Institutionalization through Reciprocal Habitualization and Typification

Eric Baumer[1] and Bill Tomlinson[2]

University of California, Irvine, Department of Informatics, 127B Computer Science Trailer, Irvine, CA 92697-3430
ebaumer@uci.edu
University of California, Irvine, Department of Informatics, Department of Drama, and Arts Computation Engineering (ACE) Program, Computer Science Bldg, Room 403A, Irvine, CA 92697-3440
wmt@uci.edu

When constructing multiagent systems, the designer may approach the system as a collection of individuals or may view the entire system as a whole. In addition to these approaches, it may be beneficial to consider the interactions between the individuals and the whole. Borrowing ideas from the notion of social construction and building on previous work in synthetic social construction, this paper presents a framework wherein autonomous agents engage in a dialectic relationship with the society of agents around them. In this framework, agents recognize patterns of social activity in their societies, group such patterns into institutions, and form computational representations of those institutions. The paper presents a design framework describing this method of institutionalization, some implementation suggestions, and a discussion of possible applications.

Introduction

There are many approaches to developing multiagent systems. Agent-oriented software engineering takes a top-down approach, decomposing the problem domain into components that can be addressed by individual agents or groups thereof [12]. In artificial life, a bottom-up approaches focuses on overall properties emerging from the interactions of individual agents [16]. It may be beneficial to focus not only on the individual the whole, but rather on the interactions between the individual and the whole. Social psychology has taken a variety of perspectives on the individual's relationship to the society [5, 18]. As an example of applying these different approaches, Dillenbourg, et al. [7] present a description of how changing the unit of study along the spectrum of individual to group changes the nature of educational research. This paper presents a technique for applying some of these concepts to multiagent systems, allowing the agents in the system to form computational representations of institutionalized actions.

Social psychology examines the nature of the individual's relationship to its society. Different theorists have taken a number of different approaches in trying to describe the nature of this relationship between self and society. Mead argues that the self has no meaning outside of society, that the self only exists with any meaning as a response to and interaction with the generalized other [18]. Cooley introduced the notion that the self is reflected in society in what he called the reflexive self or the looking-glass self, arguing that our self-concept comes only from the impressions we get about ourselves based on others' actions toward us [5]. Goffman posits that our social form is based on the manner in which we present ourselves [10]. Berger and Luckmann argue that the individual exists in dialectic relationship with society, and through this dialectic exchange the individual defines his or her own social reality at the same time that society is defining the individual [3]. More recent work synthesizes some of these earlier approaches. For example, Tice and Wallace present a case that even the reflected self is distorted by the lens of self-presentation, so that we see ourselves not as others actually see us but as we think others see us [22]. Although the specifics of these approaches differ, they all place the individual in an exchange of mutual sense making with society, where the society and the individual are simultaneously trying to understand one another.

This paper focuses on the work of Berger and Luckmann, specifically the social construction of reality [3]. This approach places the individual in a dialectic relationship with society in which the individual experiences society both as an objective and a subjective reality. With society as an objective reality, the individual experiences society as a unified whole, an external entity that dictates the individual's social reality. With society as a subjective reality, the individual internalizes society, making sense of it by relating it to his or her own subjective experiences, and then acts on that subjective interpretation. By experiencing society through this dual nature, individuals can reciprocally typify patterns of each other's actions and interactions to form institutions, which describe the behavioral patterns present in that society. This concept is related to some previous work that allows synthetic characters to form context-specific emotional memories [23]. The present framework, however, places those memories in a broader institutional context rather than the context of interaction with a specific individual. There has also been previous work has been done in the area of electronic institutions [9]. This work uses electronic institutions as a way to specify the interactions and commitments between different agents in an open environment [6, 21, 24]. However, all these approaches focus on using institutions as a method of specifying agent behavior to assist in the creation of multiagent systems. The work described here uses institutions to recognize and reason about emergent patterns of actions in multiagent systems.

This paper presents a framework for institutionalization that applies these concepts to multiagent systems, provides some aspects of an implementation, describes difficulties and limitations that might arise in that implementation, and discusses possible applications in domains such as human-machine interaction, societal or behavioral studies, and analysis of multiagent systems.

