Re-conceptualizing Mathematics Education as a Design Science

Richard Lesh and Bharath Sriraman

In this chapter we propose re-conceptualizing the field of mathematics education research as that of a design science akin to engineering and other emerging interdisciplinary fields which involve the interaction of "subjects", conceptual systems and technology influenced by social constraints and affordances. Numerous examples from the history and philosophy of science and mathematics and ongoing findings of M&M research are drawn to illustrate our notion of mathematics education research as a design science. Our ideas are intended as a framework and do not constitute a "grand" theory. That is, we provide a framework (a system of thinking together with accompanying concepts, language, methodologies, tools, and so on) that provides structure to help mathematics education researchers develop both models and theories, which encourage diversity and emphasize Darwinian processes such as: (a) selection (rigorous testing), (b) communication (so that productive ways of thinking spread throughout relevant communities), and (c) accumulation (so that productive ways of thinking are not lost and get integrated into future developments).

A Brief History of Our Field

Mathematics education is still in its "infancy" as a field of scientific inquiry. This is evident in the fact that the first journals devoted purely to research only started appearing in the 1960's, prominent among which were the *Zentralblatt für Didaktik der Mathematik* (ZDM) and *Educational Studies in Mathematics* (ESM). In the early 1970's, there was an explosion of new journals devoted to research—including *the Journal for Research in Mathematics Education* (JRME) and the *Journal für Mathematik Didaktik* (JMD). Until this time period we had no professional organization for researchers; and, we had few sharable tools to facilitate research.

Erich Ch. Wittmann has authored several papers with the same title and has been instrumental in proposing the design science approach to mathematics education.

R. Lesh
School of Education, Indiana University, Bloomington, USA
e-mail: ralesh@indiana.edu

B. Sriraman (✉)
Department of Mathematical Sciences, The University of Montana, Missoula, USA
e-mail: sriramanb@mso.umt.edu

B. Sriraman, L. English (eds.), *Theories of Mathematics Education,*
Advances in Mathematics Education,
DOI 10.1007/978-3-642-00742-2_14,

Arguably there were journals such as *the l'Enseignment Mathematique* (founded in 1899 in Geneva), *The Mathematics Teacher* (founded in 1901 by the NCTM) and *The Mathematical Gazette* (founded in 1894 in the UK), and the *Zeitschrift für Mathematischen und Naturwissenschaftlichen Unterricht* (founded in 1870 in Germany) all of which were supposed to address the teaching and learning of mathematics. However, a survey of the papers appearing in these journals suggests that few were aimed at advancing what is known about mathematics problem solving, leaning, or teaching. A detailed history of the birth of journals worldwide is found in Coray et al. (2003), which the interested reader is urged to look up.

For the purpose of our discussion, research as we mean it today only started in the 1960's and depended mainly on theory borrowing (from other fields such as developmental psychology or cognitive science). We really had no stable research community—with a distinct identity in terms of theory, methodologies, tools, or coherent and well-defined collections of priority problems to be addressed. Only recently have we begun to clarify the nature of research methodologies that are distinctive to our field (Biehler et al. 1994; Bishop et al. 2003; Kelly and Lesh 2000; Kelly et al. 2008; English 2003); and, in general, assessment instruments have not been developed to measure most of the constructs that we believe to be important. These facts tend to be somewhat shocking to those who were not firsthand witnesses to the birth of our field or familiar with its history—and whose training seldom prepares them to think in terms of growing a new field of inquiry.

One of the most important challenges that nearly every newly evolving field confronts is to develop a sense of its own identity and the inability of our field to do so has been a source of criticism (Steen 1999). Should mathematics education researchers think of themselves as being applied educational psychologists, or applied cognitive psychologists, or applied social scientists? Should they think of themselves as being like scientists in physics or other "pure" sciences? Or, should they think of themselves as being more like engineers or other "design scientists" whose research draws on multiple practical and disciplinary perspectives—and whose work is driven by the need to solve real problems as much as by the need to advance relevant theories? In this chapter, we argue that mathematics education should be viewed as being a design science.

What Is a Design Science?

The following characteristics of "design sciences" are especially relevant to mathematics education.

(a) The "Subjects" being Investigated tend to be Partly Products of Human Creativity. Unlike physics and other natural sciences (where the main "subjects" being investigated were on-the-scene before the dawn of human history, and have not changed throughout human history), the most important "subjects" (or systems) that design scientists need to understand and explain tend to be partly or entirely designed, developed, or constructed by humans. For example, in mathematics education, these "subjects" range from the conceptual systems that we try to help students

or teachers develop, to the ways of thinking that are embodied in curriculum materials or innovative programs of instruction.

(b) The "Subjects" being Investigated are (or Embody) Complex Systems.[1] In engineering and biology, such systems often are visible in the design documents that are developed for artifacts such as space shuttles, skyscrapers, growth models and computer information processing systems; and, in mathematics education, similar systems sometimes can be seen in the design documents that describe when, where, why, how and with whom curriculum materials or programs of instruction need to be modified for use in a variety of continually changing situations. However, in mathematics education, it also may be the case that attention focuses on the development of conceptual systems themselves, rather than on artifacts or tools in which these conceptual systems may be expressed. Thus, there are two basic types of situations where "design research" is especially useful.

- *Attention focuses on concrete artifacts or tools*. In these cases, the researcher may want to develop (and/or study the development of) resources that can be used to support teaching, learning, or assessment. But, in both engineering and education, complex artifacts and conceptual tools seldom are worthwhile to develop if they are only intended to be used a single time, by a single person, for a single purpose, in a single situation. In general, high quality products need to be sharable (with others) and reusable (in a variety of continually changing situations). So, they need to be modularized and in other ways made easy to modify. This is one reason why underlying design principles are important components of the artifact + design that needs to be produced.
- *Attention focuses on conceptual systems*. In these cases, the researcher may want to develop (and/or study the development of) some complex conceptual system which underlies the thinking of student(s), teacher(s), curriculum developer(s) or some other educational decision maker(s). But, in order to develop useful conceptions about the nature of relevant conceptual systems, the "subjects" need to express their thinking in the form of some thought-revealing artifact (or conceptual tool), which goes through a series of iterative design cycles of testing and revision in order to be sufficiently useful for specified purposes. In this way, when the artifact is tested, so are the underlying conceptual systems; and, an auditable trail of documentation tends to be generated that reveals significant information about the evolving ways of thinking of the "subject". We can draw on the evolution of mathematics to show evolving ways of thinking of the community over time. The evolution of mathematics reveals the series of iterative designs what artifacts went through before the dawn of symbolism. Moreno and Sriraman (2005) have argued that human evolution is coextensive with tool development and reveals the series of iterative designs, which artifacts undergo over time. They write:

> Take the example of a stone tool: The communal production of those tools implied that a shared conception of them was present. But eventually, somebody could discover a

[1] Here, the term "complex" is being interpreted close to the mathematical sense of being a system-as-a-whole which has emergent properties that cannot simply be deduced from properties of elements (or agents) within the system.

> new use of the tool. This new experience becomes part of a personal reference field that re-defines the tool for the discoverer; eventually, that experience can be shared and the reference field becomes more complex as it unfolds a deeper level of reference…thus, tool production was not only important for plain survival, but also for broadening the mental world of our ancestors–and introducing a higher level of objectivity.

(c) Researchers should DESIGN for Power, Sharability and Reusability; They don't just TEST for It. Survival of the useful is a main law that determines the continuing existence of innovative programs and curriculum materials; and, usefulness usually involves going beyond being powerful (in a specific situation and for a specific purposes) to also be sharable (with other people) and re-usuable (in other situations). … What is the half-live of a good textbook or course syllabus? Truly excellent teachers continually make changes to the materials that they use; and, truly excellent curriculum materials are designed to be easy to modify and adapt to suit the continually changing needs of students in specific courses. So, even if the teacher wrote the book that is being used in a given course, significant changes tend to be made each year – so that the materials often are nearly unrecognizable after only a few years.

