Emergence of Learning in Computer-Supported, Large-Scale Collective Dynamics: A Research Agenda

Manu Kapur[1], **David Hung**[2], **Michael Jacobson**[3]
National Institute of Education, Singapore
[1]manu.kapur@nie.edu.sg, [2] david.hung@nie.edu.sg, [3]michael.jacobson@nie.edu.sg,

John Voiklis[4], **Charles K Kinzer**[5]
Teachers College, Columbia University, New York, USA
[4]jv37@tc.edu, [5]kinzer@tc.edu

Chen Der-Thanq Victor
National Institute of Education, Singapore, victor.chen@nie.edu.sg

Abstract: Seen through the lens of complexity theory, past CSCL research may largely be characterized as *small-scale* (i.e., small-group) collective dynamics. While this research tradition is substantive and meaningful in its own right, we propose a line of inquiry that seeks to understand computer-supported, *large-scale* collective dynamics: how large groups of interacting people leverage technology to create emergent organizations (knowledge, structures, norms, values, etc.) at the collective level that are not reducible to any individual, e.g., Wikipedia, online communities etc. How does learning emerge in such large-scale collectives? Understanding the interactional dynamics of large-scale collectives is a critical and an open research question especially in an increasingly participatory, inter-connected, media-convergent culture of today. Recent CSCL research has alluded to this; we, however, develop the case further in terms of what it means for how one conceives learning, as well as methodologies for seeking understandings of how learning emerges in these large-scale networks. In the final analysis, we leverage complexity theory to advance computational *agent-based models* (*ABMs*) as *part* of an integrated, iteratively-validated *phenomenological-ABM* inquiry cycle to understand emergent phenomenon from the "bottom up".

Introduction

The past few decades have seen a consistent and persistent evolution from an individualized conception of learning to a more collectivist, participatory conception (Lave & Wenger, 1991; Scardamalia & Bereiter, 2003; Hutchins, 1995). By and large, this forms the epistemological core of CSCL research (Stahl et al., 2006). Concomitantly, educational technology has also evolved from the days of individual computer-based instruction to interactive and participatory online environments; the latter ranging from small-group, CSCL environments to large-scale, multi-user, interactive environments such as 2nd Life, Wikipedia, topical online communities, etc. At the small-scale or small-group level (typically 2 - 30 people), CSCL research has made substantive progress into unpacking the interactional dynamics, though, much important work still needs to be accomplished. At the large-scale level (typically in the 100s and the 1000s), however, we understand very little of the phenomenon of how people leverage technology to come together to interact, participate, collaborate, and form emergent structures and patterns (Sawyer, 2005; Jenkins, 2006); a phenomenon hereinafter referred to as large-scale *collective dynamics*.

Indeed, one of the most intractable problems in the social sciences, in general, and the learning sciences, in particular, is how interacting groups of agents (e.g., people) create emergent organizations at the collective level that are not reducible to any individual (Epstein & Axtell, 1996; Goldstone, 2006). For example, cognition and learning emerge from a collective set of interacting neurons yet the notion of cognition is absent and incomprehensible at the neural level. Highly coordinated behavior emerges in flocks of birds yet the notion of coordination is absent at the individual bird level. Traffic jams propagate backward but they emerge from interactions between individual cars (controlled by their drivers) going forward; movement at the collective level runs in a direction opposite to that at the individual level. Social structures, beliefs, values, and norms emerge in groups of people—offline and online—that cannot be attributed to or dictated by any individual per se, e.g., communities of practice, Wikipedia, and so on.

How does such collective behavior emerge? How does learning emerge in these large-scale collectives? More importantly, how are some large-scale collectives able to adapt, learn, and persist, while others perish?

Understanding the interactional dynamics of how large-scale collectives function and perform is a critical and an open research question, especially in an increasingly computer-supported, inter-connected, media-convergent world (Jenkins, 2006). It is also a subject of increasing scientific importance given its prolific coverage in the premier scientific journals such as *Nature* and *Science* (for a review of this trend, see Goldstone & Janssen, 2005; also see the Proceedings of the National Academy of the Sciences, 2002).

CSCL researchers such as Stahl et al. (2006) and Suthers (2006) have also alluded to this. For example, Suthers (2006) argued that, "small groups should not be the only social granularity studied. For example, the emergence of social and knowledge capital in a community of practice may require tracing out the evolution of relationships and the formation and spread of ideas in networks of individuals larger than the small group" (p. 16). Notwithstanding, CSCL and learning sciences research in this area remains in its infancy (Jacobson & Wilensky, 2006; Goldstone, 2006). This is, in part, due to the focus of CSCL research primarily being on small-scale collective dynamics (hereinafter synonymous with and a referent to small-group CSCL), which, of course, is substantive and meaningful in its own right. The purpose of this paper, however, is to *extend* the CSCL research agenda by proposing a line of inquiry that seeks to understand computer-supported, large-scale collective dynamics.