Social Construction

As stated above, Berger and Luckmann's social construction approach places the individual in a dialectic relationship with society [3]. In this dialectic relationship, the individual experiences society both as an external, objective reality and as an internal, subjective reality. Through these multiple experiences of society, society acts to define the individual, while the individual simultaneously acts to define society. "Society is a human product. Society is an objective reality. Man is a societal product." (p. 61) [3]

It is important to note that these three moments of definition are indeed three distinct moments, but they are also three simultaneous and cyclical events. The events of the dialectic relationship between the individual and society are continuous and coincident. This work focuses on the individual's experience of the society both as objective reality and as subjective reality in the process of forming institutions.

Society as Objective Reality

One component of the individual experience is that of society as objective reality, where in the individual experiences society as an external entity. Society dictates certain norms and patterns of behavior in the form of institutions, to which the individual must subscribe. "The institution posits that actions of type X will be performed by actors of type X." (p. 54) [3] The institution serve to dictate which types of actors take what types of actions. However, it also functions to give context to those prescriptions, so that actions of type X are performed by actors of type X in the context of other types of actors performing other types of actions. These patterns of actions form institutions that dictate to the individual the patterns of action that are to be taken within the society.

Consider the example of a courtroom trial. One would not expect to see the bailiff overseeing the court atop the bench, the lawyers sitting in the jury box listening, and the judge transcribing the events. The institution of a trial dictates different roles for different types of actors in the context of a trial. By knowing about an institution, actors in a situation know how to behave, and an individual can make sense of a situation. This leads to the individual forming its own impression of the scene.

Society as Subjective Reality

The other component of individual experience is that of society as subjective reality, wherein the individual goes through a process of sense making about the social occurrences around him or her. The individual attempts to understand societal activities based on his or her previous experiences, arriving at his or her own subjective interpretation of society. The individual then acts based on this subjective interpretation, thus influencing society. The individual's actions are then experienced as objective reality by other individuals.

When the individual takes actions based on his or her subjective reality, the individual is essentially presenting that subjective version of reality as an objective reality to other members of the society. The subjective reality is then confirmed or denied by the continued actions of others. Only when an individual's subjective reality is confirmed by the actions of others, when an individual has the same conceptualization of its society as others do, is that individual is a member of the society.

Framework

The main focus of this paper is to present a framework for institutionalization among autonomous agents based on some of these ideas from social construction. This framework builds on previous work in synthetic social construction [2], but rather than designing for a specific application this paper presents a general framework that can be applied to forming institutions in a wide range of multiagent systems. This section describes the components of the framework and how they interact to allow agents to form institutions via their interactions.

Habitualization

Habitualization[1] is the process by which actions that are frequently repeated with the same temporal relationships to one another are cast into a pattern. These patterns form groups of actions, so that an agent can reorganize its assessment of its own actions to treat these groups of actions individually or as a single unit. Any pattern of actions repeated at least once is habitualized to some degree, with those actions performed more frequently having a higher degree of habitualization. Habitualization is a non-social process by which the agent recognizes the habituality of its own actions and thus requires no interaction with other agents. This is not to say that an agent does not habitualize actions it takes in conjunction with other agents. Quite to the contrary, habitualization is more beneficial in the company of other agents, as it works hand-in-hand with typification, described below.

Typification

Typification is a process similar to habitualization, but rather than recognizes internally occurring patterns of behavior as in habitualization, typification recognizes externally occurring patterns enacted by other agents. Typification allows an agent to observe other agents and determine what types of actions are typical of what types of agents, and what types of agents typically perform what types of actions. Just as any pattern of actions repeated at least once is habitualized to some degree, any observed pattern of actions creates some degree of typification. Typification is most important in that reciprocal typification, which entails two agents typifying each other's actions, combined with habitualization leads to the formation of institutions.

Institutionalization

In the context of multiagent systems, an institution is a computational representation of a certain behavioral pattern implicitly agreed upon through the actions and interac-

[1] This is not to be confused with habituation, wherein repeated exposure to a specific stimulus decreases the degree or frequency of the associated response.

tions of the members of a society to which the institution applies. Institutions are formed by means of institutionalization.