(d) The "Subjects" to be Understood are Continually Changing—and so are the Conceptual Systems needed to Understand and Explain Them. One reason this is true is because the conceptual systems that are developed to make sense of a relevant systems also are used to make, mold, and manipulate new system. Therefore, as soon as we understand a system, we tend to change it; and, when we change it, our understandings also generally need to evolve.

(e) "Subjects" being Investigated are Influenced by Social Constraints and Affordances. Design "specs" for the systems (and accompanying artifacts or tools) that engineers develop are influenced as much by peoples' purposes as by physical or economic aspects of the contexts in which they are used. Therefore, because peoples' purposes continually change, and because people often use tools and artifacts in ways that their developers never imagined, the artifacts and tools tend to change as they are used. For examples of this phenomenon, consider personal computers, microwave ovens, and sport utility vehicles. Similarly, in mathematics education, the nature of artifacts (software, curriculum materials, instructional programs) is influenced as much by socially generated capital, constraints, and affordances as by the capabilities of individuals who created them—or by characteristics of the contexts in which they originally were designed to be used.

(f) No Single "Grand Theory" is likely to Provide Realistic Solutions to Realistically Complex Problems. This claim is true even for the "hard" sciences. In realistic decision-making situations that involve the kind of complex systems that occur in engineering and mathematics education, there almost never exist unlimited resources (time, money, tools, consultants). Furthermore, relevant stake holders often have partly conflicting goals (such as low costs but high quality). Therefore, in such situations, useful ways of thinking usually need to integrate concepts and conceptual systems drawn from more a single practical or disciplinary perspective. Most will need to involve models which integrate ways of thinking drawn from a variety of theories and practices.

(g) Development Usually Involves a Series of Iterative Design Cycles. In order to develop artifacts + designs that are sufficiently powerful, sharable, and reusable, it usually is necessary for designers to go through a series of design cycles in which trial products are iteratively tested and revised for specified purposes. Then, the development cycles automatically generate auditable trails of documentation which reveal significant information about the products that evolve. The birth of numerical analysis and analysis of algorithms as domains of applied mathematics research provides numerous examples of the revision of historical products for use today. Consider the Archimedean technique for approximating the value of π that relies on the two lemmas (traceable to Proposition 3 of *The Elements Book VI*). Archimedes essentially inscribed and circumscribed the circle with regular polygons up to 96 sides to compute an approximation for π. Traditional history of mathematics courses typically involve the exercise of employing the algorithms outlined in Lemma 1 and Lemma 2 to hand calculate the value of π. This exercise leads one to the realization of the superb computational abilities Archimedes must have possessed! However, the necessity of a 21st century computational tool becomes very obvious when one analyzes the computational complexity of Archimedes' algorithm. It is clear that each step of this algorithm requires taking an additional square root, which was dealt by Archimedes via the use of a "magical" rational approximation. It was magical in the sense that it required knowing how to compute square roots in that period, something Archimedes never explicitly revealed. Comparing the computational efficiency (or inefficiency) of the Archimedean technique to that of modern recursion techniques is a very useful mathematical exercise. The computational inefficiency becomes obvious when one sees that a nine-digit approximation of π requires 16 iterations and requires a polygon of 393216 sides! Extensions of the Archimedean algorithm include generating a class of geometric figures to which the technique would be applicable and result in an approximation of a related platonic constant. Besides the domain of approximation techniques and computation, there is an abundance of problems in the history of mathematics that reveal the need for the continual creation of better and powerful abstract and computational tools.

The arguments made in (a)–(g) in our view of mathematics education research as a design science also parallel the mutation of methodological perspectives in the history of science. Modern science, especially the progression of research in physics and biology reveals that learning is a complex phenomenon in which the classical separation of subject, object, and situation is no longer viable. Instead, reality is characterized by a "non-linear" totality in which the observer, the observed, and the situation are in fact inseparable. Yet, at the dawn of the 21st century, researchers in our field are still using theories and research methodologies grounded in the information-processing premise that learning is reducible to a list of condition-action rules. While physicists and biologists are involved in the study of complex systems (in nature) via observation, experiment, and explanation, design scientists are involved in studying and understanding the growth of knowledge that occurs when students, teachers and researchers are confronted with problem situations involving making sense of complex situations. Complex systems are those which involve numerous elements, "arranged in structure(s) which can exist on many scales. These go through processes of change that are not describable by a single rule nor are reducible to only one level of explanation … these levels often include features whose emergence cannot be predicted from their current specifications" (Kirschbaum). In other words scientists today have embraced a view of nature in which processes have supplanted things in descriptions and explanations and reaffirmed the dynamic nature of the "whole" reflected in paradoxes encountered by ancient cultures. For instance biologists have found that methodological reductionism, that is going to the parts to understand the whole, which was central to the classical physical sciences, is less applicable when dealing with living systems. According to the German molecular biologist, Friedrich Cramer, such an approach may lead to a study not of the 'living' but of the 'dead', because in the examination of highly complex living systems "Only by ripping apart the network at some point can we analyze life. We are therefore limited to the study of 'dead' things."[2]

Analogously, the challenge confronting design scientists who hope to create models of the models (and/or underlying conceptual systems) that students, teachers and researchers develop to make sense of complex systems occurring in their lives is: the mismatch between learning science theories based on mechanistic *information processing* metaphors in which everything that students know is methodologically reduced to a list of condition-action rules, given that characteristics of complex systems cannot be explained (or modeled) using only a single function—or even a list of functions. As physicists and biologists have proposed, characteristics of complex systems arise from the *interactions* among lower-order/rule-governed agents—which function simultaneously and continuously, and which are not simply inert objects waiting to be activated by some external source.

Observations about Mathematics Education as a Distinct Field of Scientific Inquiry

Mathematics education research often is accused of not answering teachers' questions, or not addressing the priority problems of other educational decision-makers. …If this claim is true, then it surely is not because of lack of trying. Most mathematics education researchers also ARE practitioners of some type—for example, expert teachers, teacher developers, or curriculum designers. But: *When you're up to your neck in alligators, it's difficult to think about draining the swamp!* This is why, in most mature sciences, one main purpose of research is to help practitioners ask better questions—by focusing on deeper patterns and regulations rather than to surface-level pieces of information. Furthermore, the challenge to "solve practitioners' problems" ignores the fact that very few realistically complex problems are going to be solved by single isolated studies. In a survey of the impact of educational research on mathematics education, Wiliam (2003) outlines the two major

[2]Friedrich Cramer (1993): *Chaos and Order*, VCH Publishers, New York, 214.

"revolutions" in mathematics education in the recent past with the caveat that such a characterization may not be universally true given the heterogeneity of changes within different nations. However the two revolutions he mentions apply well to the United States. These are the "technological revolution" and the "constructivist revolution". The canon of studies within the former reveal the mismatch between research and practice. While specific site-based studies reveal the success of integrating technology in the teaching of mathematics, in general this remains untrue. The second revolution has resulted in "we are all constructivists now" (Wiliam 2003, p. 475). However, the tiny islands where classroom practice is "constructivist" and often reported by research, are by and large surrounded by oceans of associationist tendencies.

