: the neglect of the body, of the world, and of other agents.

Philip Agre is one of the proponents of interactionism as an alternative to mentalism, both to analyse living agents and to design artificial ones. To the interactionist alternative the central phenomenon is the interaction of agents with their environment [1:57–58]: "I propose thinking about computation in terms of machinery and dynamics. A machine (. . .) is an object in the physical world that obeys the laws of physics. [The dynamics] concerns the interactions between an individual (robot, ant, cat, or person) and its surrounding environment."

Andy Clark [3] also explains why there is a plastic frontier between mind, body, and world. On the one hand, it comes from natural evolution. Clark stresses that most of our daily behaviour is niche dependent. This means that we are not "general machines" prepared for every possible contingency, but instead sensitized to those particular aspects of the world that have special significance because of our way of life. On the other hand, there is also the impact of our culture on the world. We adapt our surroundings to our needs and life style. We perform "epistemic actions" [9], we organise things on space to unload computation

Jung, Michel, Ricci & Petta (eds.): *AT2AI-6 Working Notes, From Agent Theory to Agent Implementation, 6th Int. Workshop,* May 13, 2008, AAMAS 2008, Estoril, Portugal, EU.

to the environment. One of the many examples given by David Kirsh, and mentioned by Clark, is: ”To repair an alternator, take it apart but place the pieces in a linear or grouped array, so that the task of selecting pieces for reassembly is made easier.” Language and arithmetic are widespread cognitive scaffolding tools for human beings.

Paul Dourish [5], while sharing interactionist views, strives for “embodied interaction”. Embodiment, the central element of the perspective Dourish puts forward, focus the study of cognition on the agent’s practical action on his world. Embodiment, far from being restricted to an agent situation in a physical environment, also counts on the social and organizational environments, and stresses the participative status the agent enjoys [5:19]: “Physical environments are arranged so as to make certain kinds of activities easier (or more difficult), and in turn, those activities are tailored to the details of the environment in which they take place. The same thing happens at an organizational level; the nature of the organization in which the work takes place will affect the work itself and the ways it is done.”

According to Henry Petroski [12], a deep aspect of our worldly condition is that we are surrounded by objects that are shaping and are being shaped by the slightest aspects of our daily life. He mentions Donald Norman’s suggestion that “there are perhaps twenty thousand everyday things that we might encounter in our lives”. However, he argues against rationalist conceptions of artefacts: form does not follow function. Necessity and utility does not determine technological diversity. Already in 1867 Karl Marx was surprised to learn that five hundred different kinds of hammers were produced in Birmingham. Practical use always goes beyond rational anticipation, and the variety of entrants in any design competition shows at what extent the specification of a design problem in no way dictates its solution. Petroski’s reflections could help to generalise the notion of artefact, which, according to [13:130], has been introduced recently in Multiagent Systems (MAS) as a first-class abstraction representing devices that agents can either individually or collectively use to support their activities.

Researchers within the MAS framework are calling for an explicit recognition of the responsibilities of the environment, irrespective of the agents. Since there are lots of things in the world that are not inside the minds of the agents, there is a need to surpass the subjective view of MAS, where the environment is somehow just the sum of some data structures within agents, and to embrace an objective stance towards environment, enabling modellers to deal with systems from an external point of view of the agents [18:128].

The point is that the active character of the environment must be taken seriously: some of its processes change its 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) [17:36]. Moreover, 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 out at the same date, bankruptcy of that institution could be the unintended effect. Furthermore, taking into account environment opens new means to deal with indirect or mediated interaction, which [17:14] considers characterized by properties such as name uncoupling (interacting entities do not have to know one another explicitly), space uncoupling and time uncoupling (they do not have neither to be at the same place nor to coexist at the same time).

