Institutions in Perspective
An Extended Abstract

Pablo Noriega, Carles Sierra

Artificial Intelligence Research Institute —IIIA,
Spanish Council for Scientific Research —CSIC,
08193 Bellaterra, Barcelona, Catalonia, Spain.
pablo@iiia.csic.es, sierra@iiia.csic.es

1. There are situations where individuals interact in ways that involve:
   - *Commitment.* Interactions usually involve some promises of future actions, payments and transference of property, or other forms of obligations among participants
   - *Delegation.* Participants may act in representation of someone else.
   - *Repetition.* The same type of interaction is performed repeatedly, possibly involving different individuals, usually involving different items, issues or concerns.
   - *Liability and Risk.* The achievement of the interaction involves some sort of interest or gain for participants, and usually some transaction costs as well. Consequently, there is some risk involved that may be allocated more or less explicitly to participants.

   These situations involve participants that are
   - Autonomous.
   - Heterogeneous. Having different goals, different rationales, different moral standings.
   - Independent. Act regardless of a shared loyalty, a common authority or previous agreement.
   - Not-benevolent. One cannot assume they will be moved by the common good or any altruistic aim or consideration. These participants are probably self-motivated, egoist and may even be willing to injure other participants if doing so yields any benefits to themselves.
   - Not-reliable. Likewise, they may act as if they are able and capable, but even unknowingly, they may fail to act in any expected way out of their own will.
   - Liable. Although the characteristics above may seem overwhelmingly pessimistic, one should realize that participants ought to be liable for the damage they inflict to others.
2. Such situations are not uncommon: Markets, medical services, armies are but ready examples of forms of collective problem solving, coordinated tasks, and communities in which participants interact under the type of features mentioned above. Many more exist and many have proven successful for dealing with their intended goals for a very long time.

3. In such situations, it is not uncommon to resort to a trusted third party whose aim is to make those interactions effective by establishing and enforcing conventions that standardize interactions, allocate risks, establish safeguards and guarantee that certain intended actions actually take place and unwanted situations are prevented. Such is the intuitive notion of an *institution*, as we normally use the term when we refer to the *institutional* character of markets, political organizations, religious communities or families. Similar intuitions underlie theoretical approaches to institutions such as the economic-theoretic, sociological, legal and psychological ones.
4. While having entities that facilitate commitment making among individuals have proven useful over time, those functions that institutions provide for human interactions become ever more pertinent when we consider the possibility of letting participants to be not only humans, but software agents as well.
5. If we grant that institutions serve to articulate agent interactions, one can then say that the crucial purpose of an institution is to facilitate, oversee and enforce commitment-making among participants in a repetitive situation. Further analysis of these issues is worth undertaking. At least five concerns underlie the commitment-making functionality of an institution.
   - *Manage the identity of participants.* Validate access to the institution of only those agents that qualify, and only to those activities they may be entitled to participate in. In the case of software agents, make sure that adequate management is made of delegated authority or capabilities, and avatars, clones, and other forms of invalid replication of identities are properly expressed and taken care of. In the case of mixed-environments where humans and software agents interact. distinguish between internal (or *staff*) agents, and external agents.
   - *Define and validate requirements on participant capabilities.* The institution should make explicit the different roles that may be played by participants, that is the type of actions and commitments that an individual external agent is entitled to perform and entailed to satisfy. Make explicit, also, the requirements in terms of capability (behavioural competence, expertise, legal authorization, availability of resources, etc.) that external agents need to satisfy in order to participate in the institutionally regulated interactions.
   - *Establish interaction conventions.* This function requires from the institution to make explicit the intended meaning of all communications exchanged within the institution. In order to fix such meaning, all entities that intervene in the socially shared interactions need to be made explicit beforehand, as well as the conventions that determine what communications are needed, when and who can exchange them. Special attention has to be paid to those interactions that are repetitive, thus defining a protocol or adequate guidelines for the effective execution of those actions and their intended effects, exceptions and corrective procedures. Naturally, the choice of interaction conventions should take into account value considerations, usually transparency, accountability and efficiency

are paradigmatic. Issues such as privacy, transaction costs, information asymmetries and the degree of integration of anciliary interactions ought to be taken into account for these purposes.

- *Facilitate effective interactions.* The institution has to see to it that participating agents can achieve their intended goals within the institution, and ideally in a better way than they could by themselves without resorting to the institution. Hence, the role played by the institution should result in better trust and accountability for the agent interactions, should make encounters more likely, encounters more successful, and outcomes more productive for all participants. These aims can be achieved through different means, namely by the institution imposing regularity on the interactions, making the enactment of the interactions known and foreseeable, facilitating the presence of apt participants, and enforcing interaction conventions that are efficient for achieving the intended goals of all participants. In this respect, it should be noted that, at least in principle, the institution should make an effort to stay neutral with respect to the interest of any one of the participants, or giving any undue advantages to any type of participant.
- *Enforce satisfaction of commitments.* First through the design and implementation of mechanisms and devices that achieve adequate record keeping, risk allocation, safeguards, and liability underwriting among all participants. Next, by making an explicit standardisation of protocols, pre-conditions and post-conditions that govern commitment making within the institution. Finally, through the implementation of appropriate enforcement, corrective or compensatory mechanisms that may range from staff agents, supervisory devices, corrective measures, pre-epmtive safeguards, fines, insurance, bonds or guarantees, etc.

6. We have proposed to address those functions through the idea of *Electronic Institutions.*
7. We have defined an Electronic Institution as an entity that has three main components:
   - *Dialogical Framework.* Where most ontologic aspects of the institution are addressed: the language used to communicate and the intended meanings of illocutions, terms and entities that may be invoked in those communications *within the institution.*
   - *Performative Structure.* The conventions that establish the flow of interactions and the intended social consequences of the actions that take place within the institution.
   - *Norms for Individual Behaviour.* The conventions to which individual agents are subject to while acting within the institution. These conventions address the preconditions that need to be satisfied by a given participant in order to establish a commitment, and the effects such commitments may have on the individual's existing commitments and ulterior behaviour.

8. We have developed and tested these ideas in a number of projects (e.g. Fishmarket, SMASH, MASFIT, SLIE, ISLANDER). And they seem to be fruitful.

We will expand on these ideas in our talk. We would also like to invite interested readers to look into the following references that expand on the ideas mentioned in this extended abstract [1,2,4,3,6,5].

References

1. E-Institutor URL . http://e-institutor.iiia.csic.es.
2. P. Noriega. *Agent Mediated Auctions: The Fishmarket Metaphor.* Number 8 in Monografies de l'IIIA. IIIA-CSIC, 1999.
3. P. Noriega and C. Sierra. Auctions and multi-agent systems. In Matthias Klusch, editor, *Intelligent Information Agents*, pages 153–175. Springer, 1999.
4. J. A. Rodriguez-Aguilar. *On the design and construction of Agent-mediated Institutions.* Number 14 in Monografies de l'IIIA. IIIA-CSIC, 2002.
5. C. Sierra and P. Noriega. Agent-mediated interaction. from auctions to negotiation and argumentation. In Mark d'Inverno, Michael Luck, Michael Fisher, and Chris Preist, editors, *Foundations and Applications of Multi-Agent Systems: UK-MAS 1996-2000.* Springer, in press.
6. Carles Sierra, N. R. Jennings, Pablo Noriega, and Simon Parson. A framework for argumentation-based negotiation. In *Proceedings of the 4th International Workshop on Agent Theories, Architectures and Languages (ATAL-97)*, 1997.