The Interconnected Roles of Abstraction and Emergence in Artificial Societies

Eric Baumer and **Bill Tomlinson**
Department of Informatics, University of California, Irvine
127B Computer Science Trailer
Irvine, CA 92697
{ ebaumer, wmt } @ ics.uci.edu

Abstract

This paper presents an argument that the process of emergence, moving from simple rules to complex behavior, is the converse of the process of abstraction, moving from complex behavior to simple laws. Thus, it can be argued that similar mechanisms underlie both processes, and a greater understand of one can lead to a greater understanding of the other. Especially in the case of societies, the processes of abstraction and emergence are inextricably interconnected, such that the abstractions individuals make will determine what behaviors emerge, and the behaviors that emerge in the society determine what abstractions will be made. In addition to describing this abstraction-emergence loop, the paper offers a description of an ongoing research project aimed at exploring and understanding the nature of this loop and the emergence to which it leads. Such an understanding of the relationship between abstraction and emergence can be helpful in designing communities of autonomous agents that interact socially with each other and with humans.

Introduction

When discussing emergent phenomena and emergent behavior is society, many questions often arise. What is emergence? What are the driving mechanisms behind it? How might we go about studying it? By understanding the mechanisms that drive the process of emergence, might we be able to predict, or even direct, what properties will emerge in a given system? This paper presents one possible avenue for investigating these questions, by exploring the relationship between emergence and abstraction in artificial societies.

Many varied definitions of emergence have been offered; for a survey of the subject, see (Damper 2000). This paper will adopt the definition that emergence is the process of higher-level phenomena resulting from the interactions of lower-level rules, where the lower-level rules do not contain any direct description of the higher-level phenomena.

Why should emergence be a topic of interest to researchers? Moreover, why should researchers in AI and multiagent systems (MAS) be interested in emergence? One of the main benefits is that the system designer does not necessarily need to determine the details of every action or interaction in which the agents may engage in advance. Rather, the designer can specify rules that describe the ways in which individual agents will interact, and the agents themselves can negotiate the details of specific interactions. Because the behavior of the system as a whole results from the interactions of individual rules, Holland argues (Holland 1998), they are a prime subject for the study of emergence.

The next question becomes, why study emergence in societies? There are at least two reasons to do so. One, MAS can be viewed as societies (e.g., (Lindemann *et al.* 2005)), and so examining the social ways that agents interact may help us design better systems. Two, emergence can be seen at work in human societies, and so understanding the processes by which norms and other social patterns emerge can be helpful in understanding the processes behind human social change. However, studying this process in human societies is difficult, because the behavior of individuals is available to an observer, but the representations and processes inside the individuals' heads are not. Thus, this paper supports the approach of using agent-based simulations in order to study societal emergence. While this approach does lack some of the richness found in everyday human interaction, it opens up the door to observing the mental processes that underlie individuals' actions.

Similar to the discussion above regarding emergence, one could ask why abstraction should relate to the same set of topics? By abstraction, this paper refers to the process of removing details of specific instances of a phenomenon to find simple rules that describe that phenomenon. Abstraction is one of the primary methods by which humans learn; for example, people perform abstraction to learn to recognize the shape of a square despite the numerous different colors, sizes, or orientations it may have. AI researchers have already developed various mechanisms to perform abstraction, such as reinforcement learning (Kaelbling, Littman, & Moore 1996), Q-learning (Watkins & Dayan 1992), and other pattern classifiers, each with their own unique strengths and features. This paper seeks to explore how we can put such mechanisms for abstraction to use in multi-agent societies.

As with emergence, the next question becomes, why focus on abstraction in societies? In human societies, people create abstractions that represent the various social pat-

terns they encounter, such as norms (Deutch & Gerard 1955; Kelley 1955) and institutions (Berger & Luckmann 1966; Brinton & Nee 2001). These abstractions then inform how the individual behaves in society. This paper argues that by better understanding this process of abstraction, we may gain a better understanding about how and why certain social patterns emerge and others fade away.

This paper argues that the processes of abstraction in society are closely linked to the processes of emergence. Individuals in a society make abstractions about their various interactions, and those abstractions then serve to guide further social activity. Based on the social activity thus informed, individuals continue to make further abstractions. Exploring the nature of this connection between abstraction and emergence may help to give us a better understanding of both processes.

Previous Approaches to, and Varieties of, Emergence

Numerous authors from various fields have sought to develop a rigorous definition of, and approach to studying, emergence. Damper provides a useful review of these definitions (Damper 2000). The following section highlights a selected few, showing their connections to, and differences from, the argument put forth here about abstraction and emergence.