Organization of the Paper

The rest of this paper is organized into five sections. In the first section, we discuss what our proposal means for how we conceive learning in small-scale versus large-scale collectives. Despite some obvious differences between the two, we argue that learning in collectives is an emergent phenomenon regardless of the scale at which the phenomenon takes place. An emergent conception necessitates that we next unpack the very concept of emergence, and this forms the second section. The third section examines existing CSCL methods to see if and how they may be used to gain understandings of emergent behavior of large-scale collectives, and whether we need to integrate existing methods with others not currently used in CSCL research. Our analysis sets up an imperative for a methodology that builds on existing CSCL methodologies and is able to capture and model emergent behavior of large-scale collective dynamics. The fourth section advances computational Agent-Based Models (ABMs) as *part* of an integrated, iteratively-validated *phenomenological-ABM* inquiry cycle to understand emergent behavior of large-scale collectives from the "bottom up." The fifth and final section concludes with directions for future research.

Unpacking Learning

A focus on collective dynamics—small-scale and large—requires that one adopt a broader conception of learning. Traditional conceptions of learning in cognitive science and educational research have tended to focus on the individual as the unit of cognition and learning; only the individual perceives, thinks, and learns. However, as mentioned earlier, the past few decades have seen a consistent and persistent evolution from an individualized conception of learning to a more collectivist, participatory conception (Lave & Wenger, 1991; Scardamalia & Bereiter, 2003; Jenkins, 2006; Hutchins, 1995). While this evolution is generally in agreement with CSCL research, a careful reading of CSCL research reveals a continuum of conceptions (for an excellent review, see Stahl, 2005).

On the one end, learning is conceptualized primarily as an individual-level construct but one that may benefit from the collaboration. On the other, learning is conceptualized primarily as a group-level construct or activity in which individuals participate and interact with each other. Between these two ends, a range of conceptions exist of which we provide a sample; it is neither our intent nor is it possible to be exhaustive here. For example, if one leverages the knowledge acquisition metaphor, one could examine individual pre-to-post intervention gains as learning. Alternatively, if one leverages the participation metaphor, one could selectively examine the nature of an individual's participation in an activity or a community of practice as learning; conceptions of *legitimate peripheral participation* (Lave & Wenger, 1991) and *internalization* (Vygotsky, 1978) can be justifiably invoked as individual learning. At the collective level, one could just as well leverage the participation and interactional metaphors to conceive learning as an emergent property arising from the productive agency that drives participation and interaction between group members, e.g., knowledge building (Scardamalia & Bereiter, 2003). The interactional metaphor could even be leveraged in the radical sense to conceive learning not as something that is constituted *through* interaction but *is* the interaction (Koschmann et al., 2005). Further, and these are perhaps less frequently invoked but just as valid, conceptions of learning at the collective level may be tied to how groups adapt and reorganize in response to changing environmental and selection pressures (Kauffman, 1995). Here, emergent organizations and structures can be conceived as collective learning; the more a group is able to reorganize in a flexible and effective manner, the more it can be seen as learning.

Our own conception of learning is grounded in the science of complexity (Kauffman, 1995). We adopt the broader conception to include cognition and learning at the collective level as well. This conception clearly falls on the abovementioned continuum but is different in the sense that the "group" is not an *a priori* entity; it emerges (sometimes) from the intersubjective interactions between individuals, and once such an organization emerges, it shapes and constrains the very interactions it emerged from. While this dialectic is at the core of our conception, complexity theory also suggests a rather clear ontological and epistemological position: both individual and group cognitions co-exist; both the individual and the group co-learn; the nature of such learning being dialectical, dynamic, and emergent (Kapur et al., 2006; Voiklis et al., 2006).

Invoking emergence, however, only begs the question: how, when, and why does collective cognition emerge? Among cognitive and learning scientists, interest in collective cognition and its emergence is a recent phenomenon (e.g. Goldstone, 2006); existing theories (e.g. Hutchins, 1995) detail how collective cognition propagates *once* structured and organized, but for a theory of its emergence which does not presuppose these structures *ab initio*, one needs to look towards complexity theory (Kauffman, 1995; Epstein & Axtell, 1996). This need also stems from the cumulative effect of empirical research indicating that intersubjective processes at the local (individual) level yield cognitions—e.g opinions (Isenberg, 1986), knowledge representations (Schwartz, 1995), among others—that differ, both in complexity and kind, from those produced by any collaborating agent or those expected from the central tendency among collaborators (Vallacher & Nowak, 2004). Moreover, these group-level cognitions emerge spontaneously, without forethought or awareness among collaborating agents (Goldstone, 2006). Apparently, both the individual and the group learn; this learning being at once distinct, dialectical, and emergent.

From the preceding paragraph, it perhaps follows that the closest match for our conception with CSCL research is in the intersubjective learning epistemology. This is because intersubjective learning (as we understand it) focuses on the process of meaning-making in the social interactions distributed across actions, actors, and artifacts. It conceives learning at both the individual and collective levels although the ontology underpinning this epistemology (e.g., is there really a collective mind? How is it possible for learning to be distributed?) is still being debated (e.g., see Stahl, 2005; Koschmann et al., 2005). Notwithstanding this ontological debate (which we maintain is healthy), we do see sufficient epistemological coherence and consistency between the complexity and intersubjective conception of learning to move the conversation forward.

