If We Want to Get Ahead,

We Should Get Some Theories*

Andrea A. diSessa
Graduate School of Education
University of California
Berkeley, California 94720
September, 1991

*I am "borrowing" the essence of the title from the justifiably well-known paper by Karmiloff-Smith and Inhelder (1974/75).

Abstract

In education, or in the learning sciences generally, theory is in a poor state. We have not reached deep theoretical understanding of knowledge or of the learning process, and it is important that we recognize this. Even more importantly, our community does not seem particularly intent or armed to change the situation. This paper is aimed at raising the issue of intent, arguing for new dedication toward theory. It is also aimed at a modest contribution to our toolkit for a more theoretically attentive practice of education research.

Introduction

I view the educational research community as demonstrating only minor concern for theory and its development. That should not be so. Minimally, I hope with this paper to spur discussion of the issue; at best, I hope to participate in building a consensus about the importance of theoretical thinking to our goals, and about what kind of theoretical thinking makes most sense.

My approach will be personal and more than usually assertional for two reasons. First, I hope to raise issues provocatively and relatively sharply. Second, there are deep and complex epistemological issues here that I simply cannot enter into in any great detail. I recognize I will mostly be staking ground rather than uncovering, explicating or settling the issues involved.

Theory has a somewhat deservedly bad reputation in educational circles. The relation of theory to practice is problematic. Many times the best practitioners don't have any explicit theory at all. Alternatively, it may not be at all clear that the theory they espouse "does the work" in their good practice, as opposed to their practical expertise. Others with the same theory may not be nearly as good at teaching. Some of the best, or at least, *best known* theories, such as Piagetian stages, have often seemed to put a straightjacket on instruction rather than offering many productive suggestions. To practitioners, and all too often for researchers as well, "in theory" is more a lazy lament that some expectation has gone awry rather than an appeal to some felt-to-be necessary and well-elaborated set of ideas.

Along the same lines, theoretically inclined researchers seem often to ignore the most obvious common sense. They do "silly things," if they *do* anything at all, and discover those things don't work. Or they do clever things and hide their cleverness behind theoretical claims that just do not seem refined or appropriate enough to catch their own cleverness.

I want to claim that whatever might ail both theory itself and its relation to practice is not incorrigible. For many enduring reasons, theoretical development is a principal hope for the future. An uncertain relationship between theory and practice should be viewed as an indicator of too little and insufficiently sharp theoretical thinking rather than an indicator that theory is not useful.[1]

I advocate cultivating community skills and predilections for theory. In this I am certainly not alone, although I feel I am in the minority.

I begin assuming that there is face value in having good theory, and assess the current situation in that light. Then I examine in more detail the standards by which my judgments are made. At that point, I will briefly return to buttress the assumption that theory is valuable and not just an

----

1. There was no sharp boundary between Aristotle's ethics and his physics. After Newton sharply formulated his physics, it is clear to us that it helps specifically with designing effective and efficient automobiles, but it should not be expected to decide whether it is our right to pollute.

annoyance. Finally, I turn to how we might pursue being more theoretical. These last suggestions are particularly important as they help define the practice of being theoretical as I see it, and also provide some doable steps short of directly inventing deep and excellent theories.

The State of the Art

Baldly, I think the state of the art with respect to theory is, indeed, quite poor. There are two sides of this. First, there is no general agreement at the level of theories of learning or instruction. There just aren't any strong, broadly respectable and workable theories around. Tom Romberg commented on one of the most thoroughly researched areas, children's arithmetic, in a collected volume that represented the state of the art in 1982:

> This copious literature has lacked an implicit body of intertwined theoretical and methodological beliefs that permit selection, evaluation, and criticism. (p. 1)

His hopes that the situation was imminently to change, on the "route to normal science," have not been realized. As evidence, I note that several of the contributors to that volume have moved strongly away from their orientation at that time, and the rest have not converged into anything like the common frame Romberg hoped would emerge. In areas closer to my own, like "misconceptions" and conceptual change in science, I am willing to be even more aggressive in asserting the theoretical backdrop is fragmented, diverse, and, if for no other reason than that, unsatisfactory.

I strongly believe that there were theoretically interesting threads in 1982, as there are now.[2] Several of the participants in the volume noted above had and have what I judge to be insightful theoretical frames. Case and Steffe, et al., have, in their particular areas and in their own ways, done Piaget one better. Vergnaud's theoretical work on conceptual fields and "theorems in action" is related to some of my own thinking, and appeals to me. The computationally-oriented VanLehn and Greeno (Greeno, vintage 1982!) bridge to another powerful community of theoretical thinkers who deserve attention and respect.

Yet, the list is awkwardly long if it is to represent strong and broad theoretical lines. The list is also labelled mostly by individuals who, for the most part, are the only ones pursuing their theoretical lines. There is *enormous* diversity of styles and aesthetics evident, even if I limit myself to what is represented in that one volume. All these facts show severe limitations in what the research community can claim about its theoretical state.