Holland's book represents one of the most definitive works in the study of emergence (1998). Holland is most interested in emergent properties based on a simple set of rules, and so focuses largely on games as systems where a set of simple rules begets a set of complex phenomena. He is interested in the generative power of games and, to a slightly greater extent, numbers, as the two represent abstractions away from the details of a situation. Games like chess create an abstraction by "stripping away" the full detail of a battle field. Numbers create abstractions by stripping away all details about groups of things beyond the size of the group, giving rise to notions like "two-ness" to describe groups of two objects. Operating in the realm of mathematics, Holland develops what he call constrained generating procedures (cgp), which he sees as core to describing and orchestrating emergence by building it out of smaller pieces. The main difference between Holland's approach and that presented here is that Holland focuses on systems with static rules; the mechanisms and constraints are set at the beginning and then the procedure is run. However, this paper focuses on societal systems, where, first, different individuals may be governed by different sets of rules, and, second, those rules change over time to reflect the individual's interactions with others.

Slightly more recently than Holland's work in the area of emergence is Wolfram's tome A New Kind of Science (2002). In it, Wolfram lays out a research program for understanding and mapping the computational universe through empirical investigation, part of which focuses on experiments with one-dimensional cellular automata (CA). Enumerating and then running all the possible rules for such a CA, most end up leading to static, periodic, or random behavior, which he terms class 1, class 2, and class 3, respectively. However, a few CAs lead to definite structures that are neither periodic nor random, nor can such structures be described in any terms simpler than those in which they appear. Such systems are labeled as class 4; these are the systems that exhibit emergent properties. These classifications resemble Langton's λ-parameter, in that systems with the proper value for λ are the systems that exhibit emergent behavior (1990). This paper argues that the self-referential aspect of social behavior, as described in the abstraction-emergence loop, makes societies excellent candidates for classification as these types of emergent systems.

One important aspect of both Holland's and Wolfram's work is that they deal primarily with physical systems. When Holland talks about reductionism and emergence, he uses the example of quantum mechanical principles giving rise to the macrolaws of chemical bonding. This example addresses the way that we can see higher-level phenomena as the result of lower-level laws and interactions, but it does not address how that process of emergence actually occurs. Furthermore, it focuses on the workings of physical systems that are governed by a set of static rules rather than by rules that change with the functioning of the system. Wolfram has a similar focus on physical systems. Based on the assumption that all processes can be described in terms of computation, he creates graph models that exhibit certain emergent properties similar to properties of physical space. As with Holland, Wolfram focuses on systems where the rules are set and the behavior emerges, rather than on systems that allow for changing rule sets as influenced by the emergent behavior.

Cariani offers a classification system wherein he describes three different types of emergence (1990). The first, computational emergence, "is the view that complex global forms can arise from local computational interactions." (ibid, p. 776) Largely, this is the sort of emergence with which Holland, Wolfram, and others pursuing similar research programs are interested. The second, thermodynamic emergence, is couched in physical systems, such as the way that chemical bonding emerges from the laws of quantum mechanics. The third, emergence-relative-to-a-model, occurs when the behavior of a system differs from the current model for the system and the model must then be changed. Of the possible definitions, this last one is the closest to what this paper seeks to examine. Individual agents may have models of the way that social interaction is supposed to occur. However, emergence-relative-to-a-model is a strictly top-down view of emergence; it address the way an observer changes its model to reflect the phenomenon being observed, but implies a dichotomy between the observer and the phenomenon. This view of emergence does not take into account situations in which the observer is part of the system it is modeling, nor does it take into account situations in which multiple observers are present as different parts of the system. In this case, the observer may change its model, which causes it to change its behavior, which in turn causes others to change their models and behaviors, and so on. Societies are an example of such self-referential systems, and for this reason merit study of the unique forms of emergence they may exhibit.

This paper is certainly not the first to argue that social systems present a prime candidate through which to study emergence. A number of other researchers have pointed to the unique research opportunities with respect to studying emergence in social simulation, artificial societies, and "infosocieties" (Gilbert 1995; Conte 2001; Axelrod 2003). Such complex systems often have the ability to react to their own emergent properties, a trait called by some "second order emergence" (Baas 1993; Gilbert 1995; Steels 1995). First order emergent properties arise directly from interactions between parts of the system. Second order emergence occurs when the system reacts to its own first order emergent properties and often changes the structure of the system itself. The concept of second order emergence is similar to the ideas being explored here. Specifically, this paper seeks to explore the nature of the connection between first order and second order emergence. How do the ways in which a complex system reacts to its own emergent properties influence the further emergent aspects of the system? By allowing parts of the system to react in different ways by forming different abstractions, can we influence what properties emerge and the manner in which they do so?

The definition of abstraction adopted for this paper implies the ability to view phenomena at varying levels of abstraction. Thus, whether or not a phenomenon can be considered emergent is dependent upon whether or not it can be viewed at different levels of detail. It could be argued that some phenomena cannot be viewed at different levels of detail, which would automatically preclude them from being considered emergent. However, Damper argues that "emergence is best considered from the perspective of the understanding which can stem from viewing complex phenomena and systems at different levels of abstraction, as opposed to the difficulty or impossibility of doing so" (2000) (p. 816, original emphasis). This paper takes the argument a step farther, positing that the greatest understanding can come from not only viewing systems at different levels of description, but also understanding the processes by which those various levels are developed and interact with one another.