Moving forward, therefore, we articulate five dimensions along which small- and large-scale collectives exhibit critical differences in how they function and perform. In turn, these differences have implications for how learning emerges in these collectives. Note that while we present these differences as analytical dichotomies, in our conception they represent more of a difference of degree than of kind. Conceiving these analytical dichotomies as continuums, the five dimensions are:

i. <u>Spatial-Temporal landscape</u>: The space (both real and virtual) and time over which large-scale collective dynamics evolve is relatively larger; individual members are typically distributed over a much larger space and the phenomenon typically unfolds over relatively longer life spans (typically months, often years). In contrast, small-scale collectives, while also distributed, are limited to a relatively smaller space (both real and virtual). Plus, their timescale or the life span is relatively shorter (typically days or weeks, sometimes months, rarely years). Thus, the spatial-temporal landscape on which the phenomenon unfolds differs to a substantive degree.
ii. <u>Open vs. Closed systems</u>: Small-scale collectives in CSCL research typically tend to be relatively *closed* systems; once the group members and the mediating tools and artifacts (including scaffolds) are set up, there are typically few, if any, additions within the spatial-temporal landscape over which the group's collaboration unfolds. Large-scale collective dynamics, in contrast, are relatively *open* systems, where the number of group members, mediating tools and artifacts, and the spatial-temporal landscape are typically co-evolutionary.
iii. <u>Level of a priori structure</u>: The level of *a priori* structure in small-scale collectives in CSCL research is relatively high; these are intentionally designed learning environments (witness the significant research efforts on the efficacy of interactional scripting and scaffolding in CSCL research). Large-scale collectives, on the other hand, are typically decentralized and less structured. This does not mean that there are no structures in large-scale collective dynamics. What it means is that many (though not all) of these structures emerge spontaneously from within than being designed for from the beginning. Of course, structures emerge in small-scale collectives as well; we do not deny this. As stated earlier, it is a matter of degree; the likelihood of such structures emerging from within is greater in the large scale collectives than small-scale ones in CSCL research.
iv. <u>Individual agency</u>: In small-scale collective dynamics, individual members have relatively less agency compared to individuals in open, decentralized, large-scale collectives. In the latter, individuals can choose to

participate and collaborate if and when they want, and interact with whomever they want. This is relatively less so in the former. This difference in individual agency is important when one considers research on the importance of productive agency in collaboration (Schwartz, 1999). Combined with the previous dimension, the structure-agency dialectic leans more towards a priori structure for small-scale collectives whereas it leans more towards individual agency for large-scale collectives, where individual agency can even result in a dynamic reorganization of collective structures. Again, this is a difference in degree, not kind.

v. Diversity: Large-scale collectives also allow for relatively greater diversity among its individual members. Such diversity is relatively low for small-scale collectives mainly because of their relatively smaller size. Lack of sufficient diversity also makes the emergence of distributions in small-scale collectives highly unlikely (e.g., consider if it is even meaningful to speak of a normal or power-law distribution in a small group)

The relatively longer-term, open systems formed by computer-supported, large-scale collectives together with lower a priori structure, higher individual agency, and greater diversity make for an exponentially greater system complexity when compared to the small-scale collectives in CSCL research. It is this very complexity that sets up the stage for the *emergence* of learning—structures, interactional patterns, participation patterns and culture, knowledge, values, norms, etc.—making large-scale collective dynamics an intriguing phenomenon worthy of inquiry in its own right. If so, it becomes necessary that we unpack the very concept of emergence first before seeking methodologies for understanding it in the context of large-scale collective dynamics.

Unpacking Emergence

The concept of emergent behavior is, however, rather paradoxical. On the one hand, it arises from the interactions between agents in a system, e.g., individuals in a collective. On the other hand, it constrains subsequent interactions between agents and in so doing, seems to have a life of its own independent of the local interactions (Kauffman, 1995). For example, a traffic jam emerges from the local interactions between individual drivers; at the same time, it constrains the subsequent local interactions between individual drivers. Traffic jams, once underway, do seem to have a life of their own. Similarly, structures within social networks emerge from the local interactions between individual actors. At the same time these structures constrain the subsequent local interactions between individual actors (Watts & Strogatz, 1998), and so on. It becomes fundamentally important to understand *how* macro-level behaviors emerge from and constrain micro-level interactions of individual agents.

Understanding the “how,” however, requires an understanding of a cardinal principle in complexity: simple rules at the local level can sufficiently generate complex emergent behavior at the collective level (Bar-Yam, 2003; Kapur et al., 2006). For example, consider the brain as a collection of neurons (agents). These neurons are complex themselves, but exhibit simple binary behavior in their synaptic interactions. This type of emergent behavior, when complexity at the individual-level results in simplicity at the collective-level, is called *emergent simplicity* (Bar-Yam, 2003). Further, these simple (binary) synaptic interactions between neurons collectively give rise to complex brain "behaviors"—memory, cognition, etc.—that cannot be seen in the behavior of individual neurons. This type of emergent behavior, when simplicity at the individual-level results in complexity at the collective-level, is called *emergent complexity* (Bar-Yam, 2003).