Rather than theories, there are broad communities with similar and, arguably, strong meta-theoretical commitments. Certainly there is an unmistakable family resemblance among "Pittsburgh school" computationalists, although you must chose among ACT*, SOAR, etc. Closer to home, many call themselves constructivists these days. However, constructivism is not a well-developed theory, or even a class of theories. It lacks specificity, to take one obvious and important measure. It never really comes down to saying, as far as I can tell, exactly what and when people will learn. That is why Case, Steffe, von Glasersfeld, myself and others who are, in some ways, dyed-in-the-wool constructivists all pursue different theoretical lines.

Social constructivists, who are increasingly visible in the cognitively oriented education community, or those who advocate a situated view of cognition, also share meta-theoretical commitments. Yet there is precious little that even claims to be a compactly articulated theory,

----

2. Looking at the contributions, it's striking how little, in some sense, the situation has changed in 9 years.

as opposed to an elaborated point of view, and I am skeptical about the well-formedness and clarity of these views.

So, we have precious little in the way of "hard core" theory. I am not demeaning pre-theoretical or "mere" meta-theoretical points of view. As a matter of fact, I expect that theories can only emerge as elaborations of these points of view, so we *need* to cultivate them as a means to better theories. But they are not the theories we need.

There is no shame in the fact that we do not have broad and deep theories. I believe theory development about learning and instruction is among the deepest and most difficult topics of contemporary investigation. That anyone has only paltry theories to offer is disappointing, but not surprising.

The second feature of the contemporary landscape of theory development is less cosmic than the inherent difficulty of understanding knowledge and its development, and our current "pre-Galilean" state with respect to this. That feature is, therefore, perhaps more something about which we can and should immediately do something. The general level of theoretical awareness and concern in education and learning-oriented communities is quite impoverished. In the extreme, investigators don't know or care that they have no systematic framework to guide their work, let alone a theory. They feel the most schematic principle deserves the name "theory."

I have been particularly struck with both the lack of theory and the lack of concern and critical judgment with respect to theory in the context of reviewing papers for journals. The influence of experimental psychology is strong. Experimental methods are well-developed, and there are good criteria for having adequately carried out an experiment. Reviewers are attentive to the aptness of particular statistical tests and general experimental design principles. Even most standard paper organizational formats derive from what is needed to present an experiment coherently. Or course, this is not troublesome except in contrast to the way theoretical ideas are handled. Ad hoc criteria abound, if any are applied at all. As I suggested, I think quite incoherent or simply unclear points of view are proposed as theories. Almost anything may get past reviewers theoretically, while experiments are thoroughly vetted for cultivated community practices and standards. Experimentally, confounds in experiment 1 are acknowledged and inevitably lead to a revised control in experiment 2. Theoretically, I long for the day that we similarly acknowledge familiar gaps in our positions and invoke standard repair strategies for future work.

I can cite a couple of other points at which the lack of concern for theory is vexing to me. I find it amazing that graduate school requirements are filled with "methodology" courses, while I've not yet heard of one that focused on the development of theory. That indicates a feeling that theory is either too easy to deserve attention, or else it is hopeless, at best an art that only the tiniest fraction of researchers will develop.

I also find that the way literature is cited betrays a deeply empiricist and a-theoretical bent. Articles are cited as "X showed that Y," where Y is some easily statable fact. My own reading of these articles is almost always full of nuance. They might have suggested terms for analysis and interpretations of data, but it is hardly ever compellingly clear that their terms of analysis are optimally appropriate, or that very different interpretations might not be as apt. Almost all the work in providing other interpretations and, more important, pursuing the meanings of terms, their integrity and general utility is left to the theoretically reflective reader. Similarly, much research provides phenomena without explanations. Experts do this; novices do that. Any theoretically inclined reader wants to know *why*?

In a nutshell, not many people care much about theories. Standards of practice are sorely lacking.

How Do We Know Theoretical Work When We See It?

Given the diversity of standards of theory, I feel obligated to elaborate mine. All my educational training was in physics, which may be the best developed empirical science in terms of theories. There is danger in saying any social science should be in any respect like any physical science, but standards do not arise by fiat.

I take three things from my experiences with physics. Each of these provides a "place to look" and a "judgment to make" with respect to the state of theory in an empirical science. The first has to do with the "texture" of theories, their scope and structure as complex systems of knowledge. The second concerns how the quality of theory may be judged by the quality of data that is acquired in its service. The third concerns some signs that indicate genuine theoretical progress over common sense.

Theories are richly interconnected collections of ideas and are substantial precisely because of their unusual integration. I learned from physics how much it takes to create an adequate theoretical frame. This is not done in a day of thinking or in a flash of insight. It is not explained in a paragraph or two. When scientists seem to have flashes and create revolutions, usually it is easy to see how much his/her own work and that of the community has gone into preparing for the "flash." It is trivial, I think, to understand how even Einstein's stunning "de nova" creations were tied in many and deep ways to cumulative work. And filling out the system or cleaning up the foundations has typically taken at least decades, if not generations.

Fundamental physical theories are as rich and compelling (to those who hold them) as world views. They are intricately connected to a stunning degree. There are many ways to present them, yet there is such a solidity in their interconnected nature that, among adherants, some experiments at least have entirely unambiguous interpretation and cleanly prescribed results.[3] Every Newtonian knows the outcome of billiard ball collisions.