A Note on Structuralism

As the reader may have noticed, much of the work done on studying emergence has structuralist underpinnings. As first presented by Saussure and later pursued by the Prague School, structuralism briefly is characterized by the view that for any observable phenomenon, there is an underlying structure generating that phenomenon. Saussure's main interest was linguistics; one of his contributions was making a distinction between actual observed speech, which he called parole, and the grammar or language that produces it, which he called langue (Saussure 1916). This distinction led the way for Chomsky and others to develop formal, mathematical notions of generative grammars, which work very well as tools for linguistic analysis.

One can easily see how, in our various approaches to emergence, we take the approach that there is some set of underlying rules from which the phenomenon in question emerges. However, later theorists have questioned the validity of treating other areas of study with the same structuralist assumptions (Foucault 1969; Derrida 1973 1967). In the computational and mathematical forms of abstraction studied by Holland, Wolfram, Langton, and others, these assumptions clearly hold; there is a set of logical formalisms that describe exactly how computation is to proceed. However, post-structuralists question whether the assumption of structure holds for society; there is no clear indication, they argue that there is a generative social grammar that leads to the formation of norms, culture, or other social forms. Rather, society is the result of discourse between individuals, not any overarching structure (Foucault 1969).

By taking the approach to studying abstraction and emergence in society advocated here, one must ascribe, at least in a limited way, to some structuralist assumptions. Such assumptions are neither necessarily positive nor negative, but it is generally beneficial to be aware of one's assumptions and how they may bias one's approach to a particular problem or situation. Structuralism posits that there is one overarching framework that structures all social interaction. There are two subtle but important distinctions between this view and the approach advocated here that help to mitigate some of the problems of structuralist thinking. First, this paper argues not that there is a single overarching social structure, but rather that individuals maintain their own notions of the structures present in their society. Thus, we can circumvent the usual problem of explaining change when there exists only one overarching framework structuring social interaction, because the approach here focuses on each individual having its own, changing abstractions. Second, the individual's abstractions do not generate but rather guide its behavior, much the way that plans can be seen as resources for situated action rather than exact mappings of what actions will be taken (Suchman 1987). It is impossible to be utterly free from assumptions, so rather than trying to be free of them, the assumptions in this paper and the thinking that guides them are here made explicit for the reader.

Emergence, Abstraction, and Complexity

This paper puts forth the argument that the processes of emergence and abstraction are intimately and inextricably interconnected with one another. This section describes the nature of that interconnection.

The space of all possible systems may be organized along a continuum of complexity. In order to deal with vastly complex systems, people create abstractions that characterize the salient aspects of the system but do not incorporate all of the fine details. This process of abstraction (see the bottom arrow of figure 1) moves us from a realm of greater complexity to one of less complexity. In doing so, we move from a complex phenomenon to a simpler set of rules that describe the phenomenon.

Not every system lends itself to study by the method of moving from complex phenomena to simpler describing rules. Some systems, on the other hand, begin with a simple set of rules, which then generate a greater degree of complex phenomena. The canonical example of this kind of system in discussions of emergence is Conway's Game of Life (Gardner 1970). In this CA system, a few simple rules determine, at each successive tick on the clock of life, which

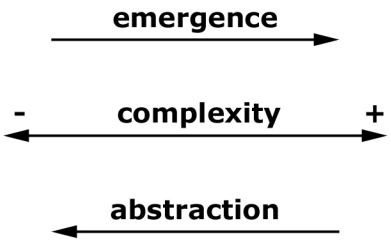

Figure 1: Abstraction and emergence move along a continuum of complexity.

cells live and which cells die. However, from these simple rules emerge myriad complex behaviors, moving along the continuum from less complexity to more (top arrow of figure 1). In this way, we can see the processes of abstraction and emergence as conversely related to one another, each moving the opposite direction along the continuum of complexity.

Holland makes a similar connection between emergence and reductionism (Holland 1998). Focusing on the ways in which we derive laws about our physical world, reductionism says that physical phenomena can be described by a set of laws, and finding these laws is the goal of basic science. This is distinct from the view of abstraction presented here, in that reductionism seeks to reduce a phenomenon to a set of laws from which the behavior of the phenomenon may be extrapolated. Abstraction, on the other hand, is the process of creating a set of rules that describe a system's behavior, but not necessarily finding the laws that generate the behavior. Holland ties this back to levels of description, saying that we could technically describe chemical reactions in terms of the quantum mechanics underlying the subatomic processes at work. However, it is more useful to develop macrolaws that describe these phenomena not in terms of quantum mechanics but in terms of the phenomena that emerge from the rules of quantum mechanics, phenomena such as chemical bonding.