Trying to understand environment mediated interaction, stigmergy is another worth mentioning point. The term “stigmergy” captures the notion that, if multiple agents leave signs in a shared environment and their subsequent actions are determined by they sensing those signs, coordination within large populations is achievable by simple means, namely without any direct communication. Most common examples coming from insects and ant societies, stigmergy is usually associated with simple agents with severely bounded computational resources. Yet, Parunak, along with researchers talking of self-organisation emerging just from mere local interaction as a widespread phenomenon, even for more sophisticated agents, claims that stigmergy is pervasive also in human societies. ”It would be more difficult to show a functioning human institution that is not stigmergic, than it is to find examples of human stigmergy” [11:163].

Contributing the “cognitive stigmergy” notion, [13] converges on this view. The point is that, since the agents we work with have not just reactive, but also cognitive activities and can adapt and learn, there is a need to generalise from stigmergy to cognitive stigmergy. Now, cognitive stigmergy asks for more sophisticated environments, being “in general more articulated than a mere pheromone container”, where “the effects of agent actions on the environment are understood as signs”, and “hold a symbolic value” [13:127,132].

We have just mentioned a few examples of recent interesting developments on the role of environments for systems of multiple agents. But some difficulties associated with these developments are worth mentioning.

A difficulty that must be a concern for all systems with just software environments is raised, e.g., by [17]. Contrary, for example, to real robots systems evolving on physical environments, all aspects of a purely virtual environment (and of a purely virtual agent) must be modelled explicitly. This raises conceptual concerns related to the role of the modeller, and asks for a clarification of the very concept of environment. Because a computational environment that is part of a software system should not be confused with the environment with which the system interacts, the different levels and dynamics at stake must be made explicit.

That point is mentioned by [19]. Discussing the Human-Computer Interaction issue, the authors say: “the role of humans in multiagent systems can be very diverse. In some applications, humans can play the role of agents and interact (. . .) with the application environment” [19:21].

Another promising issue is raised by the same researchers, talking of a “reflective level”. Writing that “Such reflective interface enables cognitive agents to modify the functional behaviour of the environment”, and that the reflective level can be seen as “a means for self-organizing MAS” [19:11], they are opening new

frontiers for artificial collective systems, promising more careful attention to the real meaning of the ”autonomy” of the agents.

Our institutional environment approach could give some ways to deal with these difficulties. And, additionally, incorporate a factor with easily recognisable importance to human societies but usually forgotten in systems of multiple artificial agents. It is all about history and accumulation. Throughout the centuries, humans have been accumulating small modifications to myriads aspects of our physical and social world, not necessarily being aware of all them. In a wholly different attitude, designers of artificial systems pretend to be able to play gods and genesis anew each time they start modelling another version of their systems. Our institutional approach also intends to respond to that situation, giving place to history and accumulation within systems of multiple artificial agents.

In the next Section we present some global aspects of this institutional approach, so paving the way to their concrete application in Section 4.

3. INSTITUTIONAL ENVIRONMENTS

We have proposed Institutional Robotics [15] as a new approach to the design of multi-robot systems, mainly inspired by concepts from Institutional Economics, an alternative to mainstream neoclassical economic theory [7]. The Institutional Robotics approach intends to sophisticate the design of collectives of artificial agents by adding, to the currently popular emergentist view, the concepts of physically and socially bounded autonomy of cognitive agents, and deliberately set up coordination devices.

On the one hand, full autonomy is not attainable. Autonomous agents are not necessarily self-sufficient. Most of the time agents depend on resources and on other agents to achieve some of their goals. Dependences imply interests: world states that objectively favour the achievement of an agent’s goals are interests of that agent. Limited autonomy of agents comes from these dependences and interests relations [4].

On the other hand, collective order does not always emerge from individual decisions alone. A set of experiences within MAS, reported in [2], proved that, at least in some situations, merely emergent processes may lead to inefficient solutions to collective problems. Due to the absence of any opportunity for individuals to agree on a joint strategy, this is true even in some situations where the best for each individual is also the best for the collective. Thus, coordination devices deliberately set up by agents could be useful and must be considered. Still, this approach does not preclude emergence. Bounded rationality combines with bounded autonomy to give place to emergent phenomena: there are deliberate planned actions but they may produce unintended effects beyond reach of the agents’ understanding.