That kind of clarity sometimes allows decisive experiments within the general theoretical frame.[4] Consider that so many scientists can agree that a little quiver of a meter reading can mean a theory of stellar evolution has been substantially confirmed. Here, I'm thinking of the detection of less than a score of neutrinos which has recently contributed vital substantiation to hypotheses related to stellar evolution and super novas. That "little quiver" (metaphorically) represents the detection of a neutrino, a massless particle that travels at the speed of light and can easily penetrate the earth. The quiver rests on a strong fulcrum consisting of a stunningly reliable understanding of the contexts of quivering, a transparent understanding of so many interconnected, invisible but theoretically sensible ideas (like neutrinos), and a web of thousands of experiments in which basic facts of quantum mechanics, relativity and particle physics have not given us enough pause for concern that one would ordinarily think the experiments were even about those fundamental theories. The fulcrum is so strong that it can be leveraged to confirm a theory about stars, where we have never been. How remarkable!

----

3. I am not talking about paradigms being overthrown (or confirmed) by critical experiments. Instead, I am more referring to experiments whose outcome are so obvious that no practitioner would bother performing them except to illustrate a fundamental point to a student. It would be extremely unlikely that a competing theoretician would bother trying to upset a theory on these core grounds.

4. Again, these are decisive *within* the paradigm.

I will be critical of learning theories until they have some similar integrity. As a consequence, for a long time to come we will be able, if we chose to, to critique the adequacy of given and proposed theories. We should chose to do so as a means of advancing our understanding.

A practical implication of this position is that it should be natural and acceptable, if not expected, that those advancing theories should spend as much time explaining the limits of their ideas as expounding them. Much more than where the theory is empirically weak (e.g., what experiment should be done next), this means exploring where it is conceptually weak, where it is unsharp, hard to articulate, in danger of incoherence, and so on. Only if we lower our standards substantially do these critical pursuits not seem worthwhile. Only if we pretend we are much farther along than we are can it be seen as a sign of weakness to discuss these issues with respect to our theoretical proposals.

There is no data without theory. As much as science involves experiment, it is not a purely inductive enterprise. This is so obviously true in contemporary physics that it hardly bears remarking on. If one didn't have a very well-developed notion of what those invisible neutrinos were all about, the "data" of meter twitching I remarked on above would not be data at all. The whole rationale for the experiment and set of observations would not exist, nor would the fabric of reasoning that makes the observations informative. Nobody would have been looking for the quiver, and it would have been incomprehensible if they had accidentally seen it.

There are two things that tend to undermine the influence of the above observation. First, scientific formulations in physics *look like* empirical generalities that one could stumble on by doing a lot of measurements and finding a pattern in the results. One just has to measure a bunch of forces, masses and accelerations and find out that, reliably, $F = ma$. Or you make a bunch of resistors and "discover" Ohm's Law. Why can't we find the laws of learning by correlating parameters? I have only space for a "one-liner": It made no sense and would have been impossible to measure forces or mass before at least some features of the theoretical framework of which they were part existed. Measuring X requires a lot of commitments about the nature of X, the very first, but highly non-trivial part of which, is to believe X exists.[5]

The power of intuitive or commonsense knowledge also undermines the appreciation of how important and necessary theoretical frames are in the production of data. That is, common sense, or some slightly refined species, can substitute for a theoretical frame so easily that we just don't notice it. Every one of us is full of intuitions about the mind and learning. Some of these are cultivated by the language we inherit -- "concepts," "beliefs," even "to know" and "knowledge" -- that have adequate purchase on the world to justify their everyday use. Some roots of these frameworks are probably more private, extrapolations of our own experiences in thinking and learning, or extrapolations of what we observe in others. We can, in these intuitive frames, "observe things" and draw fairly adequate conclusions under some circumstances. For example, we are not outstripping the power of common sense when we say with conviction, "He doesn't *know* I went out with his girl friend."

It is common to say any observation implies a theory. Observations certainly imply a framework of ideas, but not at all a deep theory by the standards implied above. (Hence a-theoretical empiricism does not mean without a framework, but without an adequate scientific one.) The problem is that intuitive frames are not powerful enough to constitute sufficient theories of the mind in general and of learning in particular. We should draw them out when we rely on them, and critique and refine them to produce more scientifically adequate frameworks.

----

5. See diSessa, 1991-b, for an articulation of what might be involved in thinking to measure a quantity and carrying that process out.

Since theory, in some respects and on some occasions, defines data, we can sometimes judge the quality of theory by the quality of its data. I provide a brief and clearly elliptical example where I judge the problem with learning data is in the theory on which the data depends. In this case, the problem seems to me to be both the clarity and integrity of the ideas themselves, and also hidden intuitive presumptions that, when brought to light, seem dubious.

