Coupling Human and Non Human Agents

Goran Trajkovski, Samuel Collins, James Braman and Mark Goldberg

Cognitive Agency and Robotics Laboratory
Towson University, Towson, MD 21252, USA
{gtrajkovski, scollins, jbrama1, mgoldb2} @towson.edu

Abstract

Using methods from computer science, anthropology and cultural studies, we present the principles of Izbushka, an intelligent agent being constructed with the goal of understanding interactions in heterogeneous multiagent systems, and observing emergent phenomena in such agent societies.

Introduction

It is a common perception that in human-computer interactions, humans perceive the agents as extensions of their physical lives into virtual spaces. In the interaction process metaphors of physical spaces are commonly used, and the characteristics of social and cultural life that are simulated are known and static, rather than one of understanding the social and cultural life to be simulated. The attitude is not much different when we talk about human-agent interaction, where the agents are not necessarily in the cyberspace.

In this project we focus on human-agent interaction issues, from the perspective of observing a heterogeneous multiagent society, and emergent phenomena within it. Following John Johnston (2002:277), we seek to "investigate how certain human cognitive functions are always made possible by technical artifacts, and thus how what we consider to be human cognition cannot be considered outside of their shaping force and the environments they constitute."

Applying this to Human-Computer Interaction (HCI) means – for us – realizing the truth of the "cyborg" as an amalgam of human and machine in embodied contexts. We suggest that the "human" in Human-Computer Interaction represents a cultural variable and shifts along with changes in technologies, and that we should be designing interactive agents that not only simulate the human and human environments, but that they should be designing agents and environments that simulate the cyborg, itself simulating the computer. We believe that this approach may result in different ways of thinking about things like online education and social interaction of all kinds.

Rationale

We started developing an interactive environment that acts as a "0-context" tool where we might observe emergent interactions between human- and non-human agents in such a way that results are not simply artifacts of pre-existing metaphorical constructs of physical, social and cultural notions. By "emergence" we refer not to the more traditional understanding of the term that locates higher-order phenomena in the behavior of lower-order, individual components (as in rational choice theory), nor in some mystical process wherein research results are "inexplicable" (a charge sometimes leveled at the sciences of complexity), but to the possibility that we may find new, generative metaphors for understanding online, multiagent systems in the course of our analysis, i.e., that properly emergent phenomena, neither reducible to individual elements nor higher-order collectivities, presents the possibility of the novel and unexpected. Likewise, while we recognize that no context can ever be at zero ("0") (there's always contexts, schema, syntax and so on), we would suggest that there are nevertheless ways to introduce the uncanny, what Walter Benjamin called the unheimlich (literally "un-home-like"). We propose placing human and non-human agents into interactions that have no immediately apparent goals beyond the self-organization of the human- non-human multiagent system itself with the ultimate goal of studying the representations produced by the interacting agents.

These sessions will result in bodies of quantitative data and qualitative data that help us understand: 1) ways that people and machines "learn" together; 2) the ways in which human- and non-human agents combine to form multiagent societies, both intra- and inter-culturally, and 3) the ways people make sense of computer interaction, the metaphors they use and the ways in which they characterize computer agency.

Interaction

The environment-agent Izbushka structurally couples with human agents, the goals of which will emerge in the course of human- and non-human interaction. Human agents interacting with Izbushka are first confronted by a blank screen where only the current position in the environment is shown. The human agent interacting in this environment may traverse the simulated space using only the arrow keys that represent the four cardinal directions: North, South, East and West. Through navigation, squares and patterns on the grid are displayed in proportion to the current level of context. Each type of square also has its own associated sound. The "goal" of human agents is to discern the goal of the simulation. Is it to traverse the environment using only a certain pattern of movements based on the condition of the environment? Or is it to avoid certain sounds and colors or is it to visit certain locations? Is "success" signaled by certain sounds, failure signaled by others? What human agents do not know is that Izbushka's goals (its "drives") depend upon the choices of human agents, i.e., the "goal" of the simulation depends upon the interaction of human and non-human agent and, additionally, may change in the space of the simulation. In this way, we hope to avoid the self-fulfilling prophecy of construing interactions as a metaphoric extension of "real" interaction. This research involves a series of related topics/questions, each addressed in a series of interrelated stages in our research design. These include:

1) The design issues (How can we build an autonomous, dynamic environment-agent);

2) Learning in a multiagent environment (How do people interact with information technologies in the process of learning? How do people socially organize in online environments? Are there cross-cultural differences in forms of social organization? Can people who do not share a common language interact together in an online environment? How?);

3) Conceptualization in humans (How do people conceptualize dynamic, online environments? Do these conceptualizations vary cross-culturally?)