Institutionalization is the process by which multiple agents habitualized their own actions and reciprocally typify each other's actions. Agent A habitualizes its own actions and typifies agent B's actions, while agent B is habitualizing its own actions and typifying agent A's. When agent A's typification of agent B's actions matches angent B's habitualization of its own action and vice versa, an institution is formed. Alternately, an agent can form an institution by observing and typifying the actions of other agents without actually participating in their interactions. The institution crystallizes the repeated patterns of actions composed of actions taken by the various agents involved in those patterns. It gives context to individual habitualizations and typifications by placing them in conjunction with other habitualizations and typifications. These institutional patterns of actions describe the types of interactions undertaken by the agents in a system.

Institutions serve a number of purposes. First, they allow agents to form computational representations of patterns of behavior. Those patterns may be analyzed by the agent itself or reported to human designers wishing to examine the nature of interactions in a system. Second, they provide a social learning mechanism. If a given agent observes other agents performing institutionalized patterns of actions, the agent may be able to learn those patterns of actions and participate in such patterns at a later time. Third, their predictive power can be used to inform action selection. Given a number of possible institutions that might fit the current events in which the agent is participating, the agent can choose its actions based on the institution that has the most beneficial outcome or leads other agents into a desirable course of action.

Institutions serve to abstract away from the differences between specific occurrences to generalize about all instances of that pattern of actions. One important way in which they do this is by defining roles, sets of actions taken within the pattern of some institution.

Roles

Roles abstract away the differences between individual actors that act in specific situations to generalize about all actors that might perform a set of actions in a given institution. An agent fulfilling a role in an institution is not only "a particular actor performing an action of type X, but" is demonstrating a "type-X action as being performable by *any* actor" (p. 72) [3] who fits the description of the role. This generalization does not pigeonhole any actor who fits this description into only performing type-X actions in this context. Rather, it represents a trend that applies to most but not necessarily all agents and situations. The specific description of a given role is the set of other roles a performer of the given role is likely to assume. For example, actors who perform role P in one institution are likely to also perform role Q in some other institution, so it is a reasonable conclusion that an agent performing role P will likely perform role Q, as well.

Implementation

Thus far, this paper has described a framework to allow agents in multiagent systems to form institutions based on habitualized and typified patterns of actions. This section proposes an implementation outline for that framework and discusses what will be required for such an implementation.

Pattern Recognition

The key component to habitualization and typification is the ability to recognize patterns of actions that are repeated frequently with the same or similar temporal relationships to one another. While there has been much research done on pattern recognition, many of the algorithms are complex and were developed for some specific application domain, such as image, face, and object recognition; audio processing; or speech recognition. Current pattern recognition algorithms are not amenable to the patterns of actions and interactions being recognized in this framework.

A number of approaches could be taken to the implementation of this component. One approach could combine some type of reinforcement learning [15], such as Q-learning [25], with various prediction methods to determine which events frequently occur within temporal proximity to what other events. Another approach could be to seed a pattern recognition algorithm with arrangements of certain types of events. In this way, the pattern recognizer does not have to start from scratch but rather has a platform of known patterns upon which to build other patterns. However, this method requires a specific set of pre-made patterns and is not as open-ended in the patterns of behavior it will recognize. A third approach could be to form a pattern from any set of events that occur within close temporal proximity. Once a pattern has been formed, the agent can determine whether or not the given pattern is an accurate or effective description for the events in its environment. Methods for making this determination are discussed below in the section on pruning.

Comparing Institutions

The ability to compare a series of currently occurring events to a known set of institutions acts as a key component to this framework. It is this ability that allows an agent to determine if some known institution is being enacted, what agents are performing which roles in this manifestation of the institution, and what the given agent's role in the institution might be. There are a number of aspects that will be used to accomplish this.

Below is described how institutions can be seen as a hierarchy, where large, complex institutions are described as compositions of smaller, simpler institutions. This very naturally gives rise to a tree-like structure for the representation of institutions, possibly in a method similar to the one described by Alspaugh [1]. Even institutions that are composed only of simple events may be seen to have a tree structure, where the institution is the root and the events that compose it are the leaves. Comparing two institutions could be accomplished in part be calculating a tree-to-tree distance algorithm [17] based on the tree-like representations of those institutions.