Some "theories" of learning provide that learning occurs when the learner is disequilibrated by new ideas or observations that compete, in some sense, with old ones. I think the commonsense roots of such ideas are evident. Everyone knows the feeling of being presented with "destabilizing" information that doesn't jibe with our current take on the world. We all, also, sometimes follow that feeling with a consideration of the circumstances of our knowing what we think we know, and we sometimes "resolve" the difficulty by realigning our existing "beliefs." Some likely inadequacies of this kind of theory (as sketchily as I've presented it) are not hard to find. First, it is drawn from a particular class of experiences where we have reflective access to our epistemic state: We are aware something is wrong. I take it as the right minimal assumption that this awareness is only possible in certain circumstances where our meta-awareness of knowing processes is above a certain threshold. Second, we must also consider the generality of the processes by which we "decide" to reorganize our beliefs, and the means by which "we" carry out that reorganization. Indeed, the sense of self that is indisputable in commonsense thinking about thinking is hardly something we can, to be theoretically self-conscious, take for granted. Sometimes we can act as an agent on our thoughts in a semi-reflective way. Sometimes, I am quite sure, we cannot. More technically, we could ask what exactly constitutes the state of disequilibrium. If we deprive ourselves of the common sense that says "I've had that feeling!" how do we describe in any generic terms what constitutes that feeling, especially in such a way as to apply to every event of learning? I could also enter into discussion of the empirical limitations of such theory. To put it crudely, there are such a host of details about learning that depend on the specifics of the knowledge to be learned and the individual as he/she comes to the learning context, that it seems unlikely that disequilibration can possibly account for them. If disequilibration uniformly exists, I believe there must be hundreds of different kinds of it. At least, this is a thing to be seriously worried about.

Respectable theory, when we get it, cleanly transcends common sense. My last point of extrapolation from physics to our expectations for theory in education really follows from discussion of the above two points. Unless we can unambiguously point to how we have transcended -- in generality, precision, clarity, and justifiability -- the intuitive sense of mechanism we all build in daily life observing and thinking about psychological matters, we just won't have adequately prepared theoretical ground. I'll pick one focus for this exposition, but I think the point is much broader. Commonsense vocabulary just won't do the job of providing the technical terms of a theory of learning. When we stop with "beliefs," "knowledge," "concepts," and so on, even with a few phrases of elaboration, we are on extremely shaky ground.

To put an edge on this, physics theorizing has always involved ontological innovation. The "force" in Newton's theory is a new entity that simply does not exist in common sense. Even mass took on a much refined interpretation to make sense in that theory. More evidently, quantum wave functions did not exist before quantum mechanics. My presumption is that we will not have adequate theoretical purchase on learning until concepts, facts, beliefs, skills, and all the rest of our common sense about knowledge and learning become reinterpreted within a fabric of more precise and less intuitively loaded terms. Please, do not mistake: I'm not appealing for obscure language, or for proliferation of new words. I'm appealing for the clarity that can come with ontological innovation.

Defending Against "Social Science Is Different"

I have three defenses against the claim that the above is simply an unwarranted extrapolation from physical to social sciences, which I can only briefly pursue. First, I believe all of those foci are epistemological, not just saying "cognition should be like physics." That is, they can be given motivation independent of their appearance in physics. I don't think, for example, that the theory dependence of data is at all unique to physics. I do believe that transcending commonsense frameworks is an important task to pursue, and a reasonable measure of success for any empirical science.

Second, let me demonstrate the care involved in selecting these points to extrapolate by listing characteristics I do not extrapolate.

1. *Mathematics.* I deliberately did not pick mathematization as a core characteristic to extrapolate. In the first instance, I believe explanation is a higher priority goal than mathematization. As well, I don't believe the mathematics of mind descriptions will be very much like the mathematics of physics; I expect it will be more like the formalisms of computation. This is, of course, a long story of its own, but it at least means simpleminded expectations about the form of knowledge and learning theories are to be guarded against.

2. *Sense of mechanism.* I don't believe the basic sense of what terms and forms are explanatory can be imported from physics. In particular, I don't expect that reductionist accounts, for example, a purely "brain science" approach to mind, will prove successful. The distinction between correlation and explanation is fundamental to any science, and deciding which is which is not a matter to prejudge on the basis of other sciences. My advocacy of theory in this paper is precisely to say we must do this for ourselves.

3. *Methods.* Every science needs its own methods adapted to its own theories and to the observational circumstances available to it. We can't blindly appropriate empirical techniques that work for sciences that have much more theoretically sound, or simply different, ontologies. In contrast to physics, I believe "empathetic techniques" that use (carefully and with many qualifications) our ability to sense our own thinking, and react instinctively to aspects of others' may be quite helpful. We don't have recourse to this in most areas of physics (though we do, in some degree, in our kinesthetic senses for the case of Newtonian mechanics).

Third, I explicitly recognize the many arguments against expecting theories in social sciences to be at all like those of physical sciences: "Social sciences are too complex and contingent to admit of theories of the sort we find in physical sciences." Or, "Social sciences are and must be fundamentally interpretive, not predictive." Without pretending to argue the points, I note that I simply have not found the arguments compelling for reasons like the following:

1. Such claims are too often simply assertional, without providing a theoretical basis for the meaning of the "fundamentally differentiating attribute," or how it opposes its supposed antithesis in the physical sciences.

2. Even if the distinctions turn out to be well-founded, one has the obligation to explain why they bear on the possibility of good theories. I don't see why the observer's being like the observed means that there can be no clean conceptualization of the observed.