As we have suggested earlier single isolated studies seldom result in any large scale changes. In general, such problems will require multiple researchers and practitioners, representing multiple perspectives, and working at multiple sites over extended periods of time. This is why, in mature sciences, researchers typically devote significant amounts of time and energy to develop tools and resources for their own use. In particular, these tools and resources generally include instruments for observing and assessing "things" that are judged to be important; and, they also include the development of productive research designs, language, operational definitions of key concepts, and theory-based and experience-tested models for explaining complex systems. In particular the role of operational definitions needs to be critically examined for theories that purport to explain complex systems. In science, the role of operational definitions is to reach agreement on terms used based on a series of measurements which can be conducted experimentally. In spite of the popular misconception of the "iron-clad" nature of definitions in the physical sciences, it is important to realize that even "physical" concepts are by and large dependent on mutually agreed upon quantification. For instance the operational definition of an "electron" is "a summary term for a whole complex of measurables, namely 4.8×10^{-10} units of negative charge, 9.1×10^{-23} grams of mass etc." (Holton 1973, p. 387). Now imagine the difficulty of reaching agreement on operational definitions in quantum mechanics where the difficulty is compounded by paradoxes arising when the state of a "system" is dependent on the observer, who simplistically speaking, destroys the state in order to make a measurement. At the sub-atomic level measurements of position, momentum, etc are also not independent of one another. In spite of these profound difficulties "physicists have learned that theoretical terms have to be defined operationally, that is they have to describe nature via theories in which terms are accepted only if they can be defined/backed up via experimentation" (Dietrich 2004). The question we pose (at this stage philosophically) is: how can similar approaches be adapted by design scientists? Before defining theoretical terms, we should first attempt to gain consensus on "observational" terms. That is, how can we operationally define observational terms (namely perceived regularities that we attempt to condense into theories, or as Piaget attempted—to phylogenetically evolved mental cognitive operators)?

Operational definitions are routinely used in physics, biology, and computer science. As we mentioned earlier, in quantum mechanics, physicists are able to define (philosophically intangible) sub-atomic phenomenon by making predictions about their probability distributions. It is important to note that physicists do not assign a definite value per se to the observable phenomenon but a probability distribution. The implication for design scientists is that the notion of operational definitions can be adapted to the study of modeling by making predictions on the range of observable "processes" that students will engage in when confronted by an authentic model eliciting situation and the range of conceptual systems emerging from this engagement. Unlike psychology which has tried to operationally define intangible and controversial constructs such as intelligence, supposedly measurable by an IQ score our goal (analogous to physics) ought to be to operationally define *tangible constructs* relevant to the learning sciences, in terms of a distribution of clearly observable processes and conceptual systems within the specific model eliciting situation (see Lesh and English 2005 for further details). In this respect we preserve the whole by not attempting to measure each individual process and adhere to John Stuart Mill's wise suggestion that we move away from the belief that anything that is nameable should refer to a "thing". We later use the example of a double pendulum to demonstrate the shortcoming of traditional approaches to researching learning in mathematics education.

Preliminary Implications for Mathematics Education

In mathematics education, very few research studies are aimed at developing tools that build infrastructure (so that complex problems can be solved in the long run); and, our funding agencies, professional organizations, research journals, and doctoral education have largely ignored their responsibilities to build infrastructure—or to support those who wish to try. In fact, they largely emphasize simplistic "quick fix" interventions that are precisely the kind practitioners do NOT need.

The USA's Department of Education says: "*Show us what works!!!*" … Yet, when discussing large and complex curriculum innovations, it is misleading to label them "successes" or "failures"—as though everything successful programs did was effective, and everything unsuccessful programs did was not effective. In curriculum development and program design, it is a truism that: *"Small treatments produce small effects; and, large treatments do not get implemented fully."* "*Nothing works unless you make it work!*" … Consequently, when developing and assessing curriculum innovations, it is not enough to demonstrate THAT something works; it also is important to explain WHY and HOW it works, and to focus on interactions among participants and other parts of the systems. This is why the underlying design (which describes intended relationships and interactions among parts of the relevant systems) is one of the most important components of any curriculum innovation that is designed; and, it is why useful designs are those that are easy to modify and adapt to continually changing circumstances. So, in successful curriculum innovations, modularity, modifiability and sharability are among the most important characteristics to design in—and assess.

All programs have profiles of strengths and weaknesses; most "work" for achieving some types of results but "don't work" for others; and, most are effective for some students (or teachers, or situations) but are not effective for others. In other words, most programs "work" some of the time, for some purposes, and in some circumstances; and, none "work" all of the time, for all purposes, in all circumstances. So, what practitioners need to know is when, where, why, how, with whom, and under what circumstances are materials likely to work. For example: When the principal of a school doesn't understand or support the objectives of a program, the program seldom succeeds. Therefore, when programs are evaluated, the characteristics and roles of key administrators also should be assessed; and, these assessments should not take place in a neutral fashion. Attempts should be made to optimize understanding and support from administrators (as well as parents, school board members, and other leaders from business and the community); and, during the process of optimization, auditable documentation should be gathered to produce a simple-yet-high-fidelity trace of continuous progress.

The success of a program depends on how much and how well it is implemented. For example, if only half of a program is implemented, or if it is only implemented in a half-hearted way, then 100% success can hardly be expected. Also powerful innovations usually need to be introduced gradually over periods of several years. So, when programs are evaluated, the quality of the implementation also should be assessed; and, again, this assessment should not pretend to be done in a neutral fashion. Optimization and documentation are not incompatible processes. In fact, in business settings, it is considered to be common knowledge that "*You should expect what you inspect*!" ... In other words, all assessments tend to be self-fulfilling. That is, they are powerful parts of what educational testing enthusiasts refer to as "treatments".

Similar observations apply to teacher development. It is naive to make comparisons of teachers using only a single number on a "good-bad" scale (without identifying profiles of strengths and weaknesses, and without giving any attention to the conditions under which these profiles have been achieved, or the purposes for which the evaluation was made). No teacher can be expected to be "good" in "bad" situations (such as when students do not want to learn, or when there is no support from parents and administrators). Not everything "experts" do is effective, and not everything "novices" do is ineffective. No teacher is equally "experienced" across all grade levels (from kindergarten through calculus), with all types of students (from the gifted to those with physical, social, or mental handicaps), and in all types of settings (from those dominated by inner-city minorities to those dominated by the rural poor). Also, characteristics that lead to success in one situation often turn out to be counterproductive in other situations. Furthermore, as soon as a teacher becomes more effective, she changes her classroom in ways that require another round of adaptation. So, truly excellent teachers always need to learn and adapt; and, those who cease to learn and adapt often cease to be effective. ... Finally, even though gains in student achievement should be one factor to consider when documenting the accomplishments of teachers (or programs), it is foolish to assume that great teachers always produce larger student learning gains than their less great colleagues. What would happen if a great teacher chose to deal with only difficult students or difficult circumstances? What would happen if a great teacher chose to never deal with difficult students or difficult circumstances?

In virtually every field where researchers have investigated differences between experts and novices, it has become clear that experts not only DO things differently, but they also SEE (or interpret) things differently; and, this applies to student development as well as to teacher development or program development. Consequently, when we assess student development, we should ask *What kind of situations can they describe (or interpret) mathematically?* at least as much as we ask *What kind of computations can they do*? … Thinking mathematically involves more than computation; it also involves mathematizing experiences—by quantifying them, by coordinatizing them, or by making sense of them using other kinds of mathematical systems. Therefore, if researchers wish to investigate the nature of students' mathematical sense-making abilities, then they generally need to focus on problem solving situations in which interpretation is not trivial; and, this creates difficulties for simple-minded studies aimed at showing what works.

Most modern theories assume that interpretation is influenced by *both* (internal) conceptual systems and by (external) systems that are encountered; and, this implies that:

- Two students who encounter the same task may interpret it quite differently. So: *What does it mean to talk about a "standardized" task?*
- In non-trivial tasks that involve interpretation, several levels and types of descriptions always are possible. So, tasks that involve simple right-wrong answers are unlikely to involve significant types of interpretation.
- In a series of tasks in which similar interpretations need to be developed, the very act of developing an interpretation of early tasks implies that the nature of later tasks will change. So: *What does it mean to talk about "reliability"—if this means that repeated measures should yield the same results?*

In general, when assessment shifts attention beyond computation (and deduction) toward interpretation (and communication), then a phenomena occurs that is reminiscent of Heiserberg's *Indeterminancy Principle*. That is: To measure it is to change it! Consider high stakes standardized testing. Such tests are widely regarded as powerful leverage points which influence (for better or worse) both *what* is taught and *how* it is taught. But, when they are used to clarify (or define) the goals of instruction, such tests go beyond being neutral indicators of learning *outcomes*; and, they become powerful components of the *initiatives* themselves. Consequently, far from being passive indicators of non-adapting systems, they have powerful positive or negative effects, depending on whether they support or subvert efforts to address desirable objectives. Therefore, when assessment materials are poorly aligned with the standards for instruction, they tend to create serious impediments to student development, teacher development and curriculum development.

