T. Koschmann, D. Suthers, & T.W. Chan (Eds.), *Computer Supported Collaborative Learning 2005: The Next 10 Years!*
Mahwah, NJ: Lawrence Erlbaum Associates.

Problem Solving as a Complex, Evolutionary Activity: A Methodological Framework for Analyzing Problem-Solving Processes in a Computer-Supported Collaborative Environment

Manu Kapur[1], John Voiklis[2], Charles K. Kinzer[3]
Communication, Computing, and Technology in Education
Teachers College, Columbia University
[1]manu_kapur@hotmail.com, [2]jv37@columbia.edu, [3]kinzer@tc.columbia.edu

Abstract. Viewed through the lens of complex systems science, one may conceptualize problem-solving interactions among multiple actors, artifacts, tools, and environmental structures as goal-seeking adaptations, and problem-solving itself, as a complex adaptive activity. Theories of biological evolution point to an analogical equivalence between problem solving and evolutionary processes and, thus, introduce innovative methodological tools to the analysis of computer-supported, collaborative, problem-solving processes. In this paper, we present a methodological framework for characterizing and analyzing these processes. We describe four measures that characterize genetic evolution - *number*, *function*, *fitness*, and *persistence* - to characterize the process of collaborative problem solving, and instantiate them in a study of problem-solving interactions of collaborative groups in an online, synchronous environment. Issues relating to reliability, validity, usefulness, and limitations of the proposed methodology are discussed.

Keywords: Problem solving, complex adaptive activity, convergence, fitness, persistence

INTRODUCTION

In this paper, we present a quantitative, analytical method for characterizing and analyzing the process of computer-supported, collaborative problem solving. Underpinning our work is a shared, situative, epistemological belief that learning in general, and problem solving in particular, is a continuous, dynamic process distributed in space and time over multiple actors, actions and artifacts, influencing and being influenced by the environment in a complex, adaptive, and iterative manner. As such, understanding the *process* of how multiple actors, artifacts, and environments interact and evolve in space and time on the way to an outcome ranks among the most important challenges facing educational research (Akhras & Self, 2000; Barab, Hay, & Yamagata-Lunch, 2001), and measures and methods for tracking the evolution of problem-solving processes as well as the emergence of learning are needed (Barab et al., 2001; Barron, 2003; Collazos, Guerrero, Pino, & Ochoa, 2002; Collazos, Guerrero, Pino, & Ochoa, 2004; Derry, Gance, Gance, & Schlager, 2000). Complex systems science is put forth as a framework for understanding the evolutionary dynamics of problem-solving processes and outcomes. From there, we derive a set of micro-genetic variables - *number*, *function*, *fitness*, and *persistence* - for characterizing the problem solving process, including how one might measure these variables. We situate our discussion and illustration of the proposed measures in a study of problem-solving interactions of collaborative groups in an online, synchronous environment. Finally, we discuss issues of reliability, validity, usefulness, and limitations of the proposed methodology.

THE NEED FOR A RECONCEPTUALIZATION

Despite the obvious complexity of problem-solving processes, existing problem-solving models remain linear, relatively rigid, and limited in scope for they are unable to account for multiple, dynamically changing actors, contexts, outcomes, and processes. For example, the General Problem Solver (Newell & Simon, 1972) specifies understanding and search processes as the two sets of thinking processes associated with problem solving. Here, "understanding" involves procedural knowledge, or an algorithm for solving a problem; "search" involves a means-ends analysis, or the selection of means (routine or subroutine in the algorithm) that will take one closer to the desired end (the ultimate or intermediate goal). A second example is the IDEAL problem solver (Bransford & Stein, 1993), which describes problem solving as a process involving several phases: *Identify* potential problems, *Define* and represent the problem, *Explore* possible strategies, *Act* on those strategies, and

Look back and evaluate the effect of those activities. Another example is STAR.Legacy (Software Technology for Action and Reflection) (Schwartz, Lin, Brophy, & Bransford, 1999), which also makes use of explicit inquiry cycles to describe problem-solving processes. The STAR.Legacy software shell separates and organizes complex problem-solving activities as cyclical sub-activities in attempting to help teachers and students manage complexity, guiding problem-solvers to begin with a challenge, generate initial ideas, consult experts, study resources, pilot-test, revise, and publish outcomes.

While the above and similar problem-solving models and cycles are descriptively very useful, they are limited because they tend to treat problem solving as either linear or cyclical in an effort to articulate a generalizable problem-solving process. However, problem solving is not a uniform, step-wise activity; problems vary in content and context (Jonassen, 2000). More significantly, these problem-solving models and cycles assume that their problem-solving processes are differentially applied in different contexts and situations but do not indicate exactly *how* that occurs.