3. Claims of intrinsic difference between social and physical sciences often are drawn from caricatures of physical science, far from what I experienced as a physicist. My experience of physics was of highly integrated explanatory systems that involved important ontological innovation. It was not of "narrow and mechanized prediction."

Similarly, to think that physical systems are easy to observe simply does not jibe with the fact that the appropriate thing to "observe" may be a wave function! There was plenty of argument and "interpretation" around in the early stages of any of the foundational physical theories.

Physical theory deals with systems of 10^{23} particles and chaotic systems that are, in some ways, strictly unpredictable. How, exactly, is the complexity of human systems fundamentally different so they are intractable by theory that resembles, only in some basic epistemological senses, physical theory?

4. Many of these claims seem to be simple restatements of the fact that we don't have good theories, drawing the conclusion, somehow, that we *can't* have such theories. "History shows that learning theory has had a poor track record in its application, in education." Of course it does. It also shows this has been true of every field of inquiry before it developed deep scientific foundations.

I've explained and, to some extent, justified my standards and judgment that we don't have excellent theories yet, but that they might be achievable. It is possible to think we are so far from *that* kind of theory that applying such standards to educational or psychological theory is ludicrous. I think, in contrast, that we may develop a tremendously helpful set of at least interim, if not absolute, standards and heuristic moves to advance our understanding out of the realization that we are not done yet. Realism is almost always the best policy. Although it is exciting to believe we're on the edge of really major breakthroughs, if we have not made them already, it is probably more important to have a cultivated sense of how far we have actually gone, and how far and in what directions we need to move. I prefer to avoid accepting "wimpy" epistemological standards that claim social sciences just won't ever and shouldn't strive to meet at least some strong standards in some respect like those physics has achieved.

As I have indicated how difficult I believe it is to achieve deep theoretical understanding, I am quite sure we will never achieve it if we don't set our minds to it. This is a kind of Pascal's wager I'm prone to accept: Unless there are compelling reasons to abandon searching for deep understanding that is in some ways like what we have in physics, we ought to pursue it.

Do We Really Need Theory?

I've treated, however briefly, claims that we *can't* reach the kind of theory in social sciences that has been achieved in physical sciences. In this section, I consider what we get from theory to bolster our resolve that it will be worthwhile before getting on with the program. Much that can be said about this will sound familiar and commonsensical. Yet I believe it bears reviewing in view of the apparent undervaluing of theory in the educational community. Of the many things that could be said, I'll select only a few.

The Scientific Power Principle.

Theoretical scientific understanding reliably yields capabilities that far surpass what we can attain by experience or intuitively-based empirical methods. Physics (lasers, nuclear energy), biology (recombinant DNA techniques), medicine (controlling viral and bacterial infections), technology (materials engineering, semiconductors and computer technology), and so on, all repeatedly show that theoretical advance is the linchpin in spurring practical competence. Even when a great deal of experiment and much engineering must be done, theoretical advance defines the parameters of experimenting (e.g., the terms of materials science), and establishes entire engineering domains (e.g., modern electronics emerged out of the basic quantum and materials principles that suggested the transistor could work).[6]

It is true that many aspects of our lives are entirely adequately handled by experiential or "purely empirical" approaches. You don't need Euclid's Axioms or General Relativity to navigate your house. Reading *Consumers Reports* and finding there a statistically reliable correlation between the measured reliability of a car and its brand is probably all you need to figure out which car to buy to have the best chance at getting a durable product.

Sometimes things are not so easy. Generating adequate power for our planet is not so easy. Building machines that fly is not so easy. I strongly believe designing for human competence, ranging in my immediate concerns from designing instruction to designing information machines for comprehensibility and effective use, is not so easy. I don't think it even needs argument that getting the most from our intelligence is a worthwhile pursuit. There is plenty of value, hence motivation for spending the time and effort to understand learning well.

"Because It's There"

One needn't be so practical about pursuing deep understanding. I believe our field is dealing with almost timeless questions. Physics approaches questions like: What are space, time and matter, and what accounts for their structure? Does the universe have an end; how could it? How did this all start? In the same way, I believe we all deep down want to know things like: How do we know? What are the limits of human knowledge? Why are people different from other animals; what does it mean to be intelligent, and are there fundamentally different types of intelligence? Such questions deserve deep answers. These are grand enough pursuits to make me very happy when I feel I've taken a small step. Realizing the scope of one's goals give meaning to the enterprise beyond the limits of present understanding.

Cumulativity in Science and Overcoming Barriers.

I have suggested already that theory is important to the infrastructure of science independent of implications for practice. "There is no data without theory." I suggested that developing standards and being critical of our explicit or implicit theoretical commitments is a prime method of improving our scientific understanding. I wish to point to two general and important infrastructural issues here.

The first is cumulativity. I hear echoes of Allen Newell's (1973) "You can't play 20 questions with nature and win."[7] His sentiments strongly parallel mine. One can't simply collect ad hoc hypotheses about what might influence what, and it is boringly non-cumulative to identify one after another little experimentally valid "phenomenon." Science needs a broader woof and warp. It needs breadth in order to supply focus. One simply must take stabs at overarching views so that the pieces fit into a larger context -- or don't, in which case we need another theoretical stab.