At a time when countries throughout the world are demanding accountability in education, it is ironic that many of these same countries are adopting without question the most powerful untested curriculum theory that has ever been imposed on schools, teachers, and children. This untested theory is called teaching-to-the-test; and, it is not only untested but it also is based on exceedingly questionable assumptions. For examples, readers need only imagine the next course that they themselves are likely to teach; and, they can think about what would be likely to happen if (a) they based grades for the course entirely on the final examination, and (b) they passed out the final examination to students at the start of the course.

Most of the Systems We Need to Understand Are Complex, Dynamic, and Continually Adapting

The USA's Department of Education states: *The Secretary considers random assignment and quasi-experimental designs to be the most rigorous methods to address the question of project effectiveness.*

In most mature sciences, the most important criteria which should be that determines the scientific quality of a research methodology is based on the recognition that every methodology presupposes a model; so, above all, the scientific merit of a methodology depends on whether the model makes assumptions that are inconsistent with those associated with the "subject" being investigated.

Our observations in the preceding section suggest, *random assignment and quasi-experimental designs* tend to be based on a variety of assumptions that are inconsistent the kind of complex, dynamic, interacting, and continually adapting systems that are of greatest interest to mathematics educators. To see what we mean by this claim, consider the following situation. It involves one of the simplest systems that mathematicians describe as being a complex adaptive system. It involves a double pendulum; and, simulations of such systems can be seen at many internet web sites. For example, the one shown in Fig. 1 came from http://www.maths.tcd.ie/~plynch/SwingingSpring/doublependulum.html. We will use it to simulate a typical study that involves "control groups" in education.

We begin our simulated study by creating two identical browser windows on two identical computers; and, in each window, we set the initial state of the double pendulum so they are identical (see Fig. 1). ... We will think of these two systems as being the "control group" and the "treatment group" in a study where the "treatment" is actually a placebo. In other words, we are setting up a study where we'll investigate whether doing nothing produces a reliable and significant effect. Alternatively, we can think of ourselves as setting up a study to show that, when investigating complex adaptive systems, the whole idea of a "control group" is nonsense. To test our hypothesis, we can set the settings so that each of the two systems produces a trace to show the position of the motion of the tip of the second pendulum in each of the two windows. Then, in each of the two windows, we can punch the start buttons at exactly the same moment; and, after a brief period of time (e.g., 10 seconds in Fig. 2, 20 seconds in Fig. 3), we can stop the two systems at exactly the same time. Then, we can examine the paths of the two pendulum points. ... Clearly, the paths are not the same; and, it is easy to produce a quantitative measure of these

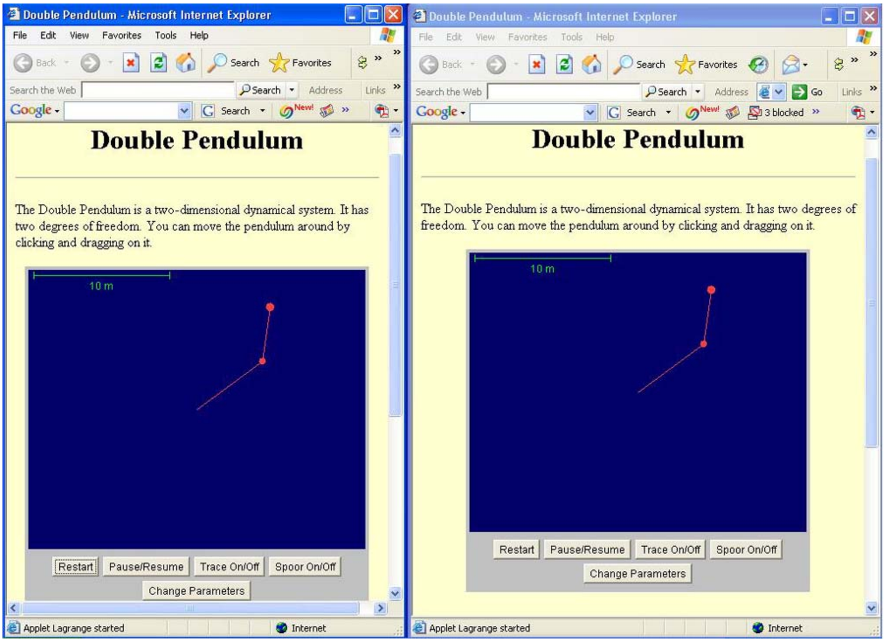

Fig. 1 Two identical starting points for a double pendulum system

differences.[3] In fact, if the two systems are allowed to run for longer periods of time (e.g., more than 20 seconds), then the differences between the two paths begin to be the same as for two paths whose initial positions were completely random (see Fig. 3). In other words, the systems begin to behave as if nothing whatsoever had been ”controlled” in the initial states of the two systems!

Why did these systems behave in this way? Like all complex adaptive systems, one significant fact about a *double pendulum* is that, even though each of its two components obeys simple rules, when the components function simultaneously and interact, the interactions lead to feedback loops that produce chaotic behavior which is unpredictable in the sense that it never repeats itself and cannot be described by a single rule.

One distinguishing characteristic of mathematically complex systems is that the systems-as-a-whole have “emergent properties” which cannot be deduced from properties of elements of the system. In particular, these “emergent properties” cannot be described using single-function models—or even using lists of single-function models. ... This is significant because researchers in the educational, social, and cognitive sciences have come to rely heavily of models that are based on

[3]One easy way to do this is to: (a) superimpose the paths of the two double pendulum systems, (b) mark the locations of the points at equal intervals (e.g., at 1 second, 2 seconds, 3 seconds, and so on), and (3) to measure the distances between corresponding pairs of points.

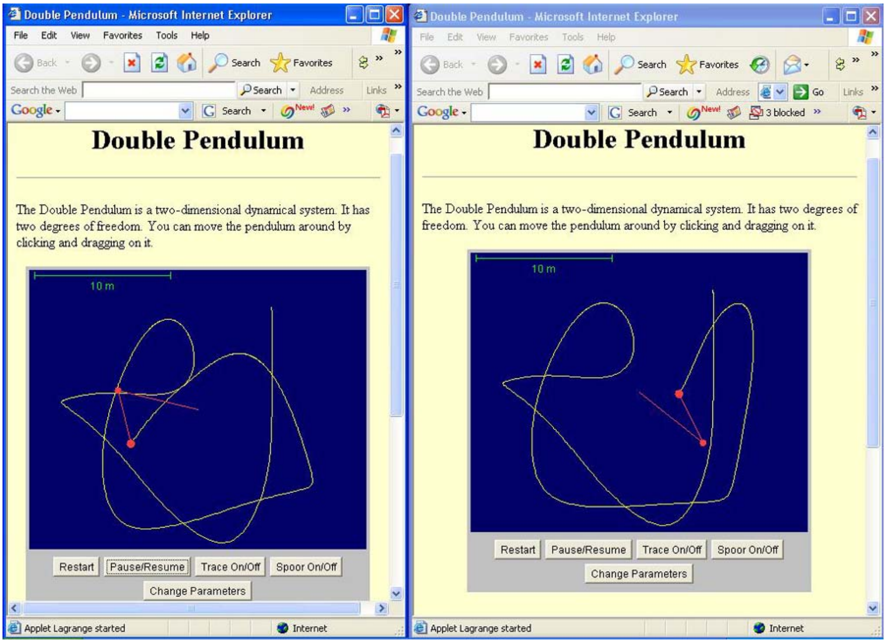

Fig. 2 Stopping the two systems after 10 seconds

simple functions—where independent variables $(A, B, C, \ldots N)$ go in, and dependent variables $(X, Y,$ and $Z)$ come out.