An ability to account for the varying contexts and situations is somewhat offered by the prevailing socio-cultural, constructivist theories such as Activity Theory (AT) (Leont'ev, 1978). This is evident in their popular and persistent use by many researchers as a framework for understanding collaborative/collective activity. AT, for instance, provides the inclusiveness and plasticity needed to describe collaborative activity as a product of complex interaction among multiple actors, artifacts, and environment. As a result, working within the framework of AT, one can describe problem solving as a process of continuous change, development, and construction. One can also describe systemic structures - people, culture, and artifacts - that emerge through the process of problem solving as well as the emergent properties of interactions within those systems. Working strictly within the framework of AT though, one is not able to explain exactly *how* interactions contribute to the change, development, and emergence of structures and solutions over time. In other words, AT states that problem solving evolves over time, but not *how* it does so. As mathematical sociologist Mark Granovetter (1978) suggested in his model of collective behavior, an analysis of "norms, preferences, motives, and beliefs" (as indeed afforded by socio-cultural, constructivist theories such as AT) can account for the *necessary but not sufficient* conditions one needs in order to explain how these conditions "interact and aggregate" on the way to an outcome. We argue that taking a step towards achieving sufficiency, i.e., the theoretical and methodological tools to explain the problem-solving process and *how* it evolves in time, one requires a lateral step into complex systems science.

COMPLEX SYSTEMS SCIENCE

Complex Systems Science provides a framework for studying how interactions among the parts of any given system culminate in the behaviors of the system as a whole (Crutchfield, 1994). As a theory, complex systems science enables researchers from both the natural and social sciences to speak the same language as they study the same or similar macroscopic behaviors and interactions in a range of natural and artificial systems (Bar-Yam, 1997). As a science, complex systems science enables researchers to reason about this uncertainty, extracting measures and constructs that allow researchers to discover, describe, and predict *how* interactions form patterns, *how* patterns form complex systems, and *how* those complex systems behave (Crutchfield, 1994).

Problem Solving and Evolution: A Two-Way Analogy

According to complex systems science, adaptation is one macroscopic behavior shared across systems - biological, physical, and cognitive. A complex adaptive system (CAS) changes its behavior in response to environmental and internal feedback, often in an attempt to achieve a goal or objective (Bar-Yam, 1997). Goal-seeking adaptations that occur on a collective scale and/or over multiple iterations emerge as evolution (Bar-Yam, 1997). Modern synthesis, the prevailing theory of evolution, combines Darwin's theory of variation and natural selection with Mendel's theory of genetics to characterize evolution as a process of development or change over time. According to complex systems science, this process extends biological organisms to include the development or change of a culture, or an idea. Problem solving, too, involves iterative goal-seeking adaptations (or operations) through which an agent (or collection of agents) tries to reduce discrepancies between an initial problematic state and an ideal goal state (Newell & Simon, 1972). Thus, a group of people collaborating to solve a problem can be seen as a complex adaptive system, and evolution entails how the group interacts to solve a problem and how this interaction develops and changes over time. This facilitates a strong two-way analogy between problem solving and evolution - evolution may be characterized as a problem-solving process, and problem solving as an evolutionary process where ordered patterns move, sometimes through seemingly random paths, toward desired goals.

Measures for Characterizing the Problem Solving Process

Modern synthesis may provide useful analogies to describe what one might observe during the process of collaborative problem solving, but not what one might measure. *Cybernetics*, a field closely allied with complex systems science, provides three measures - *number*, *function*, and *fitness* of problem states - that prove informative for describing and explaining the problem solving process (Heylighen, 1988). When one imagines the collaborative problem-solving process as a sequence of problem states, the *number* of states from the initial state to the goal state can serve both as a temporal and spatial measure: temporally, each state is a *tick* on the evolutionary clock; spatially, each state is a *step* along the evolutionary path (Heylighen, 1988). In biological evolution, each mutation reconfigures the gene. Similarly, in problem solving, each interaction reconfigures the problem state. This reconfiguration may increase or decrease the difference between the reconfigured and goal states and, thus, the distance (number of *ticks* or *steps*) required to reach the goal state. Each interaction, then, has a positive (acceleratory) or negative (deceleratory) *impact* on the problem-solving process. In other words, each interaction *aims* to perform a telic function, i.e., it operates to reduce the difference between the current, problematic state and a specified goal state. Thus, one can view problem-solving interactions as *operators*, goal directed actions, performing a means-ends analysis in the problem space (Newell & Simon, 1972). However, with the exception of the initial interaction in a problem-solving episode, the configuration of a problem-state's properties results from more than one interaction; the configuration emerges from the cumulative impact of all the interactions up to that particular state. The distance between any intermediate state and the goal state reflects the cumulative impact of interactions. If problem solving means minimizing this distance between a given state and the goal (or end) state, then cumulative impact reflects the *fitness* of the collaborative problem-solving process at the given state (Heylighen, 1988).