However, institutions are described by more than the events that compose them. They are also described by the temporal relationships between those events. Ordering the branches in the tree left-to-right based on the temporal order in which the events begin gives a starting point for the tree-to-tree distance calculation. However, the temporal relations between each pair of events must also be examined, since two institutions having the same tree structure may still be significantly different based on the temporal relations among their constituent events.

One more aspect must be compared to determine the degree to which a series of currently occurring events matches an institution. The actors performing the roles in the current manifestation of the institution must be compared to the characterizations of those roles. As stated above, a given role is characterized by the other roles that actors performing the given role are likely to take on. So for each role, there is a list of other roles, each with an associated probability indicating how likely a performer of the current role is to perform any one of a number of other roles. A similar list of roles and associated probabilities is maintained for each agent. When an agent is compared against a role it is performing, these probability matrices are compared, and their difference is the difference between the given agent and the role it is performing. However, there is a problem of scalability here, because each agent must carry around a probability matrix for each other agent of which it is aware. It may be possible for these matrices to decay over time, so that after periods with little interaction two agents no longer keep detailed information about each other's probability matrices. Similarly, the agent could perform pruning, so that rather than removing probability matrices at some given time interval they are removed based on the number and frequency of interactions with some other agent.

Another significant difficulty lies in determining to which institutions to compare the currently occurring events. Since this comparison will be a complex and likely computationally expensive process, minimizing the actual number of full comparisons will help keep the system scalable. One way to do this would be a heuristic determining which institutions to check based on their frequency of occurrence. The most commonly occurring institutions will be checked first, because the time spent checking them will yield a match much more often than the rarely occurring institutions. Along these lines, the heuristic could also incorporate a starting sequence, a series of events that always occurs at the beginning of a manifestation of the institution. Comparing currently occurring events against this starting sequence would be far less computationally intense than comparing the entire institution and would allow the full gamut of institutions to be checked rather than just the most frequently occurring ones.

Hierarchical Structure

Once a series of events has been recognized as a manifestation of an institution, those events can either be treated individually or as a whole. If treated as a whole, this manifestation of an institution can be seen as a single event, one that could possibly combine with others to form another higher-level institution. In this way, large, complex institutions can be formed from combinations of shorter, simpler institutions.

It may also be beneficial to organize not only institutions but individual events into a classification hierarchy. For example, running, walking, swimming, and crawling are all types of locomotion. In certain institutions, the method of locomotion should not matter as much as in others. In an institution such as the transportation of materials, it does not matter if an agent is using bipedal walking, caterpillar treads, or wheels, just that it is moving from one place to another while carrying some material. Using a super-event from which all these other sub-events inherit would allow any of the sub-events to be used as an instance of the super-event. This enhances the flexibility of institutions and reduces the total number of institutions needed, since a single institution can describe an action that involves any type of locomotion rather than needing a separate institution for each type.

Pruning

In this framework, pruning is analogous to the maintenance of subjective reality, wherein the individual verifies his or her subjective reality by acting on it and then determining if his or her actions are comprehensible to others. Institutions are monitored to determine how accurate they are as predictors and how frequently they are manifest. Institutions that are inaccurate predictors or infrequently manifest are either removed from the agent's collection or decreased in the frequency with which they are checked. However, this creates a possible difficulty for special purpose institutions. For example, the institution of a wedding occurs very infrequently, and so we do not walk around on a daily basis trying to understand our interactions in terms of a wedding. However, when we see a wedding, whether expected or unexpected, we have little trouble recognizing it as such. This example again brings up some of the challenges in selecting which institutions to compare against the currently occurring series of events.

Limitations and Difficulties

Some of the limitations and possible difficulties in this institutionalization framework were mentioned in the implementation section above. This section discusses two specific problems that may arise in the framework's implementation and suggests some possible approaches to addressing these difficulties.

