Preface

TOWARDS MORE LIVELY MACHINES

If there is one thing anthropology (and archaeology) teaches, it is that our tools are key to our identity as humans. In fact, this was one of the earlier—though now discarded—definitions of human (qua Carlyle). But, despite numerous other ethological examples of tool-use, we still tend to think of ourselves as ontologically grounded in the tools we use. Homo habilis is, after all, a "tool using man."

Certainly, anthropologists like Mead and Bateson (anticipating what would later become distributed cognition), have noted our embeddedness in systems composed of humans and their material culture (Bateson 1972). But, in the millennia since Acheulian hand-axes, we have not only developed more and more complex tools (and relied on them more), but, the tools themselves have begun to take on a life of their own. As Haraway wrote (in a style at once half ethnographic and half prognostication), our "machines are disturbingly lively, and we ourselves frighteningly inert" (Haraway 1985).

The fear of non-human agency is a theme in 20th century dystopian fiction—aliens, monstrous forces awakened by nuclear blasts ("The Thing," "Gojira"), robots carefully hemmed by laws hard-wired into their programming (Asimov's I, Robot), Draconian supercomputers orchestrating the end of humanity ("War Games," "Terminator"). And yet, just as compelling a case could be made that the possibility of non-human agency and intelligence represents the culmination of our human potential (Heckman 2008; Collins 2008).

If the 20th century suggested fear and unease with non-human agents, the 21st century adds a utopian edge, particularly in the hopes we have for Multi-Agent Systems (MAS). Whether optimizing resource allocation, organizing complex systems, or simulating human behaviors, the hope is that non-human agents may prove a palliative to the kinds of alienation we face in a society characterized by high degrees of mobility, tenuous relationships with place and people, and in general, unremitting complexity (Allison 2006). Thus, navigating news coverage, financial systems, traffic, internet searches, and so forth, are all thought to be assisted by systems of agent proxies (self)-organized about our individuated needs. From the fears of "Terminator" (being replaced by the robot), we all become hopeful cyborgs, variously hybrid agents embedded in our machines (Clark 2003).

Traditionally, multi-agent systems are composed of either software or robot agents, although many researchers have utilized "human agents" as a baseline in their development of non-human agents. Woolridge (2002:11) adds that they are:

at least to some extent capable of autonomous action—of deciding for themselves what they need to do in order to satisfy their design objectives. Second, they are capable of interacting with other agents, not simply by exchanging data, but, by engaging in the analogues of the kind of social activity that we all engage in every day of our lives—cooperation, coordination, negotiation, and the like.

Although Wooldridge's work grounds the articles in this volume in a common vision of MAS, we also go beyond this more engineering-inflected vision of MAS.

In Disney's retelling of *The Sorcerer's Apprentice* in its 1940 *Fantasia*, Mickey Mouse is overwhelmed by a material world over which he loses control. His broom, split into pieces, continues the (multi-agent) work of carrying (and dumping) water without Mickey, to the extent that the house floods. And yet, in MAS, the hope is exactly that—programmers and roboticists look (in some way) to lose control over the systems they've engineered, with the hope that some different kind of solution will emerge. For some, this is an adaptation to the world around us. For example, for Serugeno et al (2006: 45):

The complexity of today's applications is such (e.g., world scale) that no centralized or hierarchical control is possible. In other cases, it is the unforeseeable context, in which the application evolves or moves, which makes any supervision difficult.

In other words (qua "*The Sorcerer's Apprentice*"), we've already "lost control" in any deterministic, more Newtonian sense: financial markets, ecological catastrophe, refugee flows, and Internet traffic. Our problems are increasingly non-linear; traveling salesman-type (NP) problems are more and more the rule than the exception. The solution cannot be the reinstatement of patriarchical authority (the return of the sorcerer), but it's opposite—the manumission of control. The hapless apprentice, after all, had only granted the broom limited autonomy (to carry water and nothing else). If he had granted the broom the freedom to decide when there was enough water, than the problem could, literally, have solved itself.