The Institutional Robotics approach endeavours to reflect these aspects taking institutions as decisive elements of the environment of multi-agent systems. Within this approach, the control system for a collective of artificial agents is a network of institutions. However, in this context, we adopt a broad concept of institution [15:600]: “Institutions are coordination artefacts and come in many forms: organizations, teams, hierarchies, conventions, norms, roles played by some robots, behavioural routines, stereotyped ways of sensing and interpret certain situations, material artefacts, some material organization of the world. A particular institution can be a composite of several institutional forms.” In the next section we further refine some concepts that are crucial to future implementation of this approach.

4. A NETWORK OF INSTITUTIONS AS THE CONTROL SYSTEM FOR A COLLECTIVE OF ARTIFICIAL AGENTS

4.1 A hypothesis on the main constitutive elements of the social order dynamics

The classic problem of the social sciences, the problem of social order or the micro-macro problem, is the question that introduces [6]: “How does the heterogeneous micro-world of individual behaviours generate the global macroscopic regularities of the society?”. Our institutional approach aims to contribute to a better understanding of that problem within systems of multiple artificial agents interacting with natural ones. Our strategy consists of putting together the main constitutive elements of the complex dynamics of institutional order, let them interact and let us interact with them, draw some lessons from the experiment, and test these lessons on new generations of experiments. Our tentative hypothesis is that the main constitutive elements of the social order dynamics to experiment with are as follows.

4.1.1 The powerful engine of the interactive workings of inner life and outer life mechanisms of the agent

Agents have built-in reactive behaviours, routines, and deliberative competences. Agents have partial models of themselves (they know some, but not all, of their internal mechanisms). Some of the internal mechanisms known by the robots can be accessed and modified by themselves. These elements are constitutive of the inner life of the agent.

The continuing functioning of any agent depends on some material conditions. Basic needs drive the activity of agents and lead to modifications of both physical and social world. How an agent interprets its world and the possibilities it affords depends on the physical and social world models the agent bears upon. An agent’s links to some, and not others, available institutions on its environment influence the world models it puts to use, thus biasing its behaviour. Beyond being influenced by its links to a subset of the existing institutions, the agent also is, at some extent, able to exert some influence on institutional mechanisms. However, autonomous agents do not transcribe institutional models without (slightly or not) modifying them. So, basic needs, fundamentally disposed by nature, have strong, even if indirect, interaction with social mechanisms like institutions. These elements are at the root of the dynamics we call “outer life of agents”.

The inner life of the agent has multifaceted effects at behavioural level, and thus on its participation in social interaction. The agent’s activities on its social and material environments interact intensively with its internal mechanisms. The joint workings of inner and outer life are of paramount importance for the emergence of complex collective phenomena. The diffuse frontier between nature and nurture is also captured by our notion of interaction between inner life and outer life of an agent.

4.1.2 Agents with reactive and deliberative mechanisms in a world with mental and material aspects

Let us, following a number of researchers (e.g., [10][16]), call coordination artefacts to those artefacts shaped for coordinating the agents' actions. Now, some interesting coordination artefacts are associated not only with physical but also with cognitive opportunities and constraints (deontic mediators, such as permissions and obligations). Recognizing all of those enables a single agent to act in a coordinated way: a driver approaching a roundabout is obliged, only by physical properties of the artefact, to slow down and go right or left to proceed; traffic regulations add something more indicating which direction all drivers have to choose not to crash with others. In another example, some rules (or other kinds of mental constructs) can be associated to a material object to implement some aspect of the collective order (a wall separating two countries is taken as a border; there are some doors in the wall to let robots cross the border; some regulations apply to crossing the border).