The distinction between emergent simplicity and complexity is critical, for it demonstrates that a change of scale (individual vs. collective level) can be accompanied with a change in the type (simplicity vs. complexity) of behavior. “Rules that govern behavior at one level of analysis (the individual) can cause qualitatively different behavior at higher levels (the group)” (Gureckis & Goldstone, under review, p.1). We do not necessarily have to seek complex explanations for complex behavior; complex collective behavior may very well be explained via simple, minimal information, e.g., utility function, decision rule, or heuristic, contained in local interactions. Repeated updating, interaction, and aggregation of local interactions can sufficiently generate the phenomenon over time from the “bottom up” (Nowak, 2004). Bearing this albeit brief conceptual unpacking of emergence (within the constraints of a conference proposal) in mind, we now turn our attention to methodologies for how one might seek understandings of emergent behavior of large-scale collective dynamics.

Unpacking Collective Dynamics

A complexity-grounded focus on collective dynamics, as argued earlier, requires that one undertake an ontological and, consequently, a methodological shift. Making this methodological shift, in turn, requires that one examine existing methodologies in CSCL research to see if and how they may be used to gain understandings of large-scale collective dynamics, and whether we need to integrate existing methods with others not currently used in CSCL research. Broadly speaking, existing methodological approaches in CSCL research fall into one or more of three categories: a) *experimental*, b) *descriptive*, and c) *design* (Suthers, 2006).

For the purposes of our argument, however, the three categories may be reduced to two. This is because the third category—design-based approach—at the compositional, methodological level (as opposed to the theoretical level) uses methods that are typically descriptive, though sometimes integrative (descriptive cum experimental) to understand and explain learning in CSCL groups. Design researchers offer rich accounts of an iterative exploration of the possibility space of designs; once promising or effective design features are identified, experimental methods may be used together with descriptive methods to document and explain the emergence of learning in collaborative settings (e.g., Barab & Squire, 2004). At the methodological level however, one could reasonably posit that the design approach, in the final analysis, typically resorts to descriptive or integrative (descriptive cum experimental) approaches to gain and explain phenomenological understandings (Bielaczyc & Collins, in press). For the purposes of this paper, therefore, it suffices that we discuss the experimental and descriptive approaches, examine their usefulness and limitations in studying small-scale collective dynamics, and evaluate if and how they may be used gain understandings of large-scale collective dynamics.

Experimental Approaches

Experimental (including quasi experimental) approaches are pervasively used in CSCL research (e.g., Suthers & Hundhausen, 2003; Kapur, 2006). They typically seek to establish causal or quasi-causal explanations of design or intervention effects versus control conditions. Reductive quantification of qualitative interactional data into categories followed by counting and aggregation, and then linear statistical modeling typify this approach. While this approach allows one to draw aggregated-level interpretations and conclusions about relationships between manipulated variables and their effects, it has been criticized for over-simplifying the complexity of interactional dynamics in CSCL groups. Still, it serves a valuable purpose as a method for making quantified causal or quasi-causal generalizations, especially as a complement to descriptive methods (Stahl et al., 2006). With our focus on emergent behavior of collectives, we also need to examine this approach with regard to its assumptions of linearity. The reason for this is startlingly simple: assumptions that work at the atomic or particle level may not work at the human or social level (Kauffman 1995); one must closely examine the assumption of linearity.

Linearity is usually conceived both as a mathematical operator as well as a functional relationship. A linear operator is essentially an additive operator. For example, traditional analytical methodologies such as linear differential equations and statistical modeling, regardless of their mathematical sophistication, are essentially linear operators. They work well for *closed*, *linear* systems (or approximations thereof) where the whole is equal to the *sum* of its parts, thus allowing for a reductive analysis; one can break a system into its components or parts, study the parts individually, and then *add* the parts together to form the whole. Applying the linear operator and its associated methodologies to the study of collective dynamics means that a collective is no more than a simple aggregate of the individuals. Clearly, critical information is lost when heterogeneous actors (parts) are aggregated or averaged into factors (Eidelson, 1997). However, large-scale collective dynamics is an emergent phenomenon; emergent properties, by definition, can not be obtained and analyzed no matter how one adds or aggregates the parts. Thus, a study of collective dynamics calls for methodologies which permit the modeling of *open*, *non-linear* systems where the whole is greater than the sum of its parts, systems that exhibit emergent behavior.

Linearity may also be conceived as a functional relationship—constant proportionality or a straight line. Thus conceptualized, methodologies resting on an assumption of linearity are restricted to studies of phenomena in which the effects are proportional to their causes; a restriction that precludes a wide range of real-world social phenomena, including large-scale collective dynamics. Linearity tends to treat small changes or perturbations as temporally transient without any long-term effects. However, collective dynamics is a phenomenon that exhibits non-linear effects. One can no longer assume that effects are proportional to their causes. In fact, small changes or perturbations can and usually do have large effects. Therefore, important non-linear relationships among variables may be missed entirely, or worse, be modeled linearly since that is what the method can handle. As Holland (1995) explained it, "Nonlinearities mean that our most useful tools for generalizing observations into theory—trend analysis, determination of equilibria, sample means, and so on—are badly blunted" (p. 5).