My reference to neutrino detection above can make another point. The "strong fulcrum of well-elaborated theory" I described in that story can disconfirm as well as confirm. For example, scientists might measure a tiny shift in the orientation of an orbit to (possibly) disconfirm Einstein's theory of relativity. It *has* to be that way, if Einstein is right, no ifs or buts. In a sea of "phenomena," of correlations without rigid underlying causal mechanisms, of heuristic but

----

6. diSessa (1991-a) describes some details of how the engineering context of learning theories might relate to the theories themselves.

7. Or see the first chapter of Newell, 1990.

commonsense ideas about knowledge and learning, no such disconfirmations are possible. There are always exceptions and extenuating factors. We don't know when exactly our hypotheses must apply, nor exactly what they predict. To take a case I introduced above, I believe that current disequilibration theories of learning are not disconfirmable. (Perhaps they are tautological, which is not the worst status possible.) Until we know exactly what disequilibration is, what processes generate it, and what processes are available to "select" a new view, and "change beliefs," we will always be able to fiddle with our characterization of a learning event to make it look like disequilibration.

Problems with a Theoretical Approach

I hear a couple of "Well, OK, but..." reactions to my line of argument to which I would like to respond. The first is the feeling that only special individuals, the Einsteins, Newtons, maybe the Piagets and Skinners, and so on, create theories. I am comfortable that grand moves might always be associated with individuals. Still, a field is not all grand moves. As I suggested, I believe almost every paper I have reviewed for journals could have been improve and clarified -- putting its results and non-results in clearer relief -- by some hard thinking about its hidden or missing theoretical commitments. I think small steps at clarity, generality, even to better fix the present state of the art, can accumulate. This may be more plausible to those who habitually see theory as always coming in identifiable, "world shattering" chunks after I make some suggestions (in the section on Some Almost-Practical Steps) about small things we can do on the way to more adequately addressing the theoretical side of the requirements of science in our community. Even if we accept the grand move hypothesis about theory, our community has a much better chance of cultivating or attracting individuals who can make those moves if we are more theoretically aware and intent. Perhaps we would be better at noticing and judging important theoretical moves in the making.

I anticipate one other reaction. It is easy to imagine that if theory-building becomes a more popular sport, journals will be filled with incomprehensible jargon and unsubstantiated speculation that now tends to characterize "theoretical" work. But I'm advocating "better" as much as "more." Future theorizing should be constrained by significant advances in a critical sense, which would prune away idle speculations. Indeed, as I suggested, the first signs of a more theoretical orientation will much more likely be self and other criticism and recognition of limits rather than just more theory.

Cultivating a Theoretical Turn of Mind: Some Almost-Practical Steps

The premise of this section is that the pursuit of theory is an excellent thing to do short of producing encompassing and revolutionary theories, as usually catch our attention. I've collected a short, ad hoc list of steps we can take toward becoming better theoretical thinkers. Many of these reflect things I've said above.

These heuristics for the development of theory are actually a fairly critical part of this essay. First, this is really the place I begin to define what I mean by theoretical thinking, short of standards for "having arrived." I hope it is evident that I have a broad interpretation of theoretical thinking, and I would argue that is appropriate. Second, if appeals to be better-oriented theoretically are to have any effect, they had better have particular, doable moves associated with them. I hope to get from this section reaction from colleagues on what they think constitutes theoretical work, and whether it is important and doable (or done!).

Some of these suggestions, especially the later ones, specifically single out students. I don't mean to imply that those suggestions are only for students, or that students shouldn't expect to get anything from the other suggestions. I do mean to emphasize the importance of students' training in changing a field, and also to point out some steps toward theoretical thinking that I think are either particularly easy or particularly important.

Almost every proposition we can formulate these days is as false as it is true. Try to understand why and when they are both true and false. This is a heuristic I've cultivated myself in reviewing journal submissions. It helps us discover the hidden contextual dependencies of our ideas, hence helps to define their real generality. It combats "confirmation bias." In addition, it asks us to be more explicit about what we mean so that one can make sure we have explained what our terms mean, rather than relying on inarticulate instincts that apply ideas only where we know already they work. The heuristic can be also used to be clear on the contexts in which our ideas have their intuitive roots. Armed with that, we can understand both a bit more about why and when our claims might be valid and adequately specified.

Is learning always best done in groups? Almost certainly not. Is cognitive apprenticeship the right method to learn any material? Can't be. Are novices always concrete and experts always abstract? Not a chance.[8] For all the social roots of individual cognition, I am confident there are also individual roots of social cognition.

If you can't decide, take a line and push it until it breaks. I frequently tire of papers that list all the possibilities of how the world might be configured to explain a phenomenon. Sometimes, anyway, we should be able to make good guesses that cut away broad ranges of possibilities and hence have important consequences. These are guesses that are worth pursuing in an extended way, in contrast to meandering among the many possibilities. For example, in my work with intuitive physics, I have quite deliberately made the decision to assume that such knowledge comes in identifiable bits, "atoms of cognition" if you like. I am quite aware I have precious little evidence to establish that fact, but I expect only to know whether or not, and in what way, it is true *if* I develop an elaborate theoretical scheme that defines precisely what "knowledge in pieces" means, and can draw extensive implications.