We can say that material objects are devices for institutions when they implement some aspect of the collective order. Notwithstanding, the boundaries between institutional and purely physical aspects of the world are not sharp. Consider a wall separating two buildings: it effectively implements a prohibition of visiting neighbours if the robots are not able to climb. However, if the wall is seen as just an element of the physical world, some robots gaining access to opposite building with newly acquired tools or physical capabilities will not be minded as a breach of a prohibition. Still, modifications of the material world creating new possibilities of interaction can become institutional issues. If the collective prefers to preserve the previous situation of separated buildings, the new capability of the robots to climb the wall could give place to new regulations.

This kind of artefacts, along with the coordination purposes they serve, illustrates how much could it be difficult to separate, either in conceptual or in practical terms, material from mental aspects of our world. That difficulty is closely related to our condition as complex agents combining reactive and deliberative ties, both to the physical and the social world.

4.1.3 Nobody is born alone in the wild. Not even artificial agents. And, at times, humans act as ancestors for artificial agents.

When a natural human agent comes into world, generations of ancestors have been shaping physical and social environments for centuries. Yet, the human agent can contribute with some modifications, some of which will last; some others will vanish sooner or later. The same happens with institutions for artificial collectives. When an artificial agent comes into existence, designers have already settled most contingencies that can determine its life. But, if it enjoys some kind of autonomy, it will also contribute to the continuing evolution of its world. The institutional environment at any point in the history of a collective is always a mix of inherited and newly adopted forms. So, the designers of an artificial collective must shape the first version of an institutional network. Thus, they play the role of predecessors for the artificial agents and (at least some aspects of) their environment. And, if we want to develop a better understanding of the interaction between human and artificial agents, designers must stay involved; say "as participative gods".

4.2 Definitions

Now, we will try to capture the tentative hypothesis stated above with a set of definitions designed to guide the modelling of institutional environments: node of the institutional network, institutional agent, and institutional network.

Departing from prevalent approaches (e.g., [14],[8]), we bring forward the following tentative informal definition: «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.» Building upon this, the main constituents of institutional environments will be defined by structured tuples.

Starting with the definition of "node of the institutional network", instead of with the definition of "institutional network", deserves an explanation. Since we are not usually able to reach an external viewpoint on complex societies, especially where we enjoy a participative status, a top down approach could prove unrealistic. From an epistemological standpoint, starting with some particular institutions, and then trying to broaden our understanding of the network, looks like a more modest but reliable strategy. Additionally, this approach better accommodates the existence of genuine emergent dynamics.

Moreover, we talk of "node of the institutional network", and not of "institution", because we don't know a principled way to get general clear-cut distinctions between an institution and a network of institutions. For example: the judicial system of a country must be seen as an institution or as a net of institutions (a net of courts of justice)?

Definition 1. A Node of the Institutional Network is a tuple < *ID, Rationale, Modifiers, Network, Institutional Building, History* > where:

ID = < *Label, Form* >

Label: Unique ID for this node of the institutional network.

Form: Generic form of this node (formal organisation, informal group, role, rule (law, norm, convention, right), behavioural routine, stereotyped way of sensing and interpret certain situations, material artefact, some material organisation of the world, a composite of several basic institutional forms). To each form corresponds a specific way of communicating to agents the expectations embedded on a specific node of the Institutional Network.

Rationale = < *Goals, Activities* >

Goals: Collective goal this institution is thought to be functional to.

Activities: Specific activities of the agents this node of the institutional network is supposed to serve to.

Modifiers = < Cognitive Modifiers, Praxic Modifiers >

Cognitive Modifiers = < Ideologies-P, Ideologies-S, Material Infrastructure for Cognitive Modifiers, Mental Infrastructure for Cognitive Modifiers >

Ideologies-P: ideologies about the physical world.

Ideologies-S: ideologies about the social world.

(Ideologies are partial world models provided by institutions, and so in principle shared by the subset of all agents with links to specific institutions. One and the same institution can provide several ideologies to agents. There is no consistency requirement associated to the set of ideologies provided by one and the same institution. Ideologies include partial ontological assumptions about some regions of the multi-agent system's world: entities, their properties, relations possibly holding among them.)