The point that we want to emphasize here is NOT that such systems are completely unpredictable; they are simply not predictable using single-formula models whose inputs are initial conditions of the system and it's elements. In fact, web sites such as http://ccl.northwestern.edu/netlogo/ and http://cognitrn.psych.indiana.edu/rgoldsto/ give many examples of systems which are far more complex, and in some ways just as unpredictable, as double pendulums; yet, these same systems also often involve some highly predictable system-level behaviors. For example:

– In simulations of automobile traffic patterns in large cities, it is relatively easy to produce wave patterns, or gridlock.
– In simulations of flying geese, groups of geese end up flying in a V pattern in spite of the fact that there is no "head" goose.
– In simulations of foraging behaviors of a colony of ants, the colony-as-a-whole may exhibit intelligent foraging behaviors in spite of the fact that there is no "head ant" who is telling all of the other ants what to do.

For the purposes of this paper, the points that are most noteworthy about the preceding systems are that: (a) at one level, each system is just as unpredictable as a *double pendulum*, (b) at another level, each system has some highly predictable "emergent properties" which cannot be derived or deduced from properties of elements themselves—but which results from interactions among elements in the sys-

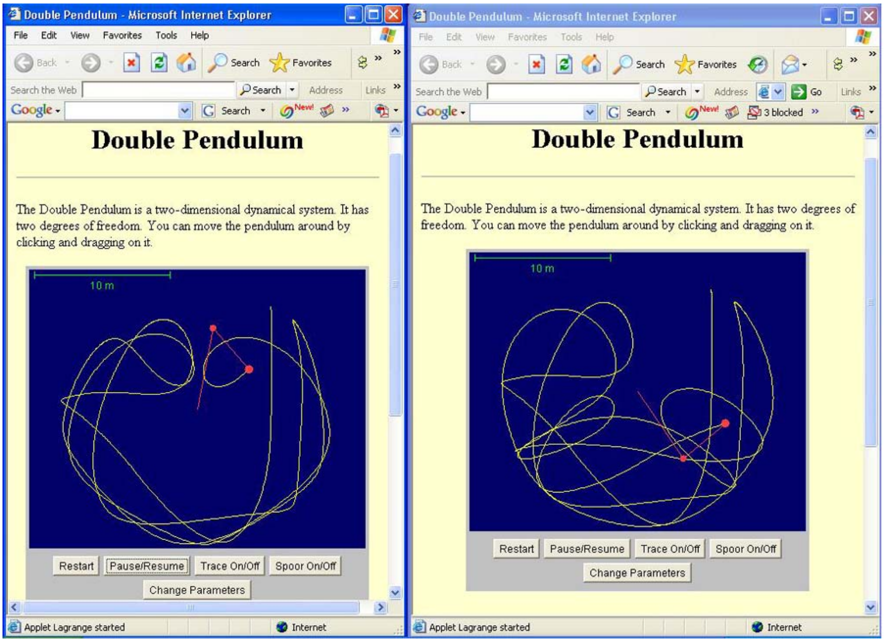

Fig. 3 Stopping the two systems after 20 seconds

tem. Consequently, if the goal is to control such systems, then what needs to be controlled are the interactions—not just the initial conditions. . . . Consider the *Paper Tearing Experiment* described in Fig. 4. Now consider the kind of systems that mathematics educators need to understand and explain—such as: complex programs of instruction, plus complex learning activities in which the complex conceptual systems of both students and teachers or students will be functioning, and interacting, and adapting. . . . Within such a systems, it is clear that the system will involve feedback loops (where A impacts B which impacts C which returns to impact A) and where the systems-as-a-whole develop patterns and properties which result from interactions among elements of the systems, and which cannot be derived or deduced from properties of elements themselves plus properties of any "treatment" that might be used.

In spite of the obvious complexities in educational systems, a prototypical study in education tends to be thought of as one that shows *what works*—even in situations where (a) nobody was clear about what "it" really was that worked, nor what "working" really should have meant, and (b) the assessments themselves were among the most powerful un-tested parts of the "treatment" that presumably were being tested. . . . In fact, as we observed earlier, most tests are chosen precisely because they were intended to influence outcomes. So, in cases where the things they assess are not consistent with the goals of curriculum innovations that they are being used to assess, then they become important parts of the treatments themselves. Furthermore, if they are only used as pre-tests and post-tests, then they neglect to measure the sin-

Take a standard 8.5″ × 11″ printer paper, and mark it as shown in the Fig. 2. Then, make a small cut at point "C" as shown. Next, hold the paper with your two hands—pinching it between your thumb and pointer finger at points A and B. Finally, close your eyes and try to tear the paper into two pieces—so that the tear ends at point "T" on the opposite side of the piece of paper. … Now, repeat the procedure several more times using several sheets of paper, and conduct the following experiment. … The goal of the experiment is to answer this question: *What is the best possible place to make the cut "C" so that the tear ends at point "T" on the opposite side of the sheet of paper*? (note: One answer is given in the footnote below.[5])

Fig. 1: 8.5″ × 11″ Printer Paper

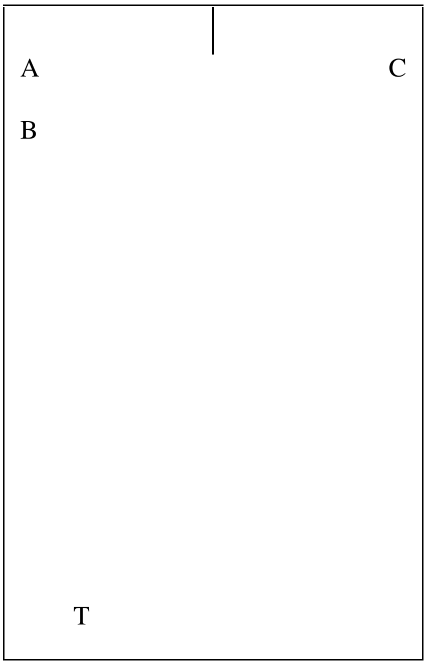

Fig. 4 A paper tearing experiment

gle most important parts of the situations being assessed—that is, the interactions. In other words, the situation becomes very similar to tearing paper with your eyes closed in the *Paper Tearing Problem*.

What Kind of Explanations are Appropriate for Comparing Two Complex Systems?

The preceding section focused mainly on complex systems such as those that characterize large curriculum innovations. But, similar observations also apply to the kind of complex systems that characterize the thinking of experts or novices in studies of students or teachers. For example, in virtually every field where ethnographic studies have been conducted to compare experts and novices, results have shown that experts not only *do* things differently than novices but they also *see* things differently. Experts not only *do things right* but they also *do the right things*—by doing them at the right time, with the right situations, and for the right purposes. Yet, in

[5]Perhaps the "best" answer to the question is: *If you open your eyes almost any point will work as well as any other for the cut "C"*.

spite of these observations, students, teachers, and programs continue to be developed (and assessed) as if "excellence" was captured in some cookbook-style list of rules.