A complementary heuristic is to understand when you have made such a commitment, as opposed to believing every aspect of your thinking is justified by the weight of evidence. Many of our working assumptions are simply not justified in this way. It's worth our taking cognizance of that fact.

Arrange your work to be thematic, cumulative. I don't think it happens without effort that each of us (and, perhaps, communities of researchers as well) plots a coherent line. I think it is particularly easy to have an empirical program that does a little of this, a little of that, and moves on. Experimental methods seem much more transportable than theory. Yet, if we are to develop theory, we shall have to work coherently at it.

I see too much opportunism in the way research topics are approached. Mental models, "misconceptions," or collaboration become "hot topics," and many jump in. But they are also as likely to leave in a year or so as to make a deep mark. Of course, we must all decide when a line

----

8. I've applied this heuristic systematically in thinking about differences we instinctively apply to naive versus expert knowledge. This has become articulated criticism of some of the "expert/novice" literature. See, for example, sections on "concrete and abstract" and on "generality and specificity" in Smith, diSessa & Roschelle (in preparation).

is progressing and where the new opportunities lie. But we should also select our foci carefully enough that we believe an extended effort will be rewarded.[9]

Question ontologies; refine categories. I've transplanted my suspicion that deep theoretical advances are always accompanied by seeing the world in new and different terms into this heuristic. What are "concepts" or "entrenched (or any other kind of) beliefs"? What is "metacognition," "a community practice," "an educational activity"? Questioning the analytic and empirical meaning and adequacy of these categories expresses skepticism about the precision of nearly commonsense ideas that substitute in much current work for what should be technical terms in well-developed theory. Questioning meanings also expresses a feeling that a pursuit of what we instinctively mean by these words can be clarifying. Of course, this could become an armchair game. The enterprise works best in the context of empirical study that tests the work more refined terms might do for us.

I find myself questioning my instinctive categorization of instances all the time. It would be a worthwhile enterprise to catalog strategies for making these tests. Such questioning episodes turn frequently into pursuing clearer meanings for terms -- operationalizing them or framing them better in order to afford both easier classification of instances and also clearer import of classifications that have been made.

Make the most of "what we know for sure." Physics has a few things that it knows for sure. Symmetry considerations are among them. As well, it knows that all physical interactions must be local in space and time. Although things "we know for sure" may seem general and bland, in the hands of the best physicists they have proved amazingly powerful and particular. They seem, especially in combination, nearly to "deduce" particular physical laws.[10] Surely we must have, or should be looking to find, similar principles in education or learning psychology. What are they? I'll leave this heuristic open as a good litmus test concerning how we think our field is or will ultimately be organized. It might be that most readers will simply not know what I am talking about. Or, alternatively, they have their list, or believe there can be no such list.

Let us think what appropriate empirical work, data collection and analysis, might be like to serve theory building. I am convinced that our arsenal of empirical methods are skewed tremendously toward confirming or disconfirming hypotheses that are assumed to be well-formulated rather than toward building an adequate basis for making hypotheses, or testing the well-formedness of our ideas in contrast to testing their truth or falsity. I believe empirical work can play a vital role in developing theory, but this role and methods that fit it are undervalued and underdeveloped. I would love to see a good course and text developed around empirically grounded theory development.[11]

----

9. Early in my professional formation, I was influenced by Howard Gruber's concept of a "network of enterprise" (Gruber, 1981) to describe how creative individuals manage to pursue a sufficiently diverse yet cumulative, and mutually reinforcing set of lines of inquiry. I sat down and designed my near-future network. I believe, in retrospect, that was an important step for me.

10. Feynman (1965) wrote a beautiful little book on this. I have also been tremendously impressed by the work of scientists like E.P. Wigner, and Einstein in this regard.

11. Perhaps I am defensive, but I believe some of my empirical work has been misunderstood as not-so-good theory confirmation, when I view it as more-than-usually-conscientious data sensitivity for the purpose of theory motivation, specification and development.

Cultivate a sense that explanation is the name of the game. When people begin to play the game of science, their first glimmers of understanding it are that science is about finding the way things are. Science finds "that" X or Y. More deeply, I take it that science is explaining "why and how" things can be the way they are. Of course, there should be a few "thats" in science, that F equals ma, for example. But these "thats" must have thousands of "and therefores" following them. In general, observations must be carefully placed in an explanatory web.

I think versions of this primitive "that" orientation are insidious and long-lived. As I mentioned, many too many papers talk about the existence of a phenomenon without pursuing underlying mechanism. In education, prescription substitutes to an amazing degree for adequate understanding of underlying mechanisms. To parody, "We know that to teach well, one should *do* X." I find this in some degree even in some of the best work in the field, or at least in the field's (if not the investigator's) take on the work. Reciprocal teaching inappropriately becomes a principle rather than a technique.

Taking instructional prescription as mechanism is essentially a category error. Instruction is an area of complex design. I don't expect deep principles of learning will often if ever show themselves on the surface of an effective design. Of course, this fact makes our job harder -- we must both understand the principles behind instructional interventions, and we must understand the contexts of application of those principles well enough to know that the principles are truly involved and do the central work we might claim for them.