Material Infrastructure for Cognitive Modifiers: Material aspects of the institution that impact the cognitive mechanisms of the agents (for example, tools for augmented computational power - like calculator or computers, or tools for modified perception, like microscopes, telescopes, sensors for sound or light waves outside the range of natural equipment of the agents - where the access to those tools is not granted to every agent and depends on institutional appurtenance or institutional position).

Mental Infrastructure for Cognitive Modifiers: Mental aspects of the institution that impact the cognitive mechanisms of the agents (for example, concepts that apply some specific distinctions to organize some region of the perceptive space - where the access to those concepts is not granted to every agent and depends on institutional appurtenance or institutional position).

Praxic Modifiers = < Material Infrastructure for Praxic Modifiers, Mental Infrastructure for Praxic Modifiers, Enforcement >

Material Infrastructure for Praxic Modifiers: Material aspects of the institution that impact the action mechanisms of the agents (for example, physical objects functioning exclusively by means of its physical characteristics given the physical characteristics of the agents: a wall separating two buildings implements the prohibition of visiting neighbours if the robots are not able to climb it).

Mental Infrastructure for Praxic Modifiers: Mental aspects of the institution that impact the action mechanisms of the agents (e.g., a program to control a sequence of operations). Some infrastructures combine material and mental aspects (for example, a traffic sign is a physical object which functioning is due to a specific link to a mental construct: a traffic rule).

Enforcement: Mechanisms associated with this node of the institutional network specifically designed to prevent or to redress negative effects of violation of expected behaviour (examples are fines and reputation) and to reward observance (examples are prizes and advancement in rank or status). Enforcement mechanisms affect future acting possibilities of agents.

Network: Links to other nodes of the institutional network (the existence of a link, its nature).

Institutional Building = <Institutional Imagination, Co-operative Decision-making >

Institutional Imagination: Mechanisms designed to facilitate "thought experiments" about possible modifications to actual institutions, or even alternative institutions (agents could test alternatives without actually implement them). Results of Institutional Imagination (possibly fuelled by access to the *Institutional Memory* of the *Institutional Network*, and to the *Lineage & Accumulation* element of *History* of a *Node of the Institutional Network*) would eventually be put forward to Co-operative Decision-making mechanisms specific to this node of the institutional network.

Co-operative Decision-making: Mechanisms designed to implement collective deliberation about possible modifications to actual institutions, or about alternative institutions.

History = < Material Leftovers, Mental Leftovers, Lineage & Accumulation >

Material Leftovers: Material objects that once served some aspect of the institutional dynamics but have gotten disconnected from it. (Because the continuing existence of a material object can be uncoupled from the continuing existence of the institutional device it implements – e.g., the wall could be demolished without eliminating the border; the border can be eliminated without demolishing the wall – a material leftover of a discarded institution can last as an obstacle in the world.)

Mental Leftovers: Mental constructs that once served some aspect of the institutional dynamics but have gotten disconnected from it (for example: norms that once served a collective goal and persist notwithstanding the goal having been relinquished).

Lineage & Accumulation: Old versions of this node of the institutional network, saved as a list of cumulative modifications to the oldest known version.

Definition 2. An Institutional Agent is a tuple < *ID, Nature, Individual Links, Institutional Links, Ideas, Praxis* > where:

ID = < Name, Natural Group Name >

Name: Specific individual identification.

Natural Group Name: (for example) Family name, for humans.

Nature =<Relatives, Species, Basic Needs, Built-in Mechanisms>

Relatives: Names of the other members of the Natural Group.

Species: Human, Non-Human Animal, Robot, ...

Basic Needs: Material conditions for continuing functioning of the agent.

Built-in Mechanisms: Built-in perceptive and motor apparatus, reactive behaviours, routines and deliberative competences.