Holland is mainly concerned with reductionism as it relates to physical systems. Later, this paper will show how the connection between abstraction and emergence is even stronger in social systems, but we can already see in this example the interconnected nature of the two processes. Through the process of abstraction, we create laws of quantum mechanics that describe the interactions of subatomic particles. Based on these lower-level laws, we then see higher-level phenomena, such as chemical bonding, as emerging from these lower-level laws. However, the view that the higher-level phenomenon is emergent comes from the lower-level laws we created by our processes of abstraction. Thus, the abstractions we make create rules that then determine our perception of what is emerging. That is, the process of abstraction drives the process of emergence.

The opposite is also true. Consider the Game of Life.

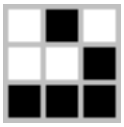

Figure 2: The initial configuration of a glider from Conway's Game of Life.

Given a set of basic rules, many complex phenomena emerge. These emergent phenomena then determine what abstractions, what macrolaws, we make about the system. For example, one of the most commonly discussed patterns is that of the glider (figure 2), a pattern of cells that, every four time steps, moves down one square and to the right one square. The higher-level phenomenon of the glider emerged from the workings of the lower-level rules of the game. Similarly, from the low-level interactions of quantum mechanics emerges the higher-level phenomenon of chemical bonding, which we then see as a phenomenon in its own right and codify into a macrolaw. The abstractions that we make about various phenomena are a direct result of the rules from which those phenomena emerge. That is, the process of emergence drives abstraction.

Thus, we can see that not only are abstraction and emergence conversely related processes, they are cyclically related as well, creating an abstraction-emergence loop (figure 3). The basic rules or laws that we formulate through our abstractions determine what phenomena emerge; these emergent phenomena then provide fuel for further abstractions. This interconnectedness of abstraction and emergence can give us new insights and new directions for research. For example, there already exists a body of knowledge on computational processes to perform abstractions, e.g. (Kaelbling, Littman, & Moore 1996). Because the two processes are conversely related, similar mechanisms may lie behind both, so that a deeper understand of one may help give us a deeper understanding of the other. In chemistry, freezing and melting can be seen as conversely related processes, one moving from liquid to solid and the other from solid to liquid. They are certainly not inverses, that is, the processes through which atoms go when crystallizing into a solid are not the exact opposite of the processes they undergo when changing into liquid from. However, because the two processes are conversely related, changing the matter's state in opposite directions, a better understanding of the process of melting can help lead to a better understanding of the process of freezing. Similarly, a better understanding of the process of abstraction can help lead us to a better understanding of the process of emergence, and vice versa.

Abstraction and Emergence in Society

At several points in his book, Holland points to agent-based models provide many opportunities for studying emergence in a general setting. By carefully choosing a few simple rules to govern the behavior of an individual agent, such as the rules that govern the behavior of individual ants within a colony, one can induce the emergence of complex behavior

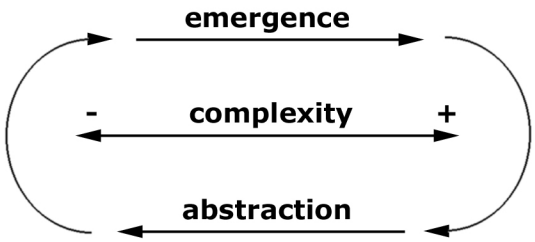

Figure 3: The abstraction-emergence loop.

by allowing such agents to interact with one another. However, rather than writing rules for these agents, what if the rules themselves allow for their own modification? To a certain extent, this is just giving the agents another level of rules. However, the nature of these rules is different, in that they are meta-rules about how to form rules. Such meta-rules allow for the type of self-reference that is key to the abstraction-emergence loop.

This feedback loop is essentially the situation we have when we consider social interaction. Individual social actors, in going about their various interactions, form representations of those interactions. Moreover, they abstract away from the details of individual interactions to formulate underlying rules that describe these interactions. In the case of societies, these rules are usually called norms (Deutch & Gerard 1955; Kelley 1955) or institutions (Berger & Luckmann 1966; Brinton & Nee 2001). However, the abstraction-emergence loop indicates that there is more going on here. When individuals form abstractions about what the normative behaviors are in their society, they begin to act on them, either by behaving differently themselves or by reacting differently to the behavior of others. Thus, the abstractions the individuals make drive the behavior that emerges from their interactions. But the loop does not end here. As the individuals continue to interact based on the abstractions they have made and new societal patterns emerge, the individuals will make abstractions about those new patterns, which in turn will give rise to a new set of emergent behavior, ad infinitum.