Example (Expert Cooks See (and Taste) Differently Than Novices!) Even if a cook's goal involves nothing more complex than making a *Margarita Pizza*, truly exceptional cooks tend to use high protein bread flour rather than all-purpose flour, baker's yeast rather than dried yeast packages, freshly ground rock salt rather than preprocessed and chemically treated salt, water that isn't simply tap water, and so on. Furthermore, if possible, they use freshly picked basil rather than dried herbs, and tomatoes that are home grown, and in season, and freshly picked. And, they recognize that: (a) there are many types of tomatoes, cheeses, olive oils, herbs, and yeasts—and that the best of these vary widely in the ways that they need to be handled, and (b) combining such ingredients isn't simply a matter of following rules, it involves tasting and adapting. Finally, outstanding cooks recognize that different ovens, sauce pans, and burners behave very differently, and that the performance of such tools often is strongly influenced by factors such as altitude and humidity. Consequently, results from a given set of ingredients may vary considerably from one season to another, from one location to another, and depending on the tastes of diners. (*Do your guests prefer strong asiago cheeses with heavy sourdough crusts; or, do they prefer mild mozzarella cheeses or non-dairy cheese-like substances with cracker-thin grilled crusts?*). So, great pizza chefs select the best available ingredients; they compose their sauces and pizzas by testing and adapting (rather than blindly following rules); they pay attention to patterns (rather than just pieces) of information; and, they continually adapt their thinking as well as their products to fit changing circumstances. They don't simply create their pizzas using fixed formulas; and, even though most of them usually end up being very skillful at using tools such as knives and sauce pans, they didn't learn to be great chefs by waiting to make whole meals until after they become skillful at using every tool at http://www.cooking.com/.

Example (Single Formulas Solutions Don't Work!) If any recipe could claim the prize for "working best", then it might be the recipe for *Toll House Chocolate Chip Cookies*. Yet, if the standard tried-and-true recipe from a bag of *Hershey's Chocolate Bits* is given to twenty professional mathematicians, then the result is sure to be twenty batches of cookies that are very different from one another—even thought the mathematicians probably are not incompetent at measuring and following rules. Conversely, if twenty superb cooks make *Toll House Chocolate Chip Cookies*, then they are sure to modify the recipe to suit their own preferences, current resources, and cooking environments—as well as the preferences of the people who are expected to eat their cookies. This is another reason why, malleability, not rigidity, tends to be one of the most important hallmarks of both great recipes, great cooks, and great curriculum materials. In fact, even if a cookbook is written by the cook who is using it, the book tends to be filled with notes about possible modifications for different situations. So, the half-life of cookbooks (as an actual plans of action) tends to be no longer than the half-life of a useful syllabus for a course that is taught by a truly excellent teacher (who continually adapts her behaviors to meet the needs of specific and continually changing students and classroom communities).

Therefore, because of the continuing power of this "cooking" metaphor, we believe that it might be useful to examine it more carefully. ... Even though few people would deny that cooking involves a great many formulaic recipes that need to be mastered, it also is obvious that cooking is an activity where a great deal more is involved than simply following fixed formulas. For example: Excellent cooks usually have large collections of cookbooks; and, they know that no recipe or cookbook "works" for all situations—or for all levels and types of chefs or guests. An entry level cookbook is not the same as an advanced cookbook; and, excellent cooks generally are not victims of a single, inflexible style. They are able to manipulate their personae to suit changing circumstances—which include the preferences of guests, and the availability of fresh and high quality ingredients.

Excellent cooks need to do more than follow recipes in cookbooks that use standardized off-the-shelf ingredients. They generally need to: (a) make substitutions and adaptations in recipes in order to use ingredients that are freshest and best, (b) understand relationships that make harmonious tastes so that exciting and creative compositions can be made, (c) taste what is being composed and adapt recipes accordingly, and (d) understand difficult-to-control things such as heat flow in their ovens and pans.

Again, examples from cooking are similar to the situation described in the Paper Tearing Problem that was described in the preceding section of this chapter. That is, a cook who doesn't taste-and-adjust is like a paper tearing by a person who only works with his eyes closed—or like non-adaptive "treatments" in curriculum innovation.

Lack of Cumulativeness is Our Foremost Problem

One of the foremost reasons why mathematics education research has failed to answer teachers' questions is because its results have a poor record of accumulation. Lack of accumulation is an important issue because most realistically complex problems will only be solved using coordinated sequences of studies, drawing on multiple practical and theoretical perspectives, at multiple sites, over long periods of time (Lesh et al. 2005; Kelly and Lesh 2000). However, this failure to accumulate tends to be portrayed as a problem in which "the field" had not agreed on basic definitions and terminology (Kilpatrick 1969a, 1969b; Begle 1979; Silver 1985; Lester 1994; Lester and Kehle 2003; Schoenfeld 1993). So, nobody in particular is to blame. Whereas, shortcomings that most need to change are more closely related to the work of individuals who: (a) continually introduce new terms to recycle old discredited ideas—without any perceivable value added, (b) continually embellish ideas that "haven't worked"—rather than going back to re-examine foundation-level assumptions, and (c) do not develop tools to document and assess the constructs they claim to be important.

Consider the research literature on problem solving. In the 1993 *Handbook for Research on Mathematics Teaching and Learning* (Grouws 1993), Schoenfeld described how, in the United States, the field of mathematics education had been subject to approximately 10-year cycles of pendulum swings between basic skills and problem solving. He concluded his chapter with optimism about the continuation of a movement that many at that time referred to as "the decade of problem solving" in mathematics education. However, since the time that the 1993 handbook was published, the worldwide emphasis on high-stakes testing has ushered in an especially virulent decade-long return to basic skills.

Assuming that the pendulum of curriculum change again swings back toward problem solving, the question that mathematics educators should consider is: *Have we learned anything new so that our next initiatives may succeed where past ones have failed?* … Consider the following facts.

Polya-style problem solving heuristics—such as *draw a picture*, *work backwards*, *look for a similar problem*, or *identify the givens and goals*—have long histories of being advocated as important abilities for students to develop (Pôlya 1945). But, what does it mean to "understand" them? Such strategies clearly have descriptive power. That is, experts often use such terms when they give after-the-fact explanations of their own problem solving behaviors—or those of other people that they observe. But, there is little evidence that general processes that experts use to describe their past problem solving behaviors should also serve well as prescriptions to guide novices' next-steps during ongoing problem solving sessions. Researchers gathering data on problem solving have the natural tendency to examine the data in front of them through the lens of *a priori* problem solving models. Although there is great value in doing so, does such an approach really move problem-solving research forward? If one examines the history of problem solving research, there have been momentous occasions when researchers have realized the restricted "heuristic" view of problem solving offered by the existing problem solving research "toolkits" and have succeeded in re-designing existing models with more descriptive processes. However the problem remains that descriptive processes are really more like names for large categories of skills rather than being well defined skills in themselves. Therefore, in attempts to go beyond "descriptive power" to make such processes more "prescriptive power", one tactic that researchers and teachers have attempted is to convert each "descriptive process" into longer lists of more-restricted-but-also-more-clearly-specified processes. But, if this approach is adopted, however, then most of what it means to "understand" such processes involves knowing when to use them. So, "higher order" managerial rules and beliefs need to be introduced which specify when and why to use "lower order" prescriptive processes. … The obvious dilemma that arises is that, on the one hand, short lists of descriptive processes have appeared to be too general to be meaningful; on the other hand long lists of prescriptive processes tend to become so numerous that knowing when to use them becomes the heart of understanding them. Furthermore, adding more metacognitive rules and beliefs only compounds these two basic difficulties.