Individual Links: Names of other agents this agent can identify by their names.

Institutional Links: Nodes of the institutional network the agent is currently linked to.

Ideas = < Current Ideologies-P, Current Ideologies-S, Current Opinions, Models of the Self, Institutional Knowledge >

Current Ideologies-P: Ideologies-P the agent adheres to at present.

Current Ideologies-S: Ideologies-S the agent adheres to at present.

(Notwithstanding the fact that Institutional Links determine in principle which ideologies the agent adheres to, actually not all agents are fully aware or fully adhere to all ideologies proposed by the institutions they are linked to.)

Current Opinions: Opinions the agent currently holds. An "opinion" is an individual deviation from world models provided by institutions. By virtue of bearing an "opinion", as well as bearing an "ideology", the behaviour of an agent can be modified.

Models of the Self: Every agent know some, but not all, of their internal mechanisms (agents have partial models of themselves).

Institutional Knowledge: Knowledge the agent has about the Institutional Network.

Praxis = < Physical World Tools, Social World Tools, Self-Improvement Tools >

Physical World Tools: Tools enabling the agent to modify the material organisation of the physical world, and thus, the material infrastructure of the institutions (including, but not restricted to, physical tools: influencing other agents is a possible delegate way of modifying the physical world).

Social World Tools: Tools enabling the agent to modify the organisation of the social world.

Self-Improvement Tools: Some of the internal mechanisms known by the agents can be accessed and modified by themselves.

Definition 3. An Institutional Network is a tuple < *Nodes, Connections, Institutional Memory, Emergency Observatory, Participative Gods* > where:

Nodes: Currently active institutional nodes.

Connections: Known/explicit links between active nodes.

Institutional Memory: Incomplete repository of old/inactive institutions which can be used to feed Institutional Building mechanisms. Each old/inactive institution is saved as a list of cumulative modifications to the oldest known version.

Emergency Observatory: Available information about emergent collective phenomena within the multi-agent system which is under control of this Institutional Network.

Participative Gods =<Customer, Designer, Rationale, Ontology>

Customer: Who ordered this control system for a collective of artificial agents.

Designer: Who designed this control system for a collective of artificial agents.

Rationale = < Goals, Activities >

Goals: Goals Customer and Designer want this multi-agent system to be functional to.

Activities: Activities Customer and Designer want this multi-agent system to serve to.

Ontology: Ontological assumptions of the Customer and the Designer about the multi-agent system's world (entities, their properties, relations possibly holding among them), given the goals they (the Customer and the Designer) place on it (the system).

4.3 How basic dynamics are represented within the tuples structure

We have tried to capture our tentative hypothesis on the main constitutive elements of the social order dynamics (see 4.1. above) with definitions 1 to 3. The tuples structure expresses the complex interaction of some basic dynamics of the social life of artificial agents in interaction with human beings. We will now underline the main components of these dynamics within the tuples structure.

The agent modifies itself as it modifies its world in ways that are not always fully intentional and that cannot be completely anticipated. The dynamics of interaction between inner life and outer life of an agent (see 4.1.1. above) is mainly represented by interactions of the following elements:

Agent → Nature → Built-In Mechanisms.

Agent → Ideas → Models of the Self.

Agent → Praxis → Self-Improvement Tools.

Agent → Nature → Basic Needs.

Agent → Praxis → Physical World Tools, Social World Tools.

Node of the I. Network → Modifiers → Cognitive Modifiers → Ideologies-P, Ideologies-S.

Agent → Ideas → Current Ideologies-P, Current Ideologies-S, Current Opinions.

Node of the I. Network → Network, Institutional Building.

Physical and cognitive opportunities and constraints represented by artefacts in the environment, and sometimes recognized by the agents, combine with internal mechanisms of the agents to give rise to complex behavioural patterns. Thus, behaviour can be modulated by way of environmental or internal mechanisms which are partly modifiable by the agents themselves. The dynamics of intertwined reactive and deliberative mechanisms of agents in a world with mental and material aspects (see 4.1.2. above) is mainly represented by interactions of the following elements:

Node of the I. Network → Modifiers → Cognitive Modifiers → Material Infrastructure for Cognitive Modifiers, Mental Infrastructure for Cognitive Modifiers.