4) Cultural models (What kinds of cultural models do people mobilize in their efforts to understand?). These are, in turn, addressed in each stage of our proposed research.

Preliminaries

Over the past decade, we have examined the artifacts of information society as 1) embedded in institutions and social relations, and 2) part of the cultural milieu in which they originate. For example, Collins (1998) argues that the "digital library" cannot be understood apart from labor-management conflict and organizational change at the Library of Congress in the 1990s, nor can ideas of "digital knowledge" be dissociated from cultural economies of information as somehow existing outside of both their containers and creators and, therefore, fungible commodities rather than epistemological processes. Computer simulations are likewise cultural artifacts. Games like SimCity (Collins 2004) may allow for infinite possibilities of play based on their novel employment of von Neumann's cellular automata, but still reflect (and instill) ideas of the city as a "growth machine" drawn from theories, experience and popular writings in the United States after World War II. Nevertheless, following anthropology of science in particular and STS (science and technology studies) in general, we believe that examining the artifacts of information society as inescapably cultural opens up possibilities for critical alternatives. One purpose of simulations may be to simply confirm reality "as-we-think-it-is," but other, more open-ended designs might gesture towards life "as-it-might-become" (Collins 1998; Helmreich 1998). This approach, we believe, turns away from technological determinism, where information technologies succeed one another towards an ultimate telos, to a more Bergsonian model of emergence, where future, information societies are unimaginable from the perspective of the present and where new technologies multiply possibilities for heterogeneity. In the past few years we have expanded our research into multiagent systems, because they allow us to focus on questions of emergent organization and social interaction. The research in the field of multiagent systems requires an integrated, not analytical, approach. And, accordingly, the majority of our research in this are has followed two directions: (a) theoretical and experimental analysis of the self-organizing mechanism of two or more interacting agents, and (b) the creation of distributed artifacts able to cope with complex tasks via collaboration and interaction.

The problem of learning has been a central notion in the AI theories of Agency for a long time. In the past several years our research team has established and studied IETAL (Interactivist-Expectative Theory on Agency and Learning), as a part of the growing trend in re-approaching two very different disciplines, robotics, on one side and developmental psychology, on the other. IETAL concentrates on exploring the concepts of learning in an autonomous agent, through interaction with the environment it inhabits. In the process the agent develops intrinsic models of its environment with a relevant emotional context for a given set of active drives. Most of the investigation in the past several years has been geared towards establishing a successful parameter base for a simulation of this paradigm with robots that are learning the environment using interaction, expectancies, and emotional contexts (Stojanov et al., 1997).

The notion of expectancy has a central role in this theory, and the agent, while being in the environment, anticipates the effects of its own actions in the world. In order to avoid any possible terminological confusion, we give the following explanation. An agent is said to be aware of the environment it inhabits if it can anticipate the results of his own actions. This means that, given some current percept p the agent can generate expectancies about the resulting percepts if it applies actions from its repertoire. After inhabiting some environment for certain time, an agent builds a network of such expectancy triplets percept-action-percept. This brings us to the second key concept in our theory of agency, namely the concept of agent environment interaction. As we see, the main problems are: first, to learn the graph and second, to know how to use it.

During its stay in the environment, the agent is interacting with the environment, and builds its intrinsic representation of the environment. When the, say, hunger drive is activated for the first time, the agent performs random walk during which expectancies are stored in the associative memory. The emotional contexts of these expectancies are neutral until food is sensed. Once this happens, current expectancy emotional context is set to positive value. This value is then exponentially decremented and propagated backwards following the recent percepts.

Every next future time the hunger drive is activated, the agent uses the context values of the expectancies to direct its actions. It chooses, of course, the action that will lead to expectancy with maximum context value. If all the expectancies for the current perceptual state are neutral, random walk is performed. Again, when food is sensed emotional contexts are adjusted in the previously described manner.