Begle's (1979) early review of the literature on problem solving concluded that

> (N)o clear cut directions for mathematics education are provided by the findings of these studies. In fact, there are enough indications that problem-solving strategies are both problem- and student-specific often enough to suggest that hopes of finding one (or a few strategies) which should be taught to all (or most students) are far too simplistic. (p. 145)

Similarly, Schoenfeld's (1993) review of the literature concluded that attempts to teach students to use general problem-solving strategies (e.g., draw a picture, identify the givens and goals, consider a similar problem) generally had not been successful. He recommended that better results might be obtained by (a) developing and teaching more *specific problem-solving strategies* (that link more clearly to classes of problems), (b) studying how to teach *metacognitive strategies* (so that students learn to effectively deploy their problem-solving strategies and content knowledge), and (c) developing and studying ways to eliminate students' counter-productive beliefs while enhancing productive *beliefs* (to improve students' views of the nature of mathematics and problem solving). Schoenfeld's classroom-based research indicated some measures of success using the preceding approach. But, when assessing these results, one needs to keep in mind that the instruction was implemented by a world-class teacher who was teaching within a complex and lengthy learning environment where many different factors were at play. Thus, even though some indicators of success were achieved, the reasons for success are difficult to sort out. As Silver (1985) pointed out long ago, even when a particular problem-solving endeavor has been shown to be successful in improving problem solving performance, it is not clear *why* performance improved. The reason may have nothing to do with problem solving heuristics.

A decade later, in another extensive review of the literature, Lester and Kehle (2003) again reported that little progress had been made in problem solving research—and that problem solving still had little to offer to school practice. Their conclusions agreed with Silver (1985), who long ago put his finger on what we consider to be the core of the problem in problem solving research. That is, the field of mathematics education needs to go "beyond process-sequence strings and coded protocols" in our research methodologies and "simple procedure-based computer models of performance" to develop ways of describing problem solving in terms of conceptual systems that influence students' performance (p. 257).

When a field has experienced more than fifty years of pendulum swings between two ideologies, both of which both have obvious fundamental flaws, perhaps it's time to consider the fact that these are not the only two options that are available. For example, one alternative to traditional problem solving perspectives is emerging from research on *models & modeling perspectives* on mathematics problem solving, learning and teaching. For the purposes of this chapter, however, the details of *models & modeling perspectives* are not important. Instead, what we will emphasize is that *models & modeling perspectives* have gone back to re-examine many of the most fundamental beliefs that have provided the foundations of problem solving research in mathematics education; and, in almost every case, what we have found is that we need to reconceptualize our most basic notions about the nature of problem solving—and about the kind of "mathematical thinking" that is needed for success beyond school classroom.

Models & modeling perspectives developed out of research on concept development more than out of research on problem solving. So, we focus on what it means to "understand" and on how these understandings develop. We also investigate how to help students function better in situations where they need to modify/adapt/extend/refine concepts and conceptual systems that ALREADY ARE AVAILABLE (at some level of development) rather than trying to help them function better in situations where relevant ways of thinking are assumed to be LOST OR MISSING (i.e., What should they do when they're stuck?).

Summary—Comparing Ideologies, Theories and Models

Having developed only slightly beyond the stage of continuous theory borrowing, the field of mathematics education currently is engaged in a period in its development which future historians surely will describe as something akin to the *dark ages*—replete with inquisitions aimed at purging those who don't vow allegiance to vague philosophies (e.g., "constructivism"—which virtually every modern theory of cognition claims to endorse, but which does little to inform most real life decision making issues that mathematics educators confront and which prides itself on not generating testable hypotheses that distinguish one theory from another)—or who don't pledge to conform to perverse psychometric notions of "scientific research" (such as pretest/posttest designs with "control groups" in situations where nothing significant is being controlled, where the most significant achievements are not being tested, and where the teaching-to-the-test is itself is the most powerful untested component of the "treatment") (also states in other chapters by editors).

With the exception of small schools of mini-theory development that occasionally have sprung up around the work a few individuals, most research in mathematics education appears to be ideology-driven rather than theory-driven or model-driven.

Ideologies are more like religions than sciences; and, the "communities of practice" that subscribe to them tend to be more like cults than continually adapting and developing learning communities (or scientific communities).

Their "axioms" are articles of faith that are often exceedingly non-obvious—and that are supposed to be believed without questioning. So, fatally flawed ideas repeatedly get recycled.

Their "theorems" aren't deducible from axioms; and, in general, they aren't even intended to inform decision-making by making predictions. Instead, they are intended mainly to be after-the-fact "cover stories" to justify decisions that already have been made. . . . They are accepted because they lead to some desirable end, not because they derive from base assumptions.

New ideas (which generally are not encouraged if they deviate from orthodoxy) are accepted mainly on the basis of being politically correct—as judged by the in-group of community leaders. So, when basic ideas don't seem to work, they are made more-and-more elaborate—rather than considering the possibility that they might be fundamentally flawed.

Theories are cleaned up bodies of knowledge that are shared by a community. They are the kind of knowledge that gets embodied in textbooks. They emphasize formal/deductive logic, and they usually try to express ideas elegantly using a single language and notation system.

The development of theory is absolutely essential in order for significant advances to be made in the thinking of communities (or individuals within them). But, theories have several shortcomings.

Not everything we know can be collapsed into a single theory. For example, models of realistically complex situations typically draw on a variety of theories.

Pragmatists (such as Dewey, James, Pierce, Meade, Holmes) argued that it is arrogant to assume that a single "grand theory" will provide an adequate basis for decision-making for most important issues that arise in life (Lesh and Sriraman 2005).

- Models are purposeful/situated/easily-modifiable/sharable/re-useable/multi-disciplinary/multi-media chunks of knowledge.
- Models are both bigger than and much smaller than theories.
- Here are some ways that models are bigger than theories.
  - They often (usually) integrate ideas from a variety of theories.
  - They often (usually) need to be expressed using a variety of representational media.
  - They are directed toward solving problems (or making decisions) which lie outside the theories themselves—so the criteria for success lie outside the relevant theories.
- Here are some ways that models are much smaller than theories.
  - They are situated. That is, they are created for a specific purpose in a specific situation. On the other hand, they not only need to be powerful for the this one specific situations. Models are seldom worth developing unless they also are intended to be:
  - Sharable (with other people)
  - Re-useable (in other situations)
- So, one of the most important characteristics of an excellent model is that it should be easy to modify and adapt.

Concluding Points

The powerful pull of ideology is becoming apparent even in the popular press—and even with respect to domains of knowledge that have nothing to do with emerging fields of scientific inquiry. For example, consider George Lakoff's best selling book, *Don't Think of an Elephant*, which attempts to explain why, in the last presidential election in the USA, so many citizens clearly voted against their own best interests.

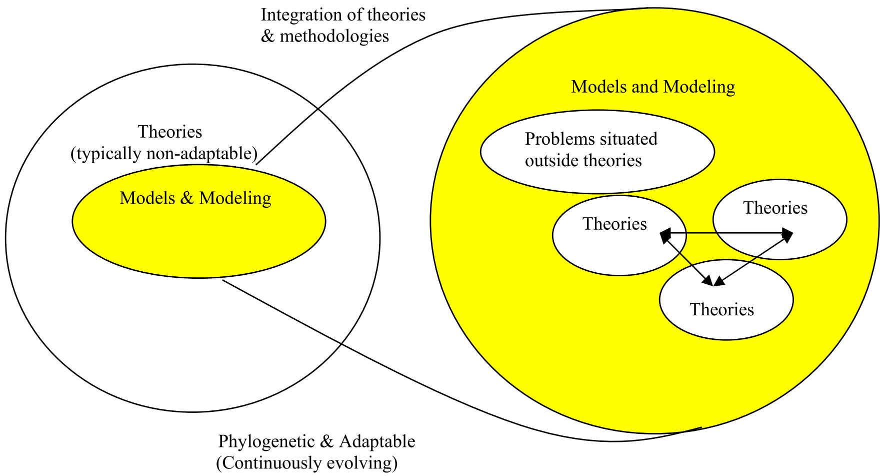

Fig. 5 A conceptual schematic of MMP's

. . . Says Lakoff:

> People make decisions . . . based on their value systems, and the language and frames that invoke those values. (p. xii)

> Frames are mental structures that shape the way we see the world. (p. xv)