Bootstrapping

One of the features listed above was that of a hierarchical structure, that is, that institutions can be composed of other institutions. At the lowest level, institutions are composed of single simple events. However, if an agent is not aware of any institutional structures, only of simple events, how does it bootstrap from simple events into basic institutions?

One approach to solving this, mentioned above, is to seed the agent with some basic institutions, on top of which the agent can construct other institutions. Using this approach limits the institutions that an agent can recognize to those that are based on institutions foreseeable by the designer. However, it gives more control and direction to the type of institutions that will be formed.

Another possible method would be giving the agent a specific list of all the possible simple events in its environment and using a pattern recognition algorithm to determine when these events occur repeatedly with the same temporal relationships to one another. As mentioned above, this would require complex and flexible pattern recognition, but it would allow for more variability in the sorts of institutions that an agent could recognize. For example, events A, B, and C might frequently occur in close temporal proximity and always with the same temporal relations to one another, but if the agent is only looking for patterns of events A and C as dictated by the designer, the agent would not be able to recognize this new pattern of events and would not learn the institution that might lie behind those events.

Both of the above approaches are possible solutions to the problem of bootstrapping if the agent is in an environment in which simple events occur in a discrete and recognizable way, such as market simulations, auction environments, or operations other than war scenarios. However, what occurs if an agent does not even have knowledge of what events are possible in its environment? In a simulation or virtual environment, it is possible for an agent to have full knowledge about what events are possible within that environment before actually interacting with it. Work is being done on activity recognition in open physical environments but has not yet been perfected in a general purpose form. Further details of this problem are discussed below.

Granularity

In very constrained environments, it would be possible to give an agent a specification of what comprises an event. However, in an open environment, this is significantly more complex. Consider an agent on a desktop computer observing the actions of the user. For simplicity's sake, suppose that the only information about the user available to the agent is the input from the mouse and keyboard. In this case, what constitutes a single event? One press of a key? One click of the mouse? Any continuous mouse movement in the same direction? Any set of inputs received within a specific temporal proximity? Figuring out exactly what comprises an event is a non-trivial matter.

Here it is possible to see the utility of having institutions that are composed of other institutions. In this example, we can allow simple repeated patterns of inputs to be learned as institutions. It may not make immediate sense to classify such activities as institutions, e.g., the institution of double clicking. However, groupings of these low-level institutions may be grouped into higher-level patterns, such as opening a file, cutting-and-pasting, or turning off the machine. In this way, institutions could be developed from the ground up. Such an approach could be applied to agents interacting with users in the physical world. Basic sensory-motor level actions could be used to form basic institutions, which would then comprise higher-level institutions that deal with interactions with users or aspects of the world.

Discussion

Up to this point, this paper has presented a framework for applying some of the concepts of social construction to multiagent systems, allowing them to form institutions in an adaptive and non-predetermined way. This section will discuss aspects of how this framework relates to other aspects of autonomous agents research.

Many multi-layered architectures have been proposed for the design of intelligent systems. In at least two of these [19] [20], there is a reflective layer that monitors the activities of the sensory-motor and deliberative layers. It is in this reflective layer that the above described framework would be implemented. As is apparent, the framework is not designed for motor-level action control, but rather to monitor actions and interactions in order to form patterns describing those interactions. Once such patterns are formed, they may be referenced by the sensory-motor or deliberative to assist in guiding future actions.

Another interesting question is that of embodiment or embodied action, which has application not only to autonomous agents but also to the broader fields of artificial intelligence, human-computer interaction, and many other disciplines. Traditionally, embodiment has been used to describe the way an agent is situated in its physical environment [4]. In recent years, embodied action has come to incorporate not only an agent's physical environment but also its social setting and how these aspects of its environment change over time [8]. While there has been a good deal of work in the field of artificial intelligence on embodiment in the physical sense, there has not been as much work focusing on embodiment in the social sense. The framework brings the concept of embodied interaction to bear on multiagent systems, giving agents a means to represent and reason about their social environment and the changes occurring therein.