This was certainly the hope embodied in something like Rodney Brooks-style reactive architectures, where the scientist looked to what kinds of behavioral phenomena might emerge from autonomous, multi-agent systems in way not reducible to individual agents or local rules. These kinds of hopes take on an almost mystic quality in something like artificial life, where emergence literally animates local rules, investing them with an elusive, black-box quality: life itself (Helmreich 1998). Although none of the contributors to this volume invest the same kind of religiosity in emergence, we nevertheless believe that MAS may be generative of novel adaptations with ultimately salutary effects for the humans who rely upon them.

This volume aims to address all of these issues and the emergence of societal phenomena in the interactions of systems of agents (software, robot or human) in particular. In a given environment, agents interact with each other, imitating, communicating, exchanging, and competing. Based on these heterogeneous modalities of interaction, a variety of socialities may emerge: language and communication, identities, economies, cultures. Tracking those emergences not only allows us to program more realistic simulations of biologies (human and otherwise), but may allows us to more effectively combine (qua hybrid agents) with our lively machines to form new socialities that are, themselves, doubly emergent—self-reflexively emergent.

We know (or, at least, think we know) what our non-human agents want. But what kinds of agency and intentionality emerge in hybrid systems composed of humans embedded in machine assemblages of non-human agents? The articles here go well beyond describing the next generation of MAS in simulations and system engineering; they gesture to the novel systems that we form (and that we might form) with our varied, lively tools. That is, the contributors to the present work are not only describing their research in the present, they are also gesturing to the kinds of MAS (with their own, attendant emergences) that may exist in the future. It is our belief that this meeting of AI research, cognitive science, and the social sciences, may constitute a novel direction for MAS that not only describes our lives in information society, but also intervenes in future assemblages of hybrid agents and agencies.

In other words, we hope the book acts as an agent in itself, in particular, what Michel Serres terms a "quasi-object," that is, an object that not only takes on agential properties, but also catalyzes agencies in others. Just as we now emulate the non-human agents we originally developed to simulate us, so the analysis of extant MAS may stimulate the development of new multiagencies, heretofore undiscovered conurbations of human and non-human, information and social sciences.

Accordingly, we have divided the books into three sections that attempt to structure this dialectic of revelation and evocation. We have tried to resist the power of the text (as an agent in its own right) to dictate the course of chapters. As artifacts of a particular way of seeing and ordering the world, texts present us with linear teleologies (in Aristotelian terms, the necessity of a "beginning, "middle" and "end"): things begin, develop according to their narrative logic, then culminate. Someone has the first word (an introduction), and someone, the last (a conclusion). We have tried to avoid characterizing this work (as well as MAS in general) as linear and have therefore ordered the book along principles which, in first-generation cybernetics, were called "circular causality": accordingly, the end of the book takes us back to the beginning, and the cascade of emergent phenomena and organizations described herein refer back and forth to teach other in a temporally chiasmic (if not emergent) fashion. Of course, we have not entirely succeeded. As Bruno Latour has pointed out of non-human "actants," our tools make reciprocal demands upon us, and to escape the kind of consciousness demanded of us by the text, we would need to leave the text behind all together.

SECTION I: INITIAL STATES

In Conway's now-apocryphal Game of "Life," "initial states" describe the configuration of cellular automata (and the rules for successive turns). Here, contributors offer insights into the foundations of MAS. But these states are themselves hardly given—they "emerge" out of the play of different disciplines, many of which come together in the space of this volume, and which we, in turn, hope may spur subsequent emergences. Along with these disciplines come diverse assumptions about psychology, social interaction, language, cognitive development, and culture, all of which form what Hegel might have called the "second nature" of MAS (Helmreich 1998). The first chapter, Sawyer's "The Science of Social Emergence," critically examines the sociology of emergence, developing an often-ignored, Durkheimian heritage into what amounts to a manifesto for a social science of emergence resting on a complex understanding of agents. Goldspink's and Kay's "Agentive Cognitive Capabilities and Orders of Social Emergence," builds, in many ways, on Sawyer's insights; interrogating the movement from agential properties to social emergence, and using an enactivist perspective to critique questions of structure and agency in sociology and to explore the challenge of modeling a social emergence that builds from cognitive to social levels. Bullington's "Agents in Social Interaction" takes up the genealogical task from the perspective of social psychology and ethology, the other two disciplines MAS research has most often drawn from, in particular, asking how different agents (human and non-human) interact together, and how insights from these studies can help researchers build more "life-like" agents to interact with us, including some of our more emergent properties (emotion, empathy and inference). Upal's "Predictive Models of Cultural Information Transmission" and Romero's "Interaction of Agent in E-Business" each take these interdisciplinary legacies into two applications—simulation and e-business, respectively—and, in the process, bridging the theoretical and conceptual configurations of this section with the emergent organizations in the next.