> They are . . . structures in our brains that we cannot consciously access, but though we know by their consequences . . . People think in frames. . . . To be accepted, the truth must fit people's frames. If the facts do not fit a frame, the frame stays and the facts bounce off. . . . Neuroscience tells us that each of the concepts we have—the long-term concepts that structure how we think—is instantiated in the synapses of our brains. Concepts are not things that can be changed just by someone telling us a fact. We may be presented with facts, but for us to make sense of them, they have to fit what is already in the synapses of the brain. (p. 18)

The experiential world of the 21st century student is characterized by complex systems such as the internet, multi-medias, sophisticated computing tools, global markets, virtual realities, access to online educational environments etc. In spite of the rapidly changing experiential world of today's student, our approaches to studying learning are still archaic. As discussed in this paper, setting up contrived experiments to understand how students' think/process mathematical content is interesting but conveys a uni-dimensional picture of learning with very limited implications for pedagogy and for future research. Today's students are more likely to be engaged in professions that calls for competencies related to understanding complex real world phenomena, team work, communication and technological skills. So, in essence there are three kinds of *complex systems*: (a) "real life" systems that occur (or are created) in everyday situations, (b) conceptual systems that humans develop in order to design, model, or make sense of the preceding "real life" systems, and (c) models that researchers develop to describe and explain students' modeling abilities. These three types of systems correspond to three reasons why the study of complex systems should be especially productive for researchers who are attempting to advance theory development in the learning sciences. In mathematics and science, conceptual systems that humans develop to make sense of their experiences generally are referred to as models. A naive notion of models is that they are simply (familiar) systems that are being used to make sense of some other (less familiar) systems—for some purpose. For example, a single algebraic equation may be referred to as a model for some system of physical objects, forces, and motions. Or, a *Cartesian Coordinate System* may be referred to as a model of space—even though a *Cartesian Coordinate System* may be so large that it seems to be more like a language for creating models rather than being a single model in itself. In mathematics and science, modeling is primarily about purposeful description, explanation, or conceptualization (quantification, dimensionalization, coordinatization, or in general mathematization)—even though computation and deduction processes also are involved. Models for designing or making sense of such complex systems are, in themselves, important "pieces of knowledge" that should be emphasized in teaching and learning—especially for students preparing for success in future-oriented fields that are heavy users of mathematics, science, and technology. Therefore, we claim that modeling students modeling is the study of a complex living system with layers of emerging ideas, sense making and a continuous evolution of knowledge, which suggests we adopt a phylogenetic approach to modeling the growth of knowledge and learning. The field of economics is an interesting case study which reveals paradigmatic shifts in theories from archaic models for simple agricultural economies to more complicated industrial economies onto the modern day integration of game theory, evolutionary biology and ecology that characterize current economic theories. A phylogenetic approach to the study of domain-specific knowledge has been embraced by linguists, biologists, physicists, political scientists, so why not the learning sciences, which attempts to study the growth of ideas. The conceptual system that we refer to as *models & modeling* (see Lesh and English 2005) is not intended to be a grand theory. Instead, it is intended to be a framework (i.e., a system of thinking together with accompanying concepts, language, methodologies, tools, and so on) that provides structure to help mathematics education researchers develop both models and theories (notice that we've used plurals here). We do not strive for orthodoxy. We encourage diversity. But, we also emphasize other Darwinian processes such as: (b) selection (rigorous testing), (c) communication (so that productive ways of thinking spread throughout relevant communities), and (d) accumulation (so that productive ways of thinking are not lost and get integrated into future developments).

References

Begle, E. G. (1979). *Critical Variables in Mathematics Education*. Washington D.C.: The Mathematics Association of America and the National Council of Teachers of Mathematics.

Biehler, R., Scholz, R. W., Strässer, R., & Winkelmann, B. (Eds.) (1994). *Didactics of Mathematics as a Scientific Discipline*. Norwell, MA: Kluwer.

Bishop, A. J. et al. (2003). *Second International Handbook on Mathematics Education*. Dordrecht: Kluwer.

Coray, D., Furinghetti, F., Gispert, H., Hodgson, B., & Schubring, G. (Eds.) (2003). *One Hundred Years of l'Enseignement Mathématique*, Monograph no. 39. Geneva.

Cramer, F. (1993). *Chaos and Order*. New York: VCH Publishers.

Dietrich, O. (2004). Cognitive evolution. In F. M. Wuketits & C. Antweiler (Eds.), *Handbook of Evolution* (pp. 25–77). Wiley-VCH Verlag GmbH & Co: Weinheim.

English, L. D. (Ed.) (2002). *Handbook of International Research in Mathematics Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

English, L. D. (2003). Reconciling theory, research, and practice: A models and modelling perspective. *Educational Studies in Mathematics*, 54(2 & 3), 225–248.

Holton, G. (1973). *Thematic Origins of Scientific Thought*. Cambridge, Massachusetts: Harvard University Press.

Kelly, A. E., & Lesh, R. (2000). *Handbook of Research Design in Mathematics and Science Education*. New Jersey: Lawrence Erlbaum & Associates.

Kelly, A. E., Lesh, R. A., & Baek, J. Y. (Eds.) (2008). *Handbook of Design Research Methods in Education: Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching*. New York: Routledge.

Kilpatrick, J. (1969a). Problem solving in mathematics. *Review of Educational Research*, 39(4), 523–533.

Kilpatrick, J. (1969b). Problem solving and creative behavior in mathematics. In J. W. Wilson & L. R. Carry (Eds.), *Studies in Mathematics: Vol. 19. Reviews of Recent Research in Mathematics Education* (pp. 153–187). Standford, CA: School Mathematics Study Group.

Kirschbaum, D. Introduction to complex systems. Retrieved from http://www.calresco.org.

Lesh, R., & English, L. (2005). Trends in the evolution of the models-modelling perspective on mathematical learning and problem solving. *International Reviews on Mathematical Education*, 37(6), 487–489.

Lesh, R., & Sriraman, B. (2005). John Dewey revisited—pragmatism and the models-modeling perspective on mathematical learning. In A. Beckmann, C. Michelsen, & B. Sriraman (Eds.), *Proceedings of the 1st International Symposium on Mathematics and its Connections to the Arts and Sciences* (pp. 32–51). May 18–21, 2005, University of Schwaebisch Gmuend. Germany: Franzbecker Verlag.

Lesh, R., Hamilton, E., & Kaput, J. (Eds.) (2005). *Foundations for the Future in Mathematics Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lester, F. K. (2005). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. *International Reviews on Mathematical Education*, 37(6), 457–467.

Lester, F. K., & Kehle, P. E. (2003). From problem solving to modeling: The evolution of thinking about research on complex mathematical activity. In R. Lesh & H. Doerr (Eds.), *Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning and Teaching* (pp. 501–518). Mahwah, NJ: Erlbaum.

Lester, F. (1994). Musings about mathematical problem-solving research: 1970–1994. *Journal for Research in Mathematics Education*, 25(6), 660–675.

Moreno, & Sriraman (2005). The articulation of symbol and mediation in mathematics education. *International Reviews on Mathematical Education*, 37(6), 476–486.

Pôlya, G. (1945). *How to Solve it*. Princeton, NJ: Princeton University Press. 2nd edition, 1957.

Schoenfeld, A. H. (1993). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 334–370). New York: Macmillan.

Silver, E. A. (Ed.) (1985). *Teaching and Learning Mathematical Problem Solving: Multiple Research Perspectives*. Hillsdale, NJ: Erlbaum.

Steen, L. (1999). Review of Mathematics Education as research domain. *Journal for Research in Mathematics Education*, 30(2), 235–241.

Steen, L. A. (2001). Revolution by stealth. In D. A. Holton (Ed.), *The Teaching and Learning of Mathematics at University Level* (pp. 303–312). Dordrecht: Kluwer Academic Publishers.

Wiliam, D. (2003). The impact of educational research on mathematics education. In A. J. Bishop et al. (Eds.), *Second International Handbook on Mathematics Education* (pp. 471–490). Dordrecht: Kluwer.