The way the agent perceives the environment is different when different drives are active. That means that it uses different instantiations, depending on the given set of drives.

In Trajkovski (2003) an experimental setup is described, where human subjects are situated in a simple virtual environment with unknown sensory-motor contingencies. This setup attempts to avoid the influence of the subjects. Previous knowledge all the way being ecologically plausible by letting subjects act and produce their stimuli. They press buttons of the computer keyboard and experience audio and/or visual sensations, and are required to produce a motor action. Actually, not knowingly, by pressing four buttons, they are moving in 2D environment, and are supposed to solve the navigation problem by avoiding obstacles and finding the place designated as a goal. Results prove that humans are capable of solving this problem even if they didn't have the phenomenological experience of being in a 2D maze. The problem is to find out what are human subjects learning and how this knowledge is represented. Based on the data from these experiments, a representation model could be derived, accounting for the problems with the classical models of representation (Brooks 1991; Stojanov and Bickhard 2004). This set-up is consistent with the implicit learning paradigm. However, in the majority of implicit learning experiments, human subjects are rather passive participants. The discussions, again, seem to focus on the way the final knowledge is represented: grammar rules, bigrams, trigrams etc., and less on the process and mechanisms of acquiring it. This is a consequence of the underlying substance ontology. With our original set-up we avoid these shortcomings by letting the subjects create their own stimuli.

In the research on the communication in the multiagent environments, we have followed the trend of renewed interest in the phenomenon of imitation. Imitation is far from a trivial phenomenon. It is a creative mapping of one's actions and their consequences onto self. Almost a decade ago Rizzolatti (1996) discovered the mirror neurons, neurons that fire when the subject is executing motor actions while watching somebody else doing them. IETAL (Stojanov et al 1997) serves as a base for the multiagent theory referred to as Multiagent Systems in Interactive Virtual Environments (MASIVE) (Trajkovski, 2003; Trajkovski, 2007), where imitation plays a key role in interagent communication. MASIVE, we have a) Established a mechanism of interagent communication in the multiagent environment, thus expanding IETAL in the domain of multiple agents societies with linguistic competence; b) Formalized the said methods via the means of the classical algebraic theories, and the fuzzy algebraic structures; c) Studied some of the emerging structures when interacting with the environment (the neighboring agents are considered a part of the environment. The agent in this system(s) has a special sensor for the other agents close to it in the environment. The drive to communicate is the top of its drive structure. Language is a key element for communication in MASIVE.

These categories are being built based on the inborn schema. The agent builds perceptual categories during its sojourn in the environment depending on the active drive of the agent. All perceptual categories that refer to a given drive compose the conceptual category for the drive. They model the trip to the satisfaction of a given drive, as well the object that satisfies it. From the perspective of agent's introspection, the concept makes it aware of the places where a certain drive can be satisfied. From the social perspective, when the concept is being fortified during the imitation conventions, the agent serves the environment it is in, and disseminate to the other agents information about the satisfaction of the drives of the fellow agents. Once we introduce a similarity measure for the concepts and observe their classes of equivalence, we could attribute a tag to the concept and talk about the phenomenon of protolexemes that would ultimately lead to the discussion of emergence of protolanguage in the system.

Izbushka

This research project is not about testing hypotheses. We are, rather, asking questions, and are extremely interested in the diversity of the differences their answers would reveal. There are several design criteria in designing Izbushka, as outlined below.

The environment is an interactive, reactive, hybrid (artificial + human coupled), intelligent agent composed of both emergent, "subsumption" architecture (emergent is not synonymous with subsumption) and a "top-down" hierarchy of drives. There are obstacles in the environment that invoke the agent to change the environment when certain conditions are met.

In a reactive fashion, the non-human environment-agent structurally "couples" to human agents as they navigate. It is "hybrid" for this reason. The 'spheres of influence' of human and non-human agents are being defined, as do the hierarchies between them in a given context.