Furthermore, another sub-domain of complex systems science - *artificial life* - provides an additional measure for characterizing the collaborative problem-solving process. In artificial life, a complex adaptive system is viewed in terms of the behavior of its constituent components. For example, in biological evolution, the components are families of genes. Analogously, in collaborative problem solving, these components can be seen as taxonomic families of interactions or functional categories. Adaptive evolution of the problem-solving process can be expected to affect the dynamics of these categories (Bedau, Snyder, & Packard, 1998) and conversely, the dynamics of these categories inform the evolutionary activity structures of the problem-solving process. The dynamics here refer to the evolutionary activity of these components, how and when they come into *existence* as well as their subsequent *usage* in the system. In other words, by using a measure of persistence one can identify the traits (functional categories) introduced by each interaction, and then track the use and usefulness of those traits. When added to fitness analyses, persistence may reveal how multiple evolutionary processes converge on similar paths without implying a single best path.

METHODOLOGY

We situate our discussion and illustration of the proposed methodology in a study of computer-supported, collaborative, problem-solving interactions. Bearing in mind that the goal of this paper is to advance a methodological framework, we briefly describe the context of the study in which the methodology was instantiated first before illustrating the process and usefulness of our methodology.

Research Context and Data Collection

Participants included sixty 11^{th} grade students (46 male, 14 female; 16-17 years old) from the science stream of a co-educational, English-medium high school in Ghaziabad, India. They were randomized into 20 groups of three and instructed to collaborate with their group members to solve two problem scenarios. Both presented an authentic car accident scenario that required the application of Newtonian kinematics. The study was carried out in the school's computer laboratory, where group members communicated with one another only through synchronous, text-only chat. The chat application allowed groups to privately and simultaneously engage in synchronous discussions and automatically archived the transcript of their discussion as a text file. These 20 transcripts, one for each group, contained the problem-solving interactions of group members as well as the final solutions produced by the groups and formed the data used in our analyses.

Data Coding: Categorizing Problem-solving Interactions

Quantitative Content Analysis (QCA) (Chi, 1997) was used to segment and code interactions using an interaction coding scheme developed by Poole and Holmes (1995), namely the Functional Category System (FCS) (see Table 1). Two trained doctoral students independently coded the interactions; inter-rater reliability was .85.

Table 1: Functional Category System (FCS)
(Adapted from Poole & Holmes (1995), p. 104)

1. Problem Definition (PD)
1a. *Problem Analysis*: Statements that define or state the causes behind a problem
1b. *Problem Critique*: Statements that evaluate problem analysis statements
2. Orientation (OO)
2a. *Orientation*: Statements that attempt to orient or guide the group's process.
2b. *Process Reflection*: Statements that reflect on or evaluate the group's process or progress
3. Solution Development (SD)
3a. *Solution Analysis*: Statements that concern criteria for decision making or general parameters for solutions
3b. *Solution Suggestion*: Suggestions of alternatives
3c. *Solution Elaboration*: Statements that provide detail or elaborate on a previously stated alternative.
3d. *Solution Evaluation*: Statements that evaluate alternatives and give reasons, explicit or implicit, for the evaluations.
3e. *Solution Confirmation*: Statements that state the decision in its final form or ask for final group confirmation of the decision.
4. Non-Task (NT)
Statements that do not have anything to do with the decision task. They include off-topic jokes and tangents
5. Simple Agreement (SA)
6. Simple Disagreement (SDA)

Table 2: Example of categorizing statements into interaction units and assigning impact values to them

Statements *(I0201, I0202, and I0203 represent the 3 group members)*	**Code**	**Impact**
I0201 DPS > physician says it was a considerable impact	1a 3a	1 1
I0203 DPS > yes as the limit was 25km/h	1a 3a	1 1
I0201 DPS > ranging bw 20g to 25g	3b	1
I0201 DPS > his medical reports r ok	1a	1
I0202 DPS > Mr rahul might have not been able to see the truck	3a 3b	1 -1
I0201 DPS > he wasn't under the influence of alcohol or drugs	3a 3b	1 1
I0201 DPS > it might be possible but he should have restricted himself to speed limit	3c 3c	1 -1
I0201 DPS > it was a blind turn	3a	-1
I0202 DPS > car has been severely struck	3c	1
I0203 DPS > he was not able to control the car i think	3b	-1
I0201 DPS > ya this proves that the impact was pretty hard and thus he was driving fast	5 3c 3c	-1 -1 -1

The *unit of analysis* was semantically (as opposed to syntactically) defined as the function(s) that an intentional statement serves in the problem-solving process. Therefore, every intentional problem-solving statement was segmented into one or more interaction unit(s) and coded into the functional categories of the FCS (see Table 2). We illustrate this with an example from our study.