Taken together, traditional experimental approaches and its underlying assumption of linearity may fail to capture let alone model emergent and self-organizing behavior of complex phenomena such as large-scale collective dynamics. This is not to suggest that we abandon their use altogether; instead, understanding the limitations of applying linear reductive methods to study non-linear emergent phenomenon requires that one exercise caution and humility in what can be accomplished with using this approach for large-scale collective dynamics.

Descriptive Approaches

One of the fundamental orientations in CSCL is the social-participatory construction of meanings as an inter-subjective, in-situ phenomenon (Koschmann et al., 2005; Stahl et al., 2006). CSCL research has focused on this emergent meaning-making process through descriptive approaches designed to gain rich, data-driven, bottom-up understandings of the phenomenon as it unfolds. These methods include conversation analysis (Sacks, Schlegoff, & Jefferson, 1974), discourse analysis (Johnstone, 2002), narrative analysis (Hermann, 2003), and the likes. In CSCL research (e.g., Roschelle & Teasley, 1995; Koschmann et al., 2005) these methods have been used to uncover the emergence of learning in small groups. Because one could use these methods at multiple scales of a phenomena (e.g., conversation or discourse analysis at a micro level, and perhaps narrative analysis at a phenomenological level), when used together, they make for a more ecologically valid understanding of emergent phenomenon. Even so, limitations of descriptive methods for small-scale collective dynamics have been well-articulated. These include, in the main, an inability to establish causal explanations/generalizations of interventions and design decisions as well as an overemphasis on theory building as opposed to theory application (Stahl et al., 2006).

Other limitations are specific to their use in large-scale collective dynamics, which forms our main concern. Specifically, the sheer spatial-temporal scale of large-scale emergent phenomenon limits (not negate) the usefulness of in-depth, descriptive analysis, which, by definition, requires that one focus on a humanly-manageable portion of the spatial-temporal landscape: the entire space and time over which the phenomenon unfolds (Eidelson, 1997). For example, if one is examining authorial dynamics in Wikipedia using descriptive methods, then the choice of the method itself limits the scope of what one might choose for an in-depth study, perhaps one or a few articles. This, of course, would not pose any problems if the spatial-temporal landscape of Wikipedia (and large-scale collective dynamics in general) was uniform so that an understanding of a small part may be applied uniformly to the whole; unfortunately though, this is rarely the case.

From a temporal standpoint, self-organizing and emergent behavior often occurs through abrupt phase transitions that tend to happen in a narrow temporal band of a phenomenon's evolution (Kauffman, 1995). A descriptive analysis likely makes it difficult (though not impossible) to detect this in a consistent and reliable manner (Kruse & Stadler, 1993). Similarly, large-scale dynamics can display drastically different characteristics in different parts of their spatial-temporal landscape. For example, an in-depth description of a small part of that landscape, while informative and meaningful in its own right, does little if what one is really seeking is an understanding of the entire landscape. Large portions of the landscape may appear highly orderly, yet the seeds of chaotic and emergent behavior may be located in a small part. Again, a descriptive analysis may make it difficult (though not impossible) to detect this in a consistent and reliable manner (Kruse & Stadler, 1993). Still, an army of descriptive studies large-enough to be distributed over varied portions of the spatial-temporal landscape of the phenomenon may yet prove to be highly useful provided one could somehow coordinate and integrate these efforts into a meaningful whole.

Moving Forward

Realizing that each approach has something to offer small-scale collective dynamics, CSCL researchers have called for greater integration of these approaches moving forward (e.g., Suthers, 2006). We second this call. However, as we have argued, both the experimental and the descriptive approaches—alone or combined—have limitations as methodologies for understanding large-scale collective dynamics. We see combining the two to be a necessary step; yet, that alone is insufficient for examining large-scale collective dynamics. The inherent complexities of large-scale emergent phenomenon place limits even on an integrative approach (Holland, 1995).

In light of our focus on large-scale collective dynamics, this sets up an imperative for a methodology that builds on the experimental and descriptive methodologies and is also able to sufficiently generate the phenomenon from the "bottom up". Given the technological advances in computational simulation power, *Agent-Based Modeling* (ABM; we use ABM as short for both agent-based models and agent-based modeling) provides a methodological *complement* that is increasingly being used not only in the natural sciences (Jackson, 1996) but also in economics (e.g., Arthur, 1990), sociology (e.g., Watts & Strogatz, 1998), socio-cultural psychology (e.g., Axelrod, 1997), organizational science (e.g., Carley, 2002), etc., just to name a few. Grounded in complexity theory, agent-based modeling has already provided significant theoretical and empirical insights into the dynamics of large-scale social systems (Eidelson, 1997). In the following section, we briefly describe agent-based modeling and examine its methodological potential for understanding the dynamics of large-scale collectives.