There are also several possible approaches in deploying this framework, depending on the purpose of the system. The framework was designed to be deployed in such a way that each agent in a system has its own set of institutions based on its subjective experiences of interacting with other agents. These institutions could be compared to determine if all agents in an interaction approach that interaction in the same way. Another possibility is setting up an observational system independent from all the agents' interactions. This observer would determine what institutions are present in a system without actually interacting with it. However, this would require some modification, because the observer is never required to act based on the institutions it forms and thus has no way of verifying the validity of these institutions. An alternate verification here could be checking how well the observer's institutions predict actions of other agents, then modifying the institutions based on the accuracy or inaccuracy of these predictions.

Applications

There are several possible ways in which institutionalization could be applied to multiagent systems. These include strategy negotiation, so that negotiated strategies do not have to be constantly renegotiated; determining societal norms and then finding behavioral outliers in societies; and user modeling to discover and facilitate commons usage patterns. Three possible applications are discussed in detail below.

Human-Machine Heterogeneous Systems

One of the main difficulties in deploying systems composed both of human and machine agents in the coordination of the two. This framework could provide a method for machine agents to learn patterns of interaction as institutions, facilitating interaction with human agents. Another similar application is human control of systems composed of heterogeneous agents. If the agents are built with specific tasks in mind, delegating tasks to them is not necessarily a complex task. For example, miner robots are given the task of mining some ore; courier robots are given the task of moving the ore; etc. However, if the system is composed of heterogeneous, general purpose agents, this task becomes more difficult. A human operator must sort through the different types of agents, or worse examine each individual agent, determine which agents are suited for a particular task, and assign that task to all the suitable agents. Using the framework described here, a different approach might be taken. The human operator could take control of one general-purpose agent, one that he or she knows is fit for a certain task, and perform the given task. During the performance of the task, the agent's motor-level functions are controlled by the human operator, but its reflective layer is still monitoring its own actions, forming institutions that describe the tasks being performed by the operator. Once the task has been completed, control is returned to the agent, and the institutions learned by the reflective layer dictate the proper course of action. Furthermore, the computational representation of this institution could be supplied to all the agents in the system. Because the institution describes the actor in terms of an abstracted role, any agent suited for that role will perform it. Thus, the agents automatically determine which of them are fit to perform a certain task.

Agent Co-learners

Some work has already been done exploring the effect of artificial co-learners on the performance of human students [14]. However, the agents in these studies act in predetermined ways and do not adjust to the behavior of human users or other agents. The behavior and performance of co-learner agents has been shown to have a significant effect on both the performance of human learners and their attitudes toward their co-learners [14]. A co-learner could adapt to the human user's patterns of behavior, both attitudinal and performance-wise, so that the agent's behavior would be relative to that of the human user.

Analyzing Emergent Behavior

Emergent behavior in complex systems has been the topic of some discussion [11, 13, 26]. However, the presence and recognition of emergent properties are often subjective; it is up to the observer, often the system builder, to recognize when a pattern of behavior has emerged as a consequence of lower-level interactions. The framework described here could provide a method for determining when an emergent pattern has formed. Furthermore, treating lower-level actions and interactions as the basis for institutions, institutions formed in this framework could be analyzed to determine how and possibly why some lower-level interactions caused certain behavioral patterns to emerge.

Conclusion

This paper presents a framework that allows agents in a multiagent system to form institutions based on their behavioral patterns and build computational models of those thereof. Institutionalization, the process of forming institutions, results from reciprocal typification and habitualization among a group of agents. Within these institutions, the patterns of actions performed by agents are described in terms of roles, which abstract away the differences between individual agents to describe which types of agents perform which types of actions and which actions are performed by what agents. Thus, the institutions form the basis for an agent's model of the interactions occurring in its environment. This institutionalization framework has broad possible applications, ranging from human-machine interaction, to user modeling, to analysis of multiagent systems.

References

[1] Alspaugh, T. Temporally Expressive Scenarios in ScenarioML *Institute for Software Research Technical Report UCI-ISR-05-06*, University of California, Irvine, 2005.

[2] Baumer, E. and Tomlinson, B., Synthetic Social Construction for Autonomous Characters. in *AAAI Workshop on Modular Construction of Human-Like Intelligence*, (Pittsburgh, PA, 2005).