The statement "*he wasn't under the influence of alcohol or drugs*" (highlighted in Table 2) was made by a participant during an interaction within a problem-solving group discussing a scenario involving a car accident. It serves two functions in the problem-solving process – first, it suggests a new parameter or criteria (intoxication, which had not been previously mentioned) for consideration in the solution and second, it asserts that the person being referred to (the driver in the problem scenario) was not intoxicated. Thus, despite being a single statement within an interaction, it contributes two units of analysis (hereafter referred to as interaction units) to two different functional categories. On the other hand, the statement "*he was not able to control the* *car I think*" serves the sole purpose of suggesting a possible factor in the solution to the problem. Hence, this statement contributed only one interaction unit. Therefore, by allowing statements to be coded into multiple interaction units, *atomicity* was achieved for the unit of analysis, i.e., interaction units can not be further divided into finer units. This, in turn, strengthens the choice of the unit of analysis (Barab et al., 2001). Bransford & Nitsch (1978) support the case for semantically-defined units by arguing that to fully comprehend a given interaction one must not only understand its words and the sentences (syntactic features), but also how it is situated in a discussion context. Chi (1997) further argued that it is often more meaningful to employ a semantic scheme, especially if it also provides for a greater correspondence between the grain size of the unit of analysis and the research questions of the study. Furthermore, recall that from the perspective of complex systems science, each interaction unit is a *functional operator* that reconfigures the problem state. Therefore, defining the unit of analysis as the *function(s)* that an intentional statement serves provides further logical correspondence with the theoretical lens used to conceptualize problem solving.

The result of coding the problem-solving interactions was a representation of each problem-solving discussion as a time-ordered sequence of functional categories or codes. Table 2 illustrates this sequence for a small sample of the coded interactions. We are now ready to describe and illustrate convergence, fitness, and persistence as measures for characterizing and analyzing the problem-solving process.

Convergence & Fitness

Convergence of problem-solving interactions may be broadly defined as the extent to which the group discussion leads to a solution as perceived by the group. To model the telic aim of problem-solving interactions and develop a measure for convergence, a two-state Markov model was used (Ross, 1996). An *a posteriori* impact value of 1, -1, or 0 was assigned to each interaction unit depending upon whether it pushed the group discussion towards (impact = 1) or away (impact = -1) from the goal of the activity - a solution state of the given problem, or maintained the status quo (impact = 0). This was done with an inter-rater reliability of .93 (see Table 2 for an example). More formally, let the problem space be defined by *n* interaction units; each assigned an impact value of 1, -1, or 0. Further, let n_1, n_{-1}, and n_0 denote the number of interaction units assigned the impact values 1, -1, and 0 respectively such that $n_1 + n_{-1} + n_0 = n$. Then convergence, $C(n)$, may be defined as:

$$C(n) = \frac{n_1 - n_{-1}}{n_1 + n_{-1}}$$

The number of zeros is not factored into the calculation of convergence. This is because interaction units assigned a zero impact, by definition, maintain the convergence level of the discussion. It is easy to see that the convergence value will always lie between -1 and 1. The closer the value is to 1, the higher is the convergence, and the closer the group is to reaching an ideal solution to the problem.

Note that the numerator in the formulation of $C(n)$ is a measure of position, $P(n) = n_1 - n_{-1}$. In other words, if the problem-solving process is a sequence of steps along a straight line - some forward (impact = 1) and others backward (impact = -1) - then the difference between the total number of forward and backward steps gives the position relative to (or distance from) the starting point, i.e., the start of the discussion. Convergence then is the mean distance from the starting point.

Convergence can also be conceptualized as measure of *fitness* of the entire discussion: the higher the convergence, the higher the fitness of the discussion. Extending this conceptualization to all problem states and not just the final one, we can define fitness as the temporal measure of convergence, i.e., at any point in time in the discussion, how close a group is to reaching the goal state – an ideal solution to the problem. Therefore, the fitness statistic at an arbitrary point in time in the problem-solving process is defined as the convergence value up to the interaction unit at that point in time, with the final fitness level of the entire problem-solving process being the convergence value itself. Recalling that time refers to *ticks* on the evolutionary clock (i.e. an arbitrary time *t* corresponds to, say, the i^{th} interaction unit), the fitness $F(t)$ at time *t* in the discussion may be defined as:

$$F(t) = C(t) = \frac{n_1(t) - n_{-1}(t)}{n_1(t) + n_{-1}(t)}$$

where $n_1(t)$ and $n_{-1}(t)$ represents the number of interaction units coded as 1 and -1 respectively, up to and including the i^{th} interaction unit. Plotting the fitness value on the vertical axis and time (as defined in this study) on the horizontal axis, one will get a representation (also called the fitness curve) of the problem-solving process as it evolves in time. Figures 1 and 2, drawn to the same scale, present four major types of fitness curves that emerged from the 20 problem-solving discussions in our study.