Agent-based Modeling (ABM)

Over the past two decades, computational agent-based modeling has emerged as an important tool for social scientists seeking to understand complex social phenomenon (Eidelson, 1997). In fact, evidence from computational ABMs is increasingly being argued and endorsed as a third legitimate source of scientific evidence, a third way of doing science (Axelrod, 1997); the other two being direct observation and mathematical manipulation (Jackson, 1996). It is not surprising then that computational ABMs are being used pervasively in both the natural and the social sciences (e.g., the work cited earlier). It is only recently though that learning sciences' researchers have begun to entertain the possibility of using computational ABMs (see Jacobson & Wilensky, 2006; Goldstone, 2006). However, their potential and use in the learning sciences and CSCL research remains largely unexplored. Therefore, a brief description of computational ABMs is in order.

ABMs shift the focus from *factors* to *actors*; one no longer has to work with homogeneous actors aggregated into factors. Instead, one could maintain the *diversity* of agents in a population as *heterogeneous* actors, each with its own set of genetic and cultural traits as well as *simple rules* of behavior (Axelrod, 1997). ABMs leverage the cardinal principle of complexity stated earlier or what is also known as the principle of dynamical minimalism (Nowak, 2004): simple rules at the local level can sufficiently generate complex emergent behavior at the collective level (Bar-Yam, 2003). Requiring complex explanations for complex behavior is not an ontological necessity (Kapur et al., 2006); complex collective behavior may very well be explained via simple, minimal information, e.g., utility function, decision rule, or heuristic, contained in local interactions. Repeated updating and interaction of local interactions can sufficiently generate the phenomenon over time from the "*bottom up*" (Nowak, 2004). Heterogeneous actors interacting with each other over *space* and *time* give rise to emergent global structures and patterns and these, in turn, dialectically shape and constrain the subsequent interactions between agents. ABMs model these emergent behaviors from the "bottom up" by computationally simulating the interactions between individual actors and letting the system evolve *in silico* (Epstein and Axtell, 1996). So, rather than positing emergent structures *ab initio*, ABM seeks to generate and understand how these structures emerge in the first place and shape the very local behaviors they emerged from (for a review, see Vallacher & Nowak, 2005). Thus, one is no longer restricted to an analysis of static equilibria in social phenomena. With ABMs, one can take a pro-active, process-oriented analysis of collective dynamics.

As an example, consider the computational ABM of Social Impact Theory (Nowak, Szamrej, & Latané, 1990) that simulates how polarized clusters naturally emerge in public opinion. Building on previous theory and empirical evidence, an ABM for social influence operating via two interlocking, dialectical mechanisms is hypothesized: the group influences each person, and each person influences the group. The intensity of the dialectic is derived from a function of three variables: group size, personal persuasiveness, and personal position in physical (or social) space. During the course of evolution, i.e., the iterative application of the social influence function to each group-on-person and person-on-group interaction, the simulation evolves from an initial random distribution of opinions into emergent organizations of islands (clusters) of minority opinion in a sea of majority opinion; an emergent organization of opinions not unlike that in the real world.

In this example as well as others cited earlier, simple agents interacting with each other using simple rules could sufficiently generate the emergent complexities that are qualitatively similar to what we observe in social phenomenon. Thus, *verisimilitude*—the plausibility of behavior and patterns—lends explanatory power to computational ABMs (and indeed to other scientific methodologies as well; the notion of sufficiency of explanation is integral to scientific inquiry though standards for what counts as sufficient vary across the fields, e.g., in the learning sciences and CSCL research, a *p*-value of .05 or less is commonly accepted as sufficient to demonstrate a causal or correlational explanation). If simple mechanisms operating on minimal variables produce realistic phenomena in a simulated world, perhaps the same simple mechanisms operating on the same minimal variables produce real phenomena in the real world (Nowak, 2004). What seems life-like could perhaps be like life (Voiklis et al., 2006). Thus, one could hypothesize theoretically-sound, computational ABMs, perform computational experiments, and validate the results against theory and empirical data (Goldstone & Janssen, 2005). In so doing, computational ABMs push the very notion of what it means to *explain* a phenomenon, what Goldstone and Janssen (2005) refer to as a "proof-by-construction." Epstein and Axtell (1996) articulate this notion succinctly in their book *Growing Artificial Societies*. They argue,

> "What constitutes an explanation of an observed social phenomenon? Perhaps one day people will interpret the question, "Can you explain it?" as asking "Can you *grow* it?" Artificial society modeling allows us to "grow" social structures *in silico* demonstrating that certain sets of microspecifications are *sufficient to generate* the macrophenomena of interest. And that, after all, is a central aim. As social scientists, we are presented with "already emerged" collective phenomena, and we seek microrules that can generate them…But the ability to grow them—greatly facilitated by modern object-oriented programming—is what is new. Indeed, it holds out the prospect of a new, *generative*, kind of social science." (Epstein & Axtell, 1996, p. 20)

While this may seem somewhat an over-enthusiastic endorsement, the ability of computational ABMs to model or "grow" emergent social phenomenon from the "bottom up" does provide an ontological coherence between the method and its object of inquiry. Epistemological and, consequently, methodological debates within the scientific discourse about the nature of knowledge and knowing from computational ABM experiments are also increasingly leaning in favor of computational ABM (PNAS, 2002). Still, these debates suggest that computational ABM is not without its own set of limitations. For example, while verisimilitude proves essential to the theory-building efforts of those trying to understand large-scale collective dynamics, clearly, an over-reliance on verisimilitude may strain one's evidentiary standards (Voiklis et al., 2006). Because of this, there is a great need for phenomenological validation of results derived from computational ABM (Goldstone & Janssen, 2005).