Now, one may ask why anything would ever change. Given the individuals in society, making abstractions about their interactions, and then using those abstractions to guide their future behavior, would the society not settle into a self-perpetuating stasis? If the same behavioral patterns emerge, are abstracted, and then emerge again, why would an individual's behavior ever change? The sociological theory of structuration calls this conflict, between the primacy of society structure and the individual's agency, the duality of structure (**?**). The individual within society is constrained to act within the societal structures as dictated by the actions of others in the society. However, the individual is not without agency, and in fact by its actions can influence those same societal structures that dictate its behavior. Indeed, it is the repeated re-enactment of societal norms on a regular basis that causes them to exist as such and perpetuates them (Berger & Luckmann 1966). Thus, the majority of the time, interactions will follow along the lines dictated by societal structure. However, there are at least two conditions in which interactions may not adhere to traditional norms, and these cases help lead to change in that social structure.

One condition occurs if a goal of the individual conflicts with a societal norm. Arguably, if a society's norms are not in line with an individual's goals, the individual might have realized this conflict and not joined the society in the first place (Lopez y Lopez & Luck 2003). However, assuming the agent is already a member of such a society (for example, by birth, or because of previous beneficial interactions), there may be a cost vs. benefit analysis, where the cost of breaking the societal norm is out-weighed by the benefit of achieving the individual's goal. If this is the case, the individual will act contrary to the norm, thus breaking other individuals' expectations.

This leads us to the second condition in which individuals may act contrary to a society's normative patterns. Societies, by definition, are composed of multiple individuals, and individuals are not necessarily homogeneous. The same interaction, observed by two individuals, may correlate to different abstractions each individual has made. Thus, the same event may be perceived differently by different observers. Furthermore, observing the same interactions may lead different individuals to make different abstractions, based on their perception of those interactions. Here, again, the abstraction-emergence loop comes into play, in that different individuals, making different abstractions, will be led to behave differently, causing different patterns of behavior to emerge.

Because of the connections between an individual's processes of abstraction and the processes of emergence in which it participates, societies make an excellent candidate for studying the abstraction-emergence loop. Since individuals' behaviors explicitly support the interconnections of abstraction and emergence described in the loop, we can observe that loop in action in order to understand the mechanisms behind it. Furthermore, in artificial societies, one can examine not only the social interactions being represented by the individuals' abstraction, but one can also observe the abstractions themselves. "It is in the interaction of representation and represented where," according to Suchman, "the action is" (1987). Lastly, in artificial societies, we have the even greater advantage of being able not only to observe the processes of the abstraction-emerge loop but in fact to engineer them. For example, agents that make abstractions via reinforcement learning rather than Bayesian belief networks will lead to different emergent patterns, because the abstractions generated by these two methods differ. Exploring the nature of those differences will be key to harnessing the power of the abstraction-emergence loop.

Applications

Thus far, this paper has described a perspective from which to study the process of emergence, that is, through the nature of its connection with the related process of abstraction,

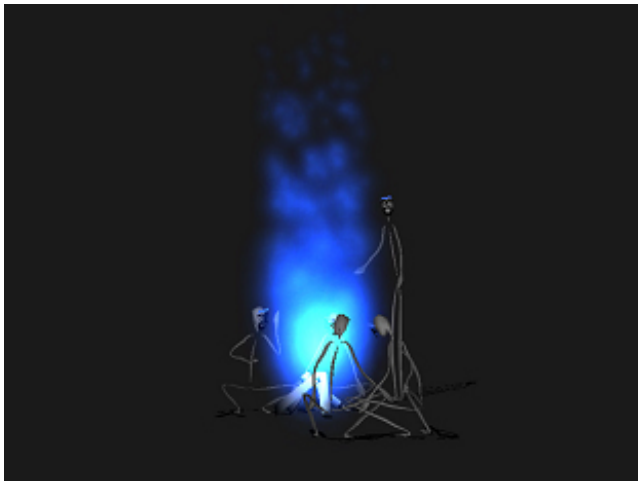

Figure 4: A community of agents inhabits a virtual island.

specifically in society. The authors and their research group are currently engaged in a number of projects that employ the approach advocated here. This section will describe one of those projects, which focuses on using the abstraction-emergence loop to generate social norms in an artificial society. Much previous work in social simulation and artificial societies has focused on simulations inspired by Conway's Game of Life, for example, the Sugarscape model (Epstein & Axtell 1996). In such models, agents inhabit individual cells on a grid, and their interactions with one another are relatively abstract: agents consume resources, reproduce, trade, and participate in a number of other activities. These activities are all very well defined and consist of one or more agents participating in the activity. To be sure, such simulations have been effective at reproducing various phenomena found in human societies. However, in human societies, what constitutes an action is not necessarily agreed upon by the whole, but rather is the result of a constant negotiation between participants in social interactions; via their social interactions, individuals are simultaneously instantiating and constructing the social structures that describe and define their social interactions (Berger & Luckmann 1966; Garfinkel & Sacks 1970; Giddens 1984). The research described here proposes a more phenomenological foundation for social simulation, so that we may understand not only how society-level phenomena emerge from individual-level interactions, but also how the definitions of those individual-level interactions emerge from the minutia of daily social life.