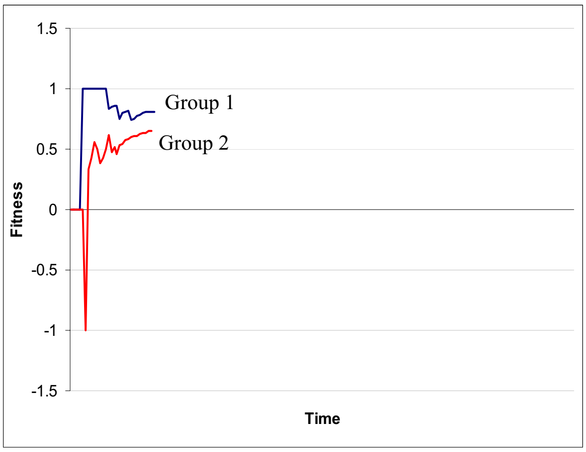

Figure 1. Fitness curves of two short discussions

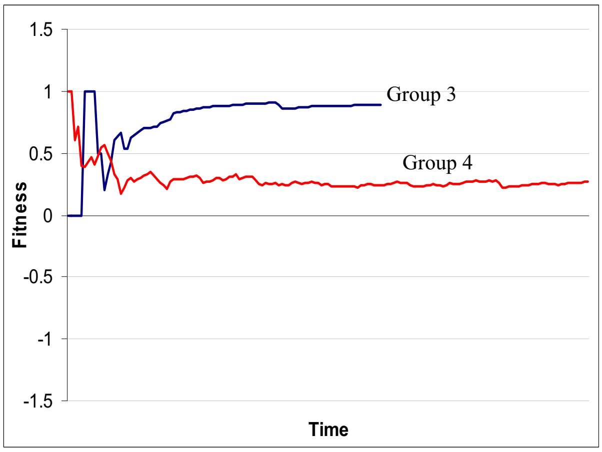

Figure 2. Fitness curves of two long discussions

Interpreting Fitness

In our view, there are five aspects to interpreting the fitness analysis. First, because the fitness value at a given time indicates proximity to an ideal solution (with higher values indicating closer proximity), fitness curves that trend upwards indicate problem-solving processes that are getting closer to an ideal solution (fitness = 1), and vice versa. Hence, fitness curves provide a quick snapshot of the entire problem-solving process in terms of how short or long it was as well as how close or far the discussion was from an ideal solution at any given point in time.

Second, the shape of the fitness curve is informative about the paths respective groups take toward problem solution. For example, groups 1 and 2 converged at approximately the same fitness levels (about 0.65, indicating positive movement toward an ideal solution), but their paths to this point were quite different. Group 1's discussion moved toward an ideal solution immediately when compared to group 2, whose initial approach seemingly took them away from the goal (indicated by the negative fitness initially) only to recover later. Similarly, comparing groups 3 and 4, we can see them settling into different plateaus of fitness albeit after some chaos (fluctuations in fitness levels) initially. Further, comparing groups 1 and 2 with groups 3 and 4, we can see that the discussions of groups 1 and 2 ended quickly whereas those of groups 3 and 4 settled into an “equilibrium” after the initial fluctuations. What is most interesting is that this interpretation of fitness curves provides a view of paths to a solution that are lost in analysis systems that consider only a given point in the solution process, thus assuming that similar behaviors or states at a given point are arrived at in similar ways. As different paths can lead to similar results, unidimensional analyses that consider only single points in time (often only the solution state) are not consistent with what we know about problem-solving processes and are not informative about movement toward a goal.

Third, the fitness curve of groups 3 and 4 also highlight the notion of “fitness inertia,” i.e., having settled into fitness equilibrium, these groups found it difficult to move in new directions. Of course, group 3 did not have a need to do so, as their high fitness value indicates movement toward an ideal solution. But implications of fitness inertia for groups that equilibrate at low fitness levels indicating no or very little movement toward higher fitness levels, such as what occurred with group 4, are grave. It follows from this that the eventual performance of groups exhibiting fitness inertia can be predicted early on in the discussion. Because our analyses showed convergence (and not the position) to be a significant predictor of group solution quality ($F = 50.245$, $p = .000$), it preliminarily suggests that the net number of positive steps (the position) is not as critical to the success of a discussion ($F = 0.012$, $p = .915$) as convergence is. This can be explained by the mathematical property a ratio, which is how convergence is operationalized, i.e., it is more sensitive to initial steps, both positive and negative, than steps that are taken later on in the process. Said another way, “good” contributions made earlier in a group discussion potentially do more good than if they were made later. Similarly, “bad” ones potentially do greater harm if they come earlier than later in the discussion. Hence, convergence takes into account not only the number of positive and negative steps (contributions), but also the order in which they are taken. This temporal order is perhaps what is missing in many studies of collaborative problem solving, which typically focus on the number of steps - both positive (such as frequency counts of higher-order thinking, questioning, etc.) and negative (such as frequency counts of errors, misconceptions, lack of cooperation, etc.) - as indicators of the quality of the discussion. Because convergence takes into account of both the number as well as the temporal order of the units of analyses, it utilizes greater amount of information present in the data, making it inherently a more powerful measure - both conceptually and statistically. Hence, as indeed our results preliminarily indicate, connections to learning and problem solving as evidenced by group solution quality are stronger when seen through the measure of convergence than through frequency measures commonly used in computer-supported collaborative learning (CSCL) research.