How might one achieve such phenomenological validation? Cioffi-Revilla (2002) suggests concrete steps for robust sensitivity and invariance analyses of computational ABMs to ensure that the simulated results are not merely "synthetic outcomes." Minimally, this may include examining the sensitivity and invariance of simulated results to *system size* (number of interacting agents), *agent geometry* (the structure of the spatial landscape on which agents interact, e.g., lattices), and *network topology*. Additionally, simulated results also need to be calibrated with respect to real-world phenomenon. Phenomenological *time calibration* would help ascertain the correspondence between notional time (the number of iterations in the simulation) and referent time (hours, days, months, years, etc.). Phenomenological *magnitude calibration* would help ensure correspondence between simulated and real phenomenon in terms of the magnitude or size or intensity of emergent behaviors. Phenomenological *distributional calibration* would help ensure that distributions (often power laws) that emerge in simulated phenomenon parametrically correspond with those in the real phenomenon.

As the field advances, this list will grow, as well it should. It is hoped that a persistent conversation between advocates and skeptics will potentially generate new ideas for phenomenological validation of computation ABMs. Over time, standards and metrics for what makes a *sufficient* explanation may emerge within this conversation. From the conversation thus far though, one thing is crystal clear: computational ABM alone too is insufficient; it cannot be done in isolation. One needs an integrative approach that builds on existing methodologies. For example, existing theoretical and empirical phenomenological understandings (gained through methods experimental, descriptive, or both) could be used to articulate critical variables and interactional rules between individuals agents in a collective. This, in turn, could be used to design agents and their interactional rules, which the computational ABMs could then simulate. Upon phenomenological validation, insights derived from the simulated collective behavior could in turn inform our theory building efforts. Importantly, a repeated, iterative application of this process cycle is most essential and forms the thrust of our methodological position for the study of computer-supported, large-scale collective dynamics. It is through such an iterative process of *building from* and *validating with* phenomenological theory and data—an *iterative* "*phenomenological-ABM-theory building*" cycle—that we seek a better understanding of large-scale collective dynamics: ABMs are hypothesized from theory and empirical data. Computational experiments using ABMs, in turn, provide new insights, explain empirical data, and inform theory building. This dialectic forms the epistemological and methodological core of our proposal.

Implications for a Research Agenda

To reiterate, our proposal for an overarching question driving the research agenda for computer-supported large-scale collective dynamics is: *How and under what conditions does learning—structures, participation patterns and culture, organizations, knowledge, values, norms, etc.—emerge from the interactions between individuals in computer-supported, large-scale collectives, and how do these emergents shape and constrain subsequent interactions between and participation of individuals in these collectives?* The research question, at this stage, remains necessarily broad. Still, it has been identified as a substantive and critical area for further inquiry not only by social scientists in general (e.g., Epstein & Axtell, 1996; PNAS, 2002) but also by cognitive scientists (e.g., Goldstone & Janssen, 2005) as well as learning scientists and CSCL researchers (e.g., Suthers, 2006). In CSCL research, the proposed research agenda might also help validate a fundamental hypothesis underpinning small-scale collective dynamics: small-groups form the critical organization that mediate between the individual and the larger community (Stahl et al., 2006). If this is so, this can be tested using the phenomenological-ABM inquiry cycle. For example, if one starts to see small group organizations emerging within the larger collective as the system evolves, then that may be one source of evidence—a generative, bottom-up, proof-by-construction—for the hypothesis.

We also argue, however, that any *comprehensive* research agenda for understanding large-scale collective dynamics will have to be multi-modal—offline and online. Why this is the case may not immediately obvious. Hence, it merits an explanation. One of the reasons for studying large-scale collective behavior in online societies and communities such as 2nd Life, Wikipedia, communities of practice, etc. is that a lot of people are increasingly participating in them (Jenkins, 2006). The emergence of an online community is contingent on such participation by real people in real networks across the world, i.e., as more people participate, they collectively create emergent structures online. In turn, this makes even more people participate setting up an *increasing returns* (Arthur, 1990) feedback loop. Clearly, emergent collective behavior in these environments not only shapes but is shaped by their spread in the real (as opposed to virtual) large-scale network, e.g., a population or a system. Thus, a study of collective behavior in large-scale online environments is incomplete without a study of how participation in such environments spreads in a real population or a system, i.e., the real and the virtual become a co-emerging, co-evolutionary phenomenon, and we argue that they must be examined and understood as such. If this is a plausible proposition, then a co-evolutionary research question becomes: *How does emergent behavior of computer-supported, large-scale collectives co-evolve with their diffusion or spread in a real (as opposed to virtual) large-scale network, e.g., a population?*

Finally, driving this research agenda is also an opportunity for data collection that is somewhat unique to computer-supported large-scale collectives. Many of these collectives have fairly complete records of their evolution automatically archived. Better still, these archival data are often freely available or one can design web-crawlers to seek required data from these archives. This clearly presents a unique opportunity to further the research agenda in this area (Goldstone & Janssen, 2005), which is precisely what the proposed research agenda aims to achieve.