To this end, the authors and their collaborators developed Normative Echoes (Baumer *et al.* 2006), an interactive installation which revolves around using the abstraction-emergence loop to allow communities of animated characters to form normative speech patterns informed by patterns spoken by users. The installation features groups of animated autonomous agents that inhabit stationary computers, which represent islands of virtual space (figure 4). These agents communicate with one another around the central bonfire, using .wav files to make their communications audible to human participants. There is no semantic content to their communications, but rather by communicating they share patterns with one another. When one agent voices a certain sequence of wave files, other agents create a scenario describing that sequence using ScenarioML (Alspaugh 2005), a scenario specification language. The agents then make abstractions from these scenarios by simple frequency, that is, scenarios representing frequently spoken sequences are cast as representative abstractions. The agents then use these abstractions to inform further communications with one another. There are also slight random perturbations to keep the communications between the agents from becoming entirely homogeneous and to encourage the development of new structures. These further communications are then subject to the same abstractions, thus harnessing the abstraction-emergence loop to create normative patterns. However, agents do not only learn to communicate from one another, but from human participants. Each stationary computer has a microphone attached through which participants can talk to the characters on that island. Using simple signal processing, the participant's speech is broken up into wave files based on drops in amplitude of the input. These individual waves files correspond to the words and phrases that compose the participant's speech. The characters who hear these utterances start off by creating a scenario that describes those specific words and phrases being spoken in that order. However, they also add these words and phrases to their vocabulary as possible components to use their random perturbations. In this way, the abstraction-emergence loop also includes the human user's actions informing the agents. Such techniques could be used to improve the learning capabilities of intelligent agents in human-computer interaction (Maes 1994).

At the time of submission, this project has been accepted to This project was shown in the demonstration program of AIIDE-06. While there, the authors are in the process of analyzing human participants' interactions with the installation by means of video analysis. The goal of this analysis is in part to understand what happens when human users try to understand the behavior of software that is, in turn, trying to understand them. In this way, we hope to explore what happens when the abstraction-emergence loop is expanded to include both artificial agents and humans interacting socially. Specifically, the humans and the computational agents use different methods to create abstractions about their interactions with one another. How will the abstractions formed differ? Will these different abstractions cause difficulty in establishing mutual understanding between the humans and computational agents? How will these different abstractions influence further interactions and affect what patterns of social exchange emerge?

Future Work

In another project, currently in the early stages of development, the authors plan to use the online game Second Life (LindenLab 2003) as a test bed to explore the application of the abstraction-emergence loop to human social behavior. There are several rationales behind this choice. Anthropolo-

gists and social scientists are already using online gaming to study social behavior (Boellstorff 2006), especially Second Life. In physical reality, human activity recognition and partition into individual actions is a very challenging problem, but online game worlds, what constitutes individual actions is determined programmatically by the system. Thus, rather than focusing on activity recognition, the research can focus on the task of evaluating the applicability of the abstraction-emergence loop to human social activity. Furthermore, Linden Lab, creators of Second Life, explicitly supports the use of their software for educational purposes. Lastly, much of the content for the game is created by players; there already exist many scripts and bots for performing meta-tasks within the game that the research team may be able to harness.

The actual plan of study consists of two main phases. In the first phase, human participants' Second Life accounts will be instrumented to collect data about their actions and interactions with other players. This data will include what actions are taken by which players, the context of those actions described in terms of what other actions are taken, and the temporal relationships between those actions. Using a variety of methods, such as forms of reinforcement learning (Kaelbling, Littman, & Moore 1996), mechanisms like Q-learning (Watkins & Dayan 1992), genetic algorithms (Koza 1992), and methods for institutionalization (Baumer & Tomlinson 2005), abstractions will be made from these data that describe the interaction patterns within that society. These abstractions will then be used to try to predict what patterns will emerge during player interactions. Based on discrepancies between the predicted and actual actions, the abstractions will be further modified. In the second phase of the study, those abstractions will be validated using the following method. The abstractions will be used to guide a simulation, where each agent in the simulation is based on one of the characters in the study. Data describing this simulation will be given to human participants, as well as data logs describing actual interactions from the game. The test will be whether or not participants can distinguish between data generated by the simulation and data gathered by the game. According to the abstraction-emergence loop, the types of abstractions individual players make about their interactions will affect what social patterns emerge, and vice versa. If the patterns that emerge from these computational abstractions resemble those that emerge from human social interactions, then the computational abstractions may be an accurate representation of those formed by human players. Furthermore, just as algorithms for creating abstractions may be classified into groups, if the properties of the abstraction-emergence loop hold, then the patterns that emerge from these various classifications of abstraction algorithms will be similarly classifiable. For example, the use of reinforcement learning should lead to the emergence of patterns that are similar to each other and simultaneously different from the patterns emerging from the use of genetic algorithms.