Fourth, the end-point of the fitness curve represents the final fitness level or convergence of the discussion. From this, the extent to which of a group was able to solve the problem can be deduced. In other words, we can deduce that, comparatively, group 3 did the best followed by group 1, group 2, and finally group 4. Furthermore, the final fitness levels can also be compared with the maximum fitness level of 1. One might imagine that an ideal fitness curve is one that has all the pushes in the right direction, i.e., a horizontal straight line with fitness equaling 1. However, the data suggests that, in reality, some level of divergence of ideas may in fact be a good thing. Note that, at present, one can only extract a comparison either between groups or with the upper and lower bounds of fitness (1 or -1). But, with repeated application in other research contexts and settings and over multiple studies, norms for absolute values of convergence and fitness will begin to emerge.

Finally, based on the above analysis of the characteristics of fitness curves and what they tell us about the problem-solving process, we can begin to conceptualize how collaborative, problem-solving discussions may be scaffolded to achieve optimal outcomes. For example, the fitness curves of groups 2 and 4 suggest a need for scaffolding early on in the discussion.

Persistence

In addition to looking at the fitness characteristics of a discussion as a whole, one can also examine how ideas or families of ideas emerge and persist during the course of the problem-solving discussion. In our study, these families of ideas are represented by the 6 major functional categories - problem definition, orientation, solution development, non-task, simple agreement, and simple disagreement (see Table 1) – into which all interactions were categorized. Treating each functional category as a component of the problem-solving system, its *usage* (or *persistence*) can be tracked as a measure of evolutionary activity. The central assumption is that components of a complex system that persist and continue to be used make greater contribution to the system. However, nothing is implied about the quality of that contribution. Equivalently, functional categories that persist and get used repeatedly make greater contribution to the collaborative, problem-solving activity. Therefore, it makes sense to choose these taxonomic functional categories as components because adaptive evolution of the problem-solving process can be expected to affect the dynamics of these categories and conversely, the dynamics of these categories inform the evolutionary activity structures of the problem-solving process (Bedau et al., 1998). Having established functional categories as the components of the problem-solving system, their contribution can be measured by their usage; the idea being that the longer a functional category persists in a system, the greater its adaptive value,. Conversely, by examining the persistence of functional categories, we can gain insights into the problem-solving process that would otherwise remain elusive. More formally, let $f_k(t)$ denote whether the k^{th} functional category exists in the problem-solving system at time *t*:

$$f_k(t) = \begin{cases} 1 & \text{if component } k \text{ exists at time } t \\ 0 & \text{otherwise} \end{cases}$$

$f_k(t)$ is simply an *activity indicator function* that “switches on” each time an interaction unit belonging to a particular functional category exists in the discussion. In order to measure the usage of a functional category, we can define a corresponding function – an *activity incrementation function* – that increases by 1 each time the indicator function “switches on.” Then, the value of the incrementation function for the k^{th} functional category at time *t*, say $a_k(t)$, reflects its cumulative *usage* up until time *t*, i.e., the persistence of the functional category up until time *t*. Formally,

$$a_k(t) = \sum_0^t f_k(t)$$

Figure 3 shows the persistence curve of the problem definition and solution development functional categories for two groups. We decided to illustrate persistence using these two categories because they had the most manifest interactions compared to the other four categories.

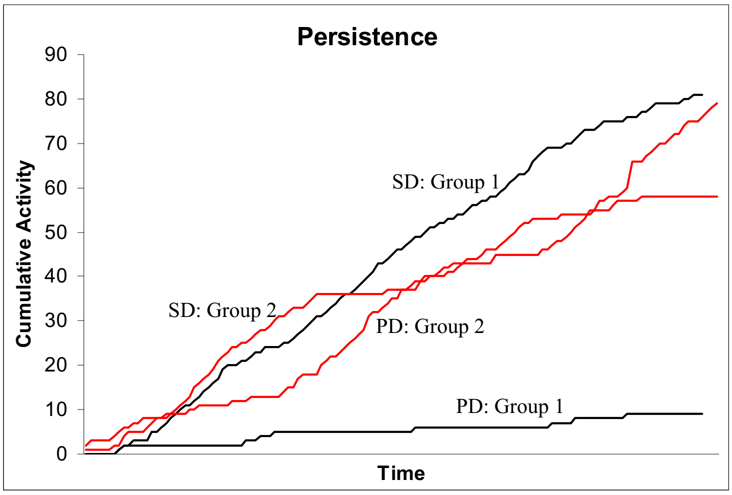

Figure 3. Persistence curves of Problem Definition (PD) & Solution Development (SD) functional categories

Interpreting Persistence

First, being a cumulative function, it is a non-decreasing curve whose end-point indicates the total activity in a given functional category, i.e., the number of interaction units in that functional category. Often, it is this number that is used as a frequency measure in quantitative content analysis. However, the number alone does not indicate anything about the evolutionary activity of the functional category it represents because, as argued earlier, it does not utilize the temporal information embedded in the data. Persistence curves utilize that information and provide a trajectory from which meaningful insights may be drawn.