SECTION II: EMERGENCES

There are at least two levels of emergence at play in this section of the book. The first, as Sawyer writes, involves applications of the central premise of social emergence, the "simultaneous focus on three levels of analysis: individuals, their interactional dynamics, and the socially emergent properties of the group."

The second is the growing awareness among people within and without the information and computing sciences regarding the utility of MAS for "solving" (keeping in mind that only sub-optimal solutions may be possible) the problems of today's world. This section is witness to the varied contexts to which MAS have

been applied, and to the possibilities for their applications in areas rather far removed from areas usually associated with MAS.

These two directions mirror the general dynamics of emergence itself—the social sciences, cognitive sciences, and AI suggest the properties non-human agents might emulate, and influence the scope of simulations produced. At the same time, developments in MAS simulations suggest answers to old problems bedeviling social theory (e.g., structure versus agency) and gesture towards new opportunities for human-non-human interaction, hybrid MAS facilitated by these human/animal behavior emulating agencies.

Conover's "A simulation of Temporally Variant Agent Interaction via Passive Inquiry" critiques the one-dimensional, temporal assumptions built into extant simulations (and, synedochically, Conway's "Game of Life") and suggests the possibility of introducing heterogeneous temporalities into simulation design. Schilling's "Agent Feedback Messaging: A Messaging Infrastructure for Distributed Message Delivery" exploits some of those diverse temporalities in order to build scalable models of agent communications based in part on biofeedback.

Zhang et al. look to interactionist models of social cognition in order to build MAS where decision-making emerges from the interactions between agents rather than through the more autonomous models of decision making in classic rational choice theory. Similarly, in Part 1 of their "Developing Relationships between Autonomous Agents: Promoting Pro-Social Behaviour through Virtual Learning Environments," Watson et al. look to social interactionism, networking, and community, in order to build "socially interactive virtual agents" for the creation of virtual learning environments (VLEs), while Takác's "Construction of Meanings in Biological and Artificial Agents" underscores the problem and promise of communicative models in MAS. Tacking back and between ethological examples and AI simulation, Takáč proposes interactionist communications premised on models of evolutionary adaptation.

Abramson's "Training Coordination Proxy Agents Using Reinforcement Learning" examines the ways agents might build on models of teamwork in order to coordinate with other agents to fulfill the needs of human agents. Likewise,

Duong's "The Generative Power of Signs: The Importance of the Autonomous Perception of Tags to the Strong Emergence of Institutions" looks to one of the relatively undeveloped directions in agent perception in order to build new models for the emergent of MAS socialities.

Sierra's and Santibáñez's "Propositional Logic Syntax Acquisition Using Induction and Self-Organisation" explores the possibility for emergent socialities between diverse agents based on almost sui generis communicative models where syntactical structures emerge in the space of agent interaction. In their "Hybrid Emotionally Aware Mediated Multiagency," on the other hand, Vincenti and Braman explore the possibilities latent in more affective communications: what advantages might an "emotion-based agent" have over other kinds of social agents? Could emotion-based agents couple more effectively with human agents? Finally, Collins's and Trajkovski's "Mapping Hybrid Agencies through Multiagent Systems" inverts the usual assumptions implicit in MAS by suggesting that it is the human agents who may be emulating non-human agents, and that the task for the researcher is as much to develop different human behaviors as much as it is different models for non-human agents. In the process, they draw a much richer, and more ambiguous, picture of agent communication (including the possibilities in miscommunication). Fittingly, the application of some of these ideas leads us to questions of second-order emergence.