The environment should be able to be static or dynamic, in order to measure the influence of the perceptual change on the learning process. The agent in this environment cannot see the whole ontology of Izbushka, he/she can only see a limited context of the locally distinct place where he/she are currently in Izbushka, and can build their cognitive map of the environment based on their perceptual/sensory resolution. To understand our context methodology, Figures 1 through 3 below show human perception at context 0, context 1 and context 2 respectively. Figure 4 shows the underlying environment in which the context levels hide from the human user. The current location of the human user in the environment is represented by the blank (white) center square. The Izbushka agent has the ability to see the entirety of the environment as the human user interacts within.

The choices of which drive Izbushka are influenced by the" moves" of human agents interacting with the environment. Due to the nature of the human drives (constant, periodic, acute and other types of drives), when several drives are active, considerations are being taken in order to develop a decision system that will determine whether two drives will work in parallel or in a sequence, and which one precedes in the quest through Izbushka.

Izbushka couples the context and actions of the human agent(s), and decides upon what to present the subject with next.

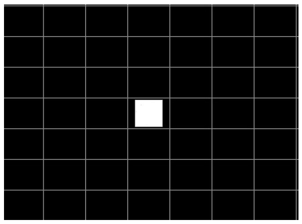

Figure 1: Context 0 Perception

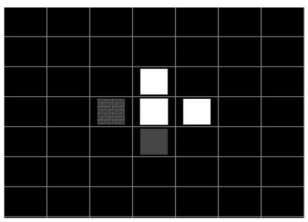

Figure 2: Context 1 Perception

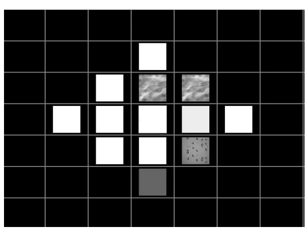

Figure 3: Context 2 Perception

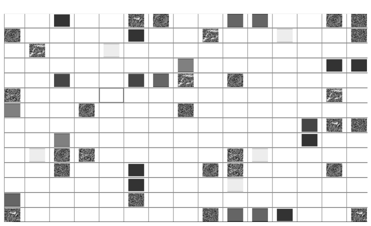

Figure 4: Underlying Sample Environment with obstacles

Izbushka learns based on the IETAL and MASIVE theories, outlined above. Based on the success of the subject in 'navigating' the environment that Izbushka is presenting him/her with, Izbushka will be building its database and changing its own actions for the current subject.

Its further actions will be determined based on the data it has collected. Izbushka records both sequences of actions and sequences of percepts, where the actions refer to keystrokes, and the percepts refer to the stimuli given to the subject when these actions were taken. Depending on whether the subject acts as Izbushka anticipates, its expectations change dynamically as surprises happen. Thus not only the subject aligns with the drive of Izbushka, but also Izbushka aligns with the subject's current drive.

During a session, records of interactions between the human and Izbushka (keystrokes, obstacles encountered, shifting environments) will be logged to a database. In the proposed research, we seek neither to reduce computer-mediated interaction individual psychology nor to cultural superstructures. Instead, we have adopted a heterogeneous series of recursive methodologies. In the spirit of Bateson's challenge, we endeavor to collect quantitative and qualitative data simultaneously and to relate these different bodies together. We will be gathering data relevant to the learning process in the human subject and the success of coupling with Izbushka. Thus the data we collect would be the keystrokes, inter-action time, and Izbushka's reaction/response. This data will give us an insight into the differences between the individuals, the teams, etc, measured by the number of moves, number of 'false' moves, and patterns developed. The Izbushka environment will heavily based on our agents developed in the IETAL and MASIVE studies.

This particular project relies strongly on emic categorizations. Hence it's important not just to the concepts people ultimately utilize, but also to look at their organization and frequency, and, finally, to check the validity in exit interviews and in journals.

After having identified hierarchies of terms relevant to interacting with Izbushka, we will conduct focus group interviews with volunteers and ask participants to record their "version" of events in a narrative journal entry. All of these will be analyzed with transcript from the discourse analysis phase above. According to Titscher (2000), conversation analysis "seeks to find those generative principles and procedures which participants use to produce the characteristic structure and order of a communicative situation." For us, this means:

1). Soliciting definitions of cover and included terms.

2). Triangulating preliminary findings on semantic relationships between terms.