Node of the I. Network → Modifiers → Praxic Modifiers → Material Infrastructure for Praxic Modifiers, Mental Infrastructure for Praxic Modifiers.

Node of the I. Network → Institutional Building.

Node of the I. Network → History → Material Leftovers, Mental Leftovers.

Agent → Nature → Built-in Mechanisms.

Agent → Ideas → Models of the Self.

Autonomous agents, coming into existence in a world shaped by generations of predecessors or designers, can also contribute to the continuing evolution of their environment. The dynamics of inherited vs. newly adopted institutions (see 4.1.3. above) is mainly represented by interactions of the following elements:

Node of the I. Network → Institutional Building.

Node of the I. Network → History → Lineage & Accumulation.

Agent → Ideas → Institutional Knowledge.

Institutional Network → Institutional Memory.

Where human beings are designers and users of collectives of artificial agents, the understanding of interaction between human and artificial agents becomes part of the understanding of the artificial system. Modelling crucial aspects of the interaction between human and artificial agents within the control system of the collective can improve that understanding. The dynamics of human/artificial agents' relationships (see 4.1.3. above) is mainly represented by interactions of the following elements:

Node of the I. Network → Rationale.

Institutional Network → Participative Gods → Rationale, Ontology.

Node of the I. Network → Modifiers → Cognitive Modifiers → Ideologies-P, Ideologies-S.

Node of the I. Network → ID → Form.

(The latter element will ease comparisons between artificial institutions and characteristic institutions of the Customer and Designer environments, thus fuelling understanding of constraints imposed by goals/activities the multi-agent system is supposed to serve.)

Our notion of interaction between inherited and newly adopted institutional forms leaves room both for deliberately set up institutional mechanisms and for emergent aspects of institutional evolution, as represented by these elements of the tuples structure:

Node of the I. Network → Institutional Building.

Node of the I. Network → History → Material Leftovers, Mental Leftovers, Lineage & Accumulation.

5. CONCLUSIONS AND FUTURE WORK

We introduced a set of definitions designed to guide the modelling of institutional environments, as part of a strategy to control collectives of artificial embodied agents (e.g., multi-robot systems), with bounded rationality and bounded autonomy, by a network of institutions. Building upon an informal definition, the main constituents of institutional environments (nodes of the institutional network, institutional agents, and institutional networks) were defined by structured tuples. The social order dynamics results of interactions among the elements of the defined tuples.

It is clear for us that deeper work must be done to gain further insight on the relevance of the constituent elements and their interactions. This will be the subject of the next steps in our research. We are working on two scenarios of different levels of complexity in order to experiment with partial aspects of our concept. The simpler scenario consists of a set of roundabouts designed to regulate urban traffic, directly associated with traffic signs and framed in a more general way by a road code. The more complex scenario consists of a "search and rescue" operation, where heterogeneous cognitive robots must cooperate, both with other species of robots and with humans, on an unstructured landscape, aiming to help victims of some kind of disaster or emergency situation.

Once the required clarifications are achieved, the tuple definitions will act as prescriptions for an ontology to be used in the software programs we plan to design and implement, so as to control a collective of real robots and their environments, including the interaction among humans and robots. Such a demonstration will act as a proof of concept of the Institutional Robotics framework.

6. ACKNOWLEDGMENTS

The research of the first author is supported by Fundação para a Ciência e a Tecnologia (grant SFRH/BPD/35862/2007). This work was partially supported by Fundação para a Ciência e a Tecnologia (ISR/IST pluri-annual funding) through the POS Conhecimento Program that includes FEDER funds. We would like to thank Fausto Ferreira, Gonçalo Neto, and Matthijs Spaan for the fruitful discussions on various aspects of Institutional

Robotics. Moreover, we thank anonymous reviewers for constructive comments and suggestions.