References

Arthur, W. B. (1990). Positive feedbacks in the economy. *Scientific American*, *Feb*, 92-99.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, *41*, 203-226.

Barab, S. & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*(1), 1-14.

Bar-Yam, Y. (2003). *Dynamics of complex systems*. New York: Perseus Publishing

Bielaczyc, K. & Collins, A. (in press). Design research: Foundational perspectives, critical tensions, and arenas for action. To appear in J. Campione, K. Metz, & A.M. Palincsar (Eds.) *Children's learning in and out of school: Essays in honor of Ann Brown.*

Carley, K. M., (2002). Computational organizational science: A new frontier. *Proceedings of the National Academy of Sciences*, *19*(3), 7257-7262.

Cioffi-Revilla, C. (2002). Invariance and universality in social agent-based simulations. *Proceedings of the National Academy of Sciences*, *19*(3), 7314-7316.

Eidelson, R. J. (1997). Complex adaptive systems in the behavioral and social sciences. *Review of General Psychology, 1*(1), 42–71.

Epstein, J. M., & Axtell, R. (1996). Growing artificial societies: Social science from the bottom up. Washington, D.C./MA: Brookings Institution Press/MIT Press.

Goldstone, R. L. (2006). The complex systems see-change in education. *The Journal of the Learning Sciences*, *15*(1), 35-43.

Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, 9(9), 424-429.

Gureckis, T. M., & Goldstone, R. L. (under review). Thinking in groups. Manuscript submitted for publication.

Hermann, D. (Ed.) (2003). *Narrative Theory and the Cognitive Sciences.* Stanford, CA: Center for the Study of Language and Information.

Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. Reading, MA: Addison Wesley.
Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
Isenberg, D. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50, 1141-1151.
Jackson, E. A. (1996). The second metamorphosis of science: A second view (SFI Working Paper No. 96-05-059). Santa Fe Institute.
Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences*, *15*(1), 11-34.
Jenkins, H., (2006). *Convergence culture: Where old and new media collide*. New York and London: New York University Press
Johnstone, B. (2002). *Discourse analysis*. Oxford: Blackwell.
Kapur, M. (2006). Productive failure. *Proceedings of the International Conference of the Learning Sciences 2006*, Bloomington, Indiana, USA.
Kapur, M., Voiklis, J., Kinzer, C., & Black J. (2006). Insights into the emergence of convergence in group discussions. *Proceedings of the International Conference of the Learning Sciences 2006*, Bloomington, Indiana, USA.
Kauffman, S. (1995). *At home in the universe: The search for the laws of self-organization and complexity*. New York: Oxford University Press.
Koschmann, T., Zemel, A., Conlee-Stevens, M., Young, N., Robbs, J. & Barnhart, A. (2005). How *do* people learn? Members' methods and communicative mediation. In R. Bromme, F.W. Hesse & H. Spada (2005), *op. cit.* (pp. 265-294.)
Kruse, P., & Stadler, M. (1993). The significance of nonlinear phenomena for the investigation of cognitive systems. In H. Haken & A. Mikhailov (Eds.), *Interdisciplinary Approaches to Nonlinear Complex Systems* (pp. 138–160). Berlin, Germany: Springer-Verlag
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.
Nowak, A. (2004). Dynamical minimalism: Why less is more in psychology. *Personality and Social Psychology Review*, *8*(2), 183-192.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer Supported Collaborative Learning* (pp. 69-197). Berlin: Springer-Verlag.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking in conversation. *Language*, *50*(4), 696-735.
Sawyer, K. R. (2005). Social emergence: Societies as complex systems. Cambridge: Cambridge University Press.
Scardamalia, M., & Bereiter, C. (2003). Knowledge Building. In J. W. Guthrie (Ed.), *Encyclopedia of Education*. New York, USA: Macmillan Reference.
Schwartz, D.L. (1995). The emergence of abstract dyad representations in dyad problem solving. *The Journal of the Learning Sciences*, *4*(3), 321-354.
Schwartz, D. L. (1999) The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 197-218). NY: Elsevier Science.
Stahl, G. (2005). Group cognition in computer assisted learning. *Journal of Computer Assisted Learning*.
Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*. Cambridge, UK: Cambridge University Press.
Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315-337.
Suthers, D., & Hundhausen, C. (2003). An empirical study of the effects of representational guidance on collaborative learning. *The Journal of the Learning Sciences, 12*(2), 183-219.
Vallacher, R. R., & Nowak, A. (2004). Dynamical social psychology: Toward coherence in human experience and scientific theory. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles*. New York: Guilford Publications.
Voiklis, J., Kapur, M., Kinzer, C., & Black, J. (2006). An emergentist account of collective cognition in collaborative problem solving. *Proceedings of the Cognitive Science conference 2006*, Vancouver, Canada.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature, 393*, 440–442.
Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.