3). Suggesting root metaphors generative of these terms. For example, much online interaction is structured by metaphorized, notional spaces. As Lakoff and Johnson (1980), Traweek (1988), Martin (1994), Helmreich (1998), Downey (1998) and others have shown the metaphors people use to relate experiences to each other facilitate certain kinds of practices and actions. Suggesting the cultural metaphors underlying participants' utterances indicates different orientations towards online interaction. In terms of our ultimate objectives, a variety of root metaphors represent resources for alternative approaches to online interactions based on cultural and social differences.

4). Identification of domains and structuring, root metaphors will allow us to make preliminary observations about the "cultural models." As Mukhopadhyay (2004:464) notes, cultural models "provide interpretive and information processing aids for creating meaning, organizing experience, think, feeling and acting." Derived from the work of Strauss and Quinn (1998), cultural models have been used to explain behaviors like African

American attitudes towards educational achievement (Fordham 1996), ways in which models of "romance" impact the career aspirations of women in the South (Holand and Eisenhart 1990). The power of the cultural models approach lies in its focus on practice: cultural models actively structure cognition and action, but, as Strauss and Quinn (1998) remind us, as "mediating devices" rather than as causally determinant superstructures. Instead, we think of them as one of a number of heterogeneous elements (i.e., the other levels of analysis in this study) through which Human-Computer Interaction emerges.

Conclusion

We see the significance of this project and its larger research context in contributing scientifically to several areas. On a more general level, the project results will have significant impact to the areas of: Cybersociology: Developing new methodologies and theories for the study of emergent, online social structure and for the study of interaction with online environments as social practice, Cyberanthropology: Developing models for the emergence of new cultural behaviors composed of networks of human and non-human agents with online environments, Multiagent systems: Simulating and observing emergent behaviors in MAS, studying the sociology of these systems, studying the communication, and the phenomenon of emergence of language, Theories of agency and learning: Building a realistic autonomous agent inhabiting uni- and multi-agent environments, Insights to cognitive science: Dynamics of learning, concepts in humans, concept formation, concept dissipation, contextual reasoning, Methodological advances: Izbushka as a tool will be a useful companion for data collection for various experiments, Developmental and cognitive robotics: Building a discrete analytical model from data collected from humans, readily implementable in robotic systems, etc.

More specific examples of impact include the following areas: Calibrating our existing interactivist-expectative model of agency, multiagency and learning, thus providing with a better cognitive model based on data and observations in human subject inhabiting simple virtual environments, Developing human-non-human multiagent experiment: Izbushka is an interface for human-machine interaction, Study of patterns and optimality in menu browsing and searching: The conclusions made in this project will have far reaching consequences in various application areas, especially in the process of design of menu/options/searching systems for optimal user usage, where the stress is not on contents or "what", but on "how", stressing the needs of an individual or a group), Applications (E-learning, E-commerce, M-Commerce, video gaming), and such.

References

Brooks, Rodney. Intelligence without representation. Artificial Intelligence 42 (1991): 139-159.

Collins, S. "Down and Out in Invisible Cities." Simcity: Mappando le Citta Virtuali, ed. by Matteo Bittanti. Ludalogica, 2004.

Collins, S. Knowledge Will Forever Govern Ignorance. Ph.D. Thesis, American University, 1998.

Johnston, John. "A Future for Autonomous Agents." Configurations 10 (2002): 473-516.

Rizzolatti, G., L. Fadiga, V. Gallese and L. Fogassi. "Premotor cortex and the recognition of motor actions." Cognitive Bran Research 3 (1996): 131-41.

Stojanov, G. and M. Birkhard. "Representation." Comments on Grush, BBS, 2004.

Stojanov, G., Bozinovski, S. and G. Trajkovski. "Interactionist-Expectative View on Agency and Learning." IMACS Journal for Mathematics and Computers in Simulation [Holland] 44 (1997).

Trajkovski, G. An Imitation-Based Approach to Modeling Homogenous Societies of Agents. Hershey: Idea-Group, 2007

Trajkovski, G.. Environment Representation in Multiagent Systems. Ph.D. Thesis, University Ss Cyril and Methodius, Skopje, Macedonia, 2003.

Varela, F., E. Thompson and E. Rosch. The Embodied Mind. Cambridge: MIT Press, 1991.