Another, more complex test would be to introduce into the actual game world autonomous characters, or bots, whose behavior is governed by the abstractions acquired during the first phase and used for the simulations described here. The test in this case is if the bots, using the abstraction-emergence loop to continuously modify their behavior, can integrate into the fabric of society within the game. Currently there do not exist well-defined quantitative metrics for measuring this form of membership in the society of a virtual world. Furthermore, the bots would run the risk of deceiving Second Life's paying clientele and betraying their trust that live humans are controlling the characters with whom they interact. Nevertheless, such in-situ social analysis of the abstraction-emergence loop would be very helpful to determine the nature of its effect on human societies.

It could be argued that this process amounts to a Turing test for the social interactions that emerge from the acquired abstractions. As mentioned above, this paper argues for more phenomenological forms of social simulation, and the tests described here present a way of evaluating the efficacy of interaction-level social simulations in reproducing the details of social exchange. Furthermore, within societies, mutual intelligibility is one of the more important products of the abstraction-emergence loop. Verifying whether such intelligibility can be obtained is core to verifying the effects of the abstraction-emergence loop.

Another question raised by the above methodology is whether the results will have any implications beyond the video gaming domain studied. Early sociological and anthropological work that focused on online interaction often portrayed the internet as a route to explore multiple alternate personalities, or express certain otherwise non-dominant facets of one's own personality. For example, some MUD (multi-user dungeon) players had several characters, each with a different personality, and chose which character to play based on they current mood or what facets of their own personality they felt like exploring (Turkle 1995). However, more recent work suggests that people use the internet not to be someone else but as new ways of being themselves. For example, Trinidadians, for whom connection with family is an important value, use the internet as a new way to be Trinidadian, such as sending email reminders to their child in London to bring an umbrella because the forecast predicts rain today (Miller & Slater 2000). This later work suggests that social dynamics in an online setting bear very close resemblance to social dynamics at work in offline situations. If the abstraction-emergence loop can be used to analyze, model, and simulate social behavior online, it may be able to do the same for a broader range of human social dynamics.

Pursuing the projects described above is the next step for this line of work. However, the most important aspect of future research will center around exploring the nature of the connections in the abstraction-emergence loop. As stated above, researchers have developed many different computational methods of forming abstractions. Can those different methods be classified in a rigorous way that reflects the types of abstractions they produce? Could such classifications be useful in further classifying various types of emergence, perhaps going beyond the schemata presented by Wolfram or Langton? Furthermore, are there certain mechanisms for abstraction that are more or less suited for use in the abstraction-emergence loop? Such future work will help us to better understand abstraction, emergence and

the nature of their interconnectedness, as well as guide us in applying that understanding to the development of multiagent systems and the study of social interaction.

Conclusion

There exist many approaches to understanding the nature of emergence. In this paper, we have argued for a treatment of the process of emergence as the converse of the process of abstraction. Because the two processes are similar, emergence moving from less complexity to more and abstraction from more to less, understanding the mechanisms that underlie one may assist in understanding those underlying the other. Furthermore, since social systems exhibit an abstraction-emergence loop, societies present an interesting and valuable avenue through which to study emergence. Individuals make abstractions about their interactions, and these abstractions guide what future behavioral patterns emerge. Simultaneously, the emergent behaviors in the system are subject to individuals' abstractions, which then lead to new emergent behavior. This self-referential aspect of societies makes them a prime target for studies of emergence; this paper has presented a number of methods by which to perform such studies. In general, the research program suggested here is to develop systems that employ various abstraction mechanisms and various ways for behavior to emerge from the interaction of rules, and then to study the ways in which different methods of abstraction and different emergent processes affect one another to gain a greater understanding of the abstraction-emergence loop. An understanding of the interconnected natures of emergence and abstraction may give us not only better means by which to create multiagent systems, but also a better understanding of the processes at work in our human societies underlying norms, culture, and social change.

Acknowledgements

Eric Baumer would like to acknowledge Eric Aaron for the conversation that helped begin this line of thinking.

The authors also thank the Donald Bren School of Information and Computer Sciences, the Claire Trevor School of the Arts and the California Institute for Telecommunications and Information Technology for their support.

Additional thanks are due to Jeff Ridenour for several useful references; Tom Boellstorff, for helping to point out some of the structuralist assumptions underlying this work; the members of the Social Code group at UC Irvine, who have served as a sounding board in refining these ideas; and several anonymous reviewers who provided helpful comments and references.

References

Alspaugh, T. A. 2005. Temporally expressive scenarios in scenarioml.

Axelrod, R. 2003. Advancing the art of simulation on the social sciences. *Journal of the Japanese Society for Management Information Systems* 12(3).

Baas, N. 1993. Second order emergence. In *Oral communication at the second European Conference on Artificial Life*.