SECTION III: SECOND ORDER EMERGENCES

Second order emergences describe the agents changing their behaviors according to their awareness of emergent phenomena or behaviors. Here, we include not only a host of reflexively understood human phenomena, which, strictly speaking, gives rise to a recursive chain of emergences, but also to the possibility that our awareness of the possibilities inherent in MAS may catalyze new combinations of hybrid agents and new applications for those combinations. "Second order emergence" also refers to the state of MAS research in

general. Now almost two decades old, MAS research has moved into a new stage involving new sites of application as well as new hybridities linking together not only different systems, but also human and non-human agents in new ways—all enabled by our growing consciousness of both the usefulness of MAS as well as their (always already) ubiquity in our lives.

In Part II of "Developing Relationships between Autonomous Agents," Watson et al. take the social theories they elaborate in Part I in their designs of Virtual Learning Environments designed to reduce the incidence (as well as mitigate the effects) of school bullying. In these hybrid agent interactions, "believability" is an emergent category—non-human agents can be "too believable" (and hence unbelievable), as are ideas about empathy and engagement. In "Reputation: Social Transmission for Partner Selection," however, Paolucci and Conte look at reputation as the "meta-belief" enabling other beliefs and, in the process, generating other, emergent socialities—cooperation, altruism, and other reciprocal behaviors. Finally, in part II of Conover's "A Simulation of Temporally Variant Agent Interaction via Belief Promulgation," the forms emerging from temporal variance in a MAS are exploited by agents who attempt to influence each other's beliefs in the process stretching Conway's cellular automata to new, and emergent, applications in both simulations and future, hybrid MAS.

Newlin applies MAS to neurophysiology, and in the process introduces a tantalizing example of second-order emergence in the self-reflexive monitoring of oneself facilitated by the imitative impulse structured into our frontal-parietal mirror neuron system. In their "Relationship Between the Processes of Emergence and Abstraction in Societies" Baumer and Tomlinson also incorporate emergent cognition into their models, in this case what the authors terms an "abstraction-emergence loop" that captures the way agents generalize on their experience and thereby influence the behavior of subsequent local behaviors. But MAS cannot only be confined to applications in what might be called "lower-levels" of cognition. In Walker's "Emergent Reasoning Structures in Law," applications of a "Default-Logic" framework result in MAS capable of both rendering legal decisions as well as deliberating on the structure of legal reasoning itself, while in the process implicating both human- and non-human agents in the future of the legal process itself.

The final articles consider reflexivity in MAS, agents examining each other for new (wanted or unwanted) properties. Richardson's "Agents in Security: a Look at the Use of Host-Based Monitoring and Protection and Network Intrusion Detection" develops a model network intrusion where "malicious" and "normal" traffic are (secondarily) emergent concepts arising from an emergent MAS consensus. North et al. detail search tools for emergent agents. As new properties emerge in MAS, the relationship of the observer changes, that is, new kinds of properties are sought after and search engines represent the boundary between one kind of emergence (emergent properties of agents) and another emergence (new foci emerges from the consciousness of emergent properties).

Recursively, that search for new properties leads us looping back to the kinds of assumptions we held about MAS and their possibilities to begin with. Hence, back to the beginning of the book!

REFERENCES

Allison, Anne (2006). *Millennial monsters.* Durham, NC: Duke University Press.

Bateson, Gregory (1972). *Steps to an ecology of mind.* NY: Ballantine.

Clark, Andy (2003). *Natural-born cyborgs.* NY: Oxford University Press.

Collins, Samuel Gerald (2008). *All tomorrow's cultures.* NY: Berghahn.

Haraway, Donna (1985). *Simians, cyborgs, and women.* NY: Routledge.

Heckman, Davin (2008). *A small world.* Durham, NC: Duke University Press.

Helmreich, Stefan (1998). *Silicon second nature.* Berkeley: University of California Press.

Serugeno, Giovanna Di Marzo, Marie-Pierre Gleizes & Anthony Karageorgos (2006). "Self-Organisation and Emergence in MAS." *Informatica, 30,* 45-54.

Woolridge, Michael (2002). *An introduction to multisgent systems.* NY: Wiley and Sons.