Second, a plateau on the persistence curve of a functional category indicates a period in a discussion where no interactions of the type that the functional category represents take place. Therefore, persistence curves that plateau often and for long periods are indicative of a passive functional category. Similarly, a persistence curve that does not plateau is indicative of an active functional category. For example, the problem definition (PD) functional category for group 1 is an example of a passive functional category whereas the PD functional category for group 2 is an active one. In other words, this suggests that group 1 either did not see the need to define the problem or was able to define it quickly and move on, whereas group 2 seemed to need much more time and discussion for problem definition. Note that, in either case, this does not indicate whether or not the problem definition was correct, which can be revealed by cross-validating persistence curves with fitness curves.

Third, persistence curves bring out the notion of competition among functional categories. For group 1, only the SD functional category is active whereas both PD and SD functional categories are active for group 2. This suggests that the problem-solving process was by and large linear for group 1: they defined the problem early on and then worked on developing a solution. There was little or no competition between the PD and SD functional categories. However, the process was quite the opposite for group 2: their attempts to define the problem and develop a solution were iterative and intermingled making the process non-linear and chaotic. There was high competition between the two functional categories. At this point, it is difficult to use the level of competition to make inferences about the quality of the discussion or the resulting solution. However, repeated application in other research contexts and settings and over multiple studies will provide greater validity for the inferences.

USEFULNESS AND LIMITATIONS

Coding Reliability and Validity

The inferences that can be drawn from the new measures are strong in so far as the coding scheme is reliable and valid. In this study, we opted to use an existing coding scheme, namely the functional category system (FCS) developed by Poole and Holmes (1995). The reasons for choosing the FCS as the interaction coding protocol for this study are:

i. The FCS was developed specifically for the purpose of studying small-group collaborative interactions in problem-solving contexts.

ii. The FCS categories are theoretically well-grounded in the cognitive and educational theories of problem solving thereby increasing their content validity.
iii. The FCS has been tried and tested in several research studies (for example, Poole & Holmes, 1995; Jonassen & Kwon, 2001) making it inherently more reliable and stable than developing an entirely new coding scheme (Gall, Borg, & Gall, 1996).
iv. From a broader perspective of research design and measurement, using a pre-existing interaction coding scheme adds to the validity of the inferences drawn from the results (Rourke & Anderson, 2004).
This gives us reason to trust the reliability and validity of our quantitative content analysis using the FCS.

Usefulness of the Methodology

A major advantage of the proposed methodology is that it takes into account temporality of the problem-solving process and extracts measures that utilize that information. As such, when compared with existing measures commonly used in CSCL research, our research suggests that the measures of convergence and persistence are potentially more powerful in characterizing and tracking the evolution of problem-solving processes and how they lead and relate to outcomes. Further, while we situated our illustration and discussion of the proposed methodology in the collaborative problem-solving efforts of science students mediated by an online, synchronous environment, the methodology can easily be applied to other settings and contexts. We argue that the proposed methodology would be applicable to the analysis of any process that is a) goal-directed, b) complex and adaptive, and c) well-manifested through rich and meaningful artifacts (which we broadly define to include not only physical behaviors, actions, and products but also conceptual artifacts such as concepts and ideas). As such, the methodology may be applied to individual or collaborative problem-solving, in domains other than physics, with other populations, in a modality other than online, synchronous chat, and using other categorization coding schemes.

Limitations

As with any new methodology, its repeated application and modification over multiple data sets is needed before strong and valid inferences about the underlying cognitive processes can be made (Rourke & Andersen, 2004). Another limitation includes the requirement of capturing rich and meaningful data in which there is ample opportunity for evolutionary structures and goal-seeking adaptations to occur. In our study, we ensured this by making the objects of the activity – the problems – rich in context. Also, we did not impose any time limit on the group discussions. While capturing the data was made easy due to the technology itself, analyzing the data was time consuming. As such, this approach is a useful analytical framework for education researchers but not for classroom teachers. However, inferences drawn by researchers while using our methodology will have meaningful implications for the classroom, especially with regard to the design and scaffolding of instruction and learning environments for problem-solving tasks.

FUTURE DIRECTIONS

As we move forward, we plan to apply the proposed methodology in other contexts and settings. From a repeated application and modification of the measures over multiple data sets will emerge indications of the validity and reliability of the proposed methodology. In turn, this will lead to fine-tuning of the measures in an iterative fashion.