Baumer, E., and Tomlinson, B. 2005. Institutionalization through reciprocal habitualization and typification. In *Second NASA Workshop on Radical Agent Concepts (WRAC)*, volume 3825. NASA Goddard Space Flight Center, Greenbelt, MD: Springer.

Baumer, E.; Tomlinson, B.; Yau, M. L.; and Alspaugh, T. A. 2006. Normative echoes: Use and manipulation of player generated content by communities of npcs. In *AI in Interactive Digital Entertainment (AIIDE-06) Demonstration Program*. Marina del Rey, CA: AAAI Press.

Berger, P. L., and Luckmann, T. 1966. *The Social Construction of Reality: A Treatise on the Sociology of Knowledge*. New York: Irvington Publishers, Inc., first irvington edition 1980 edition.

Boellstorff, T. 2006. A ludicrous discipline? ethnography and game studies. *Games and Culture* 1(1):29–35.

Brinton, M., and Nee, V. 2001. *The New Institutionalism in Sociology*. Stanford, CA: Stanford University Press.

Cariani, P. 1990. Emergence and artificial life. In Langton, C. G.; Taylor, C.; Farmer; and Rasmussen., eds., *Artificial Life II*. Santa Fe, NM: Santa Fe Institute.

Conte, R. 2001. Emergent (info) institutions. *Journal of Cognitive Systems Research* 2:97–110.

Damper, R. I. 2000. Editorial for the Special Issue on 'Emergent Properties of Complex Systems': Emergence and Levels of Abstraction. *International Journal of Systems Science* 31(7):811–818.

Derrida, J. 1973 (1967). *Speech and Phenomena and other essays on Husserl's Theory of Signs*. Evanston: Northwestern University Press.

Deutch, M., and Gerard, H. B. 1955. A study of normative and informational social influence upon judgment. *Journal of Abnormal and Social Psychology* 51:629–636.

Epstein, J. M., and Axtell, R. 1996. *Growing Artificial Societies from the Bottom Up*. Cambridge and Washington: MIT Press and Brookings Institution Press.

Foucault, M. 1969. *The Archaeology of Knowledge*. Paris: Editions Gillimard.

Gardner, M. 1970. Mathematical games: The fantastic combinations of john conway's game of 'life'. *Scientific American* 112–117.

Garfinkel, H., and Sacks, H. 1970. Formal structures of practical action. In McKinney, J. C., and Tiryakian, E. A., eds., *Theoretical Sociology*. New York: Appleton Century Crofts.

Giddens, A. 1984. *The constitution of society: Outline of the theory of structure*. Berkeley: University of California Press.

Gilbert, N. 1995. Emergence in social simulation. In Gilbert, N., and Conte, R., eds., *Artificial societies: The computer simulation of social life*. London: UCL Press. 144–156.

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

Kaelbling, L. P.; Littman, M. L.; and Moore, A. W. 1996. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research* 4:237–285.

Kelley, H. H. 1955. The two functions of reference groups. In Swanson, G. E.; Newcomb, T. M.; and Hartley, E. L., eds., *Readings in social psychology*. New York: Holt, 2nd edn edition. 410–414.

Koza, J. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Langton, C. G. 1990. Life at the edge of chaos. In Langton, C. G.; Taylor, C.; Farmer; and Rasmussen., eds., *Artificial Life II*. Santa Fe, NM: Santa Fe Institute.

Lindemann, G.; Ossowski, S.; Padget, J.; and VÃzques-Salceda, J., eds. 2005. *Agents, Norms, and Institutions for Regulated Multiagent Systems (ANIREM) workshop at AAMAS 2005*. Utrecht, The Netherlands: Springer.

LindenLab. 2003. Second life.

Lopez y Lopez, F., and Luck, M. 2003. Modelling norms for autonomous agents. In *Fourth Mexican International Conference on Computer Science (ENC'03)*, 238–245. IEEE Computer Society.

Maes, P. 1994. Agents that reduce work and information overload. *Communications of the ACM* 37(7):31–40.

Miller, D., and Slater, D. 2000. Chapter one - conclusions. In *The Internet: An Ethnographic Approach*. Oxford: Berg.

Saussure, F. d. 1916. *Cours de linguistique générale*. Lausanne and Paris: Payot.

Steels, L. 1995. The artificial life roots of artificial intelligence. In Langton, C. G., ed., *Artificial Life: An Overview*. Cambridge: MIT Press.

Suchman, L. A. 1987. *Plans and Situated Actions: The problem of human-machine communication*. Cambridge: Cambridge University Press.

Turkle, S. 1995. *Life on the Screen: Identity in the Age of the Internet*. New York: Simon and Schuster.

Watkins, C. J., and Dayan, P. 1992. Technical note: Q-learning. *Machine Learning* 8(3-4):279–292.

Wolfram, S. 2002. *A New Kind of Science*. Champaign, IL: Wolfram Media.