7. REFERENCES

[1] Agre, P. 1997. Computation and Human Experience. Cambridge University Press, Cambridge.

[2] Castro Caldas, J., Coelho, H. 1999. The Origin of Institutions: socio-economic processes, choice, norms and conventions. In Journal of Artificial Societies and Social Simulation, 2:2 (http://jasss.soc.surrey.ac.uk/2/2/1.html).

[3] Clark, A. 1997. Being There: Putting Brain, Body, and the World Together Again. The MIT Press, Cambridge, MA.

[4] Conte, R., Castelfranchi, C. 1995. Cognitive and Social Action. The University College London Press, London.

[5] Dourish, P. 2001. Where the Action Is: The Foundations of Embodied Interaction. The MIT Press, Cambridge, MA.

[6] Epstein, J.M., Axtell, R. 1996. Growing Artificial Societies: Social Science from the Bottom Up. The Brookings Institution and the MIT Press, Washington, DC.

[7] Hodgson, G.M. 1988. Economics and Institutions: A Manifesto for a Modern Institutional Economics. Polity Press, Cambridge.

[8] Hodgson, G.M. 2006. What Are Institutions? In Journal of Economic Issues, 40:1, 1-25.

[9] Kirsh, D., Maglio, P. 1994. On distinguishing epistemic from pragmatic action. In Cognitive Science, 18, 513–549.

[10] Omicini, A., Ricci, A., Viroli, M., Castelfranchi, C., Tummolini, L. 2004. Coordination Artifacts: Environment-Based Coordination for Intelligent Agents. In Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1 (New York, NY, July 19 - 23, 2004). IEEE Computer Society, Washington, DC, 286-293.

[11] Parunak, H.v.D. 2006. A Survey of Environments and Mechanisms for Human-Human Stigmergy. In Environments for Multi-Agent Systems II, E4MAS 2005, Selected Revised and Invited Papers, Springer-Verlag, Berlin Heidelberg, 163–186.

[12] Petroski, H. 1992. The Evolution of Useful Things. Vintage, New York.

[13] Ricci, A., Ominici, A., Viroli, M., Gardelli, L., Oliva, E. 2007. Cognitive Stigmergy: Towards a Framework Based on Agents and Artifacts. In Environments for Multi-Agent Systems III, E4MAS 2006, Selected Revised and Invited Papers, Springer-Verlag, Berlin Heidelberg, 124–140.

[14] Searle, J.R. 2005. What is an institution? In Journal of Institutional Economics, 1:1, 1-22.

[15] Silva, P., Lima, P.U. 2007. Institutional Robotics. In F. Almeida e Costa et al. (Eds.): ECAL 2007, LNAI 4648, Springer-Verlag, Berlin Heidelberg, 595–604.

[16] Tummolini, L., Castelfranchi, C. 2006. The cognitive and behavioral mediation of institutions: Towards an account of institutional actions. In Cognitive Systems Research, 7:2-3, 307–323.

[17] Weyns, D., Parunak, H.v.D., Michel, F., Holvoet, T., Ferber, J. 2005. Environments for Multiagent Systems, State-of-the-art and Research Challenges. In Weyns, D., Parunak, H.v.D., Michel, F. (eds.): Proceedings of the 1st International Workshop on Environments for Multi-Agent Systems. Springer-Verlag, Berlin Heidelberg, 1–47.

[18] Weyns, D., Schumacher, M., Ricci, A., Viroli, M., Holvoet, T. 2005. Environments in Multiagent Systems. In The Knowledge Engineer Review, 20:2, 127-141.

[19] Weyns, D., Omicini, A., Odell, J. 2007. Environment as a first class abstraction in multiagent systems. In International Journal on Autonomous Agents and Multi-Agent Systems, 14:1, 5-30.