Concomitantly, we also see the need for developing new measures, especially at a macroscopic level of analysis. In particular, we will focus on stable interaction phases that a discussion goes through. In other words, a problem-solving discussion can be conceptualized as a temporal sequence of phases. One can use several methods to isolate evolutionary phases, including measures of genetic entropy (Adami, Ofria, & Collier, 2000), intensity of mutation rates (Burtsev, 2003) or, in the case of problem interactions, the classification of coherent phases of interaction. Whether these phases involve genetic mutations or problem interactions, sequences of phases often alternate between stable phases, with chaotic phases interspersed throughout: these often correspond to low vs. high mutation rates, clustered vs. unclustered interactions. With the phases identified, one can calculate and predict the probabilities of moving from one phase to another using Hidden Markov Models (HMM). As a result, one may begin to understand when and why phase transitions, cascades and catastrophes (sudden mass change), as well as stable phases emerge; more importantly, one may begin to understand how the configuration of one phase may influence the likelihood of moving to any other phase. Whether one can control or temper these phases, or whether such control or temperance would prove a wise practice remains an open question which, even if only partially answered, will be a major breakthrough in characterizing and modeling the problem solving process.

Through such an endeavor, education researchers who wish to study the problem solving process will find choices among several lenses at several resolutions. With measures to analyze number, function, fitness, sequencing, and transition of states, as well as the evolutionary activity of components (functional categories), one can zoom from the micro- to macroscopic properties and behaviors of the problem-solving process.

REFERENCES

Adami, C., Ofria, C., Collier, T. C. (2000). Evolution of biological complexity. Proceedings of the National Academy of Sciences, 97, 4463-4468.

Akhras, F. N., & Self, J. A. (2000). Modeling the process, not the product, of learning. In S. P. Lajoie, *Computers as cognitive tools, volume two: No more walls* (pp. 3-28). Mahwah, NJ: Lawrence Erlbaum Associates.

Barab, S. A., Hay, K. E., & Yamagata-Lynch, L. C. (2001). Constructing networks of action-relevant episodes: An in-situ research methodology. *The Journal of the Learning Sciences*, *10*(1&2), 63-112.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*(3), 307-359.

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Perseus Publishing: New York.

Bedau, M. A., Snyder, E., & Packard, N. H. (1998). A classification of long-term evolutionary dynamics. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Artificial Life VI* (pp.189-198), Cambridge: MIT Press

Bransford, J.D., & Nitsch, K. E. (1978). Coming to understand things we could not previously understand. In J. F. Kavanaugh and W. Strange (Eds.), *Speech and language in the laboratory, school, and clinic*. Harvard, MA: MIT Press.

Burtsev, M. S. (2003). Measuring the dynamics of artificial evolution. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in Artificial Life*. Proceedings of the 7th European Conference on Artificial Life, Dortmund, Germany, September 14-17.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*(3), 271-315.

Collazos, C., Guerrero, L., Pino, J., & Ochoa, S. (September, 2002). Evaluating collaborative learning processes. Proceedings of the 8th International Workshop on Groupware (CRIWG'2002), Springer Verlag LNCS, 2440, Heidelberg, Germany.

Collazos, C., Guerrero, L.,Pino, J., & Ochoa, S. (2004). A method for evaluating computer-supported collaborative learning processes. *International Journal of Computer Applications in Technology*, *19*(3/4), 151-161.

Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics, and induction. *Physica D*, *75*, 11-54.

Derry, S. J., Gance, S., Gance, L. L., & Schlager, M. (2000). Toward assessment of knowledge-building practices in technology-mediated work group interactions. In S. P. Lajoie (Eds.), *Computers as cognitive tools, volume two: No more walls* (pp. 29-68). Mahwah, NJ: Lawrence Erlbaum Associates.

Gall, M., Borg, W., & Gall, J. (1996). *Education research: An introduction*. White Plains, NY: Longman.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*(6), 1420-1443.

Heylighen F. (1988). Formulating the problem of problem-formulation. In R. Trappl (Ed.), *Cybernetics and Systems '88* (pp. 949-957). Dordrecht, Netherlands: Kluwer Academic Publishers.

Jonassen, D. H. (2000). Towards a design theory of problem solving. *Educational Technology, Research and Development*, *48*(4), 63-85.

Jonassen, D. H. & Kwon, H. I. (2001). Communication patterns in computer-mediated vs. face-to-face group problem solving. *Educational Technology, Research and Development*, *49*(1), 35-52.

Leont'ev, A. N. (1978). *Activity, consciousness, and personality*. NJ: Prentice-Hall.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Poole, M. S., & Holmes, M. E. (1995). Decision development in computer-assisted group decision making. *Human Communications Research*, *22*(1), 90-127.

Ross, S. M. (1996). *Stochastic processes*. New York: John Wiley & Sons, Inc.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology, Research and Development*, *52*(1), 5-18.

Schwartz, D., Lin, X. D., Brophy, S., & Bransford, J. D. (1999). Toward the development of flexibly adaptive instructional designs. In C. M. Reigeluth (Eds.), *Instructional design theories and models: A new paradigm of instructional theory* (pp. 183-214). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.