[3] Berger, P.L. and Luckmann, T. *The Social Construction of Reality: A Treatise on the Sociology of Knowledge*. Irvington Publishers, Inc., New York, 1966.

[4] Brooks, R.A. Elephants Don't Play Chess. *Robotics and Autonomous Systems*, *6*.

[5] Cooley, C.H. *Human Nature and the Social Order*. Charles Scribner's Sons, New York, 1902.

[6] Cuni, G., Esteva, M., Garcia, P., Puertas, E., Sierra, C. and Solchaga, T., MASFIT: Multi-Agest System for Fish Trading. in *16th European Conference on Artificial Intelligence (ECAI 2004)*, (Valencia, Spain, 2004), 710-714.

[7] Dillenbourg, P., Baker, M., Blaye, A. and C, O.M. The evolution of research on collaborative learning. in Reimann, P. and Spada, H. eds. *Learning in Humans and Machine: Towards an interdisciplinary learning science*, Elsevier, Oxford, 1996, 189-211.

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

[9] Esteva, M. Electronic Institutions: from specification to development *Artificial Intelligence Research Institute (IIIA)*, Technical University of Catalonia (UPC), Barcelona, 2003.

[10] Goffman, E. *The Presentation of Self in Everyday Life*. Doubleday, New York, 1959.

[11] Holland, J. *Emergence: from chaos to order*. Oxford University Press, Oxford, 1998.

[12] Jennings, N.R. On Agent-Based Software Engineering. *Artificial Intelligence*, *117* (2). 277-296.

[13] Johnson, S. *Emergence: the connected lives of ants, brains, cities, and software*. Simon & Schuster, New York, 2002.

[14] Ju, W., Nickell, S., Eng, K. and Nass, C. Influence of colearner agent gehavior on learner performance and attitudes *CHI '05 extended abstracts on Human factors in computing systems*, ACM Press, Portland, OR, USA, 2005.

[15] Kaelbling, L.P., Littman, M.L. and Moore, A.W. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, *4*. 237-285.

[16] Langton, C. *Artificial Life: an overview*. MIT Press, Cambridge, MA, 1995.

[17] Lu, S.-Y. A Tree-to-Tree Distance and Its Application to Cluster Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, *1* (2). 219-224.

[18] Mead, G.H. *Mind, Self, and Society from the Perspective of a Social Behaviorist*. University of Chicago Press, Chicago, 1934.

[19] Murphy, R., Lisetti, C.L., Tardif, R., Irish, L. and Gage, A. Emotion-Based Control of Cooperating Heterogeneous Robots. *IEEE Transactions on Robotics and Automation*, *18* (5). 744-757.

[20] Norman, D., Ortony, A. and Russell, D.M. Affect and machine design: Lessons for the development of autonomous machines. *IBM Systems Journal*, *42*. 38-44.

[21] Sierra, C., Rodriguez-Aguilar, J.A., Noriega, P., Esteva, M. and Arcos, J.L. Engineering multi-agent systems as electronic institutions. *European Journal for the Informatics Professional*, *4*.

[22] Tice, D.M. and Wallace, H.M. The Reflected Self: Creating Yourself as (You Think) Others See You. in Leary, M.R. and Tangney, J.P. eds. *Handbook of Self and Identity*, The Guilford Press, New Your, 2003.

[23] Tomlinson, B. and Blumberg, B. AlphaWolf: Social Learning, Emotion and Development in Autonomous Virtual Agents. *First GSFC/JPL Workshop on Radical Agent Concepts. NASA Goddard Space Flight Center, Greenbelt, MD. January 2002. Published in Lecture Notes in Computer Science, Springer Verlag*, *2564*. 35--45.

[24] Vasconcelos, Esteva, M., Sierra, C. and Rodgriguez-Aguilar, J.A., Verifying norm consistency in electronic institutions. in *AAAI-04 Workshop on Agent Organizations: Theory and Practice*, (San Jose, CA, 2004), 8-14.

[25] Watkins, C.J.C.H. and Dayan, P. Technical Note: Q-Learning. *Machine Learning*, *8* (3-4). 279-292.

[26] Wolfram, S. *A New Kind of Science*. Wolfram Media, Champaign, IL, 2002.