I've been struck by a characteristic of most of the most creative and deep thinkers (of course, in my judgment) I have known. They are constantly on the alert for interesting phenomena, where, perhaps, a fundamental piece of the world breaks through its mundane presentation, or, as interesting and likely, where we find a deep intuition confounded. They take the time to look again, recreate, modify, and make a proposal for both an explanation and for why the phenomenon is puzzling in the first place.

In some respects, this behavior seems unprofessional. It is amateurish because these individuals frequently have no specialized interest or knowledge about the phenomenon at issue, why bottled water fizzes in a particular way, or how geological formations of a particular sort might have come into existence. But I have come to feel that these entertaining little escapades are both telling and important. They tell us that being alert to the odd moments when nature reveals herself to us is a high priority enterprise. It is an enterprise of observing, reflecting and explaining, which some people cultivate or do naturally. These people have likely acquired some generally useful skills with respect to this enterprise, and probably find it both entertaining and profitable to exercise even away from their domain expertise.

I find the instruction in cognitive science and education unusually devoid of such spontaneous pursuits. Too often students are expected only to be "library indices" to sanctioned data, knowing the results of the field, thinking to observe and comment on only things others have declared comprehensible or empirically tractable. Students don't think much about their own experiences in learning, or what they make of others', except as filtered by the sanctioned state of the art. Though the focus of this little, perhaps dubious, indicator of a more general theoretical orientation may be misplaced, I find similar indicators again and again in deep thinkers. These are almost never reflected in our training.

Create Mini-Theories. There is a slightly more professional version of the activities described above. That is to formulate little mini-theories about important issues in the field, and use them to accumulate and refine ideas about what must or might be true. The criteria for these mini-theories are not ad hoc. First, they ought to be about important things, so the time spent on them is worth the effort specifically concerning conclusions (as opposed to the process orientation, above). It also helps a lot if they are counter-intuitive, to test the strength of our "knee-jerk" dispositions that arise from implicit theoretical orientations. Frequently, mini-theories occur to me in the process of thinking, "That seems strange, but there's something appealing about it, and it might explain some very puzzling phenomena."

I find these are the kind of things from which programs and theoretical ideas grow. For example, my own "theory" of intuitive physics arose from two at the time counterintuitive (to me) mini-theories. One was that cognition is radically unsystematic. As I put it to myself, every idea is a different form. The second was to presume that we could identify a large set of what appeared to me at the time to be a few cute little intuitions you could trick people into displaying, and that, in fact, causality was constituted of a whole body of such entities, rather than being localized in general principles of cause. The latter seemed particularly counter-intuitive at the time, but I could not see how to dismiss it out of hand. And since causality had proved so elusive, maybe people were looking in the wrong place. These mini-theories developed into a fairly elaborate theoretical and empirical program, of which they are still good motivators or hooks to explain the gist of the program (diSessa, in press-a).

A recent mini-theory of mine is that the robustness of scientific "misconceptions," which is touted in the literature about them, is mostly constructed in encounters that are intended to expose and overcome them. This contrasts with presumptions that misconceptions are inherently stable, and hence *must* be attacked. Instead, people may only formulate positions when asked to. But once asked, they can build rather resilient ideas out of what might otherwise be fleeting impressions. We may then be doing exactly the wrong thing in "attacking" misconceptions. I wouldn't pretend to defend this statement scientifically at this point, but it will orient some of my thinking, and I believe it *might* turn into a collection of defensible claims. One of the properties of this mini-theory is that it challenges some of my own presumptions, as well as those I feel others have inappropriately taken up in their work. So now the game is: What could this mean? Could we demonstrate that it is definitively false, thus simply drop it?

Formulating and pursuing mini-theories strikes me as not only a reasonable practice for professionals, but, with guidance, a good and tractable finger-exercise for students.

Redescribe, redescribe, redescribe. Students particularly suffer from the feeling that the world presents itself directly to them, that intuitive characterizations define exactly the circumstances in which we can use those terms and descriptors. This is profoundly false. Our future colleagues need to understand this and need to play a better game of formulating and judging descriptions as soon as possible. I am especially fond of redescribing educational practices that students find instinctively repellent in terms that they use to describe good practices. We propagate attitudes rather than clear conceptions about instruction by only using words that sound laudable (or the reverse) to describe particular practices. Of course, redescription is not only to get students to rethink judgments and their bases, but to articulate and refine the meanings of the terms that seem clear and apt, but may not be either well-defined nor apt.

Cultivate a sense for the "big issues" in the field. I've underlined how difficult yet central I believe theoretical considerations are, and how important it is to generate a coherent program to make advances. Students especially need to know where the field is, how to measure the latest fads, and how, in general, to calibrate progress they or others might make. It is often "schoolish" and vapid to announce what a field is about. The first chapter of textbooks that explain "what physics is," or psychology, are usually crushingly boring and uninformative. Yet the responsibility of keeping track of our advances on a large scale is critical, and we should not shirk it.

Identify, practice (and give students opportunities to practice) basic theoretical moves. The subproblem here is a particularly interesting one. What *are* basic theoretical moves? This is the parent problem of several of the above suggestions. Identifying basic theoretical moves not only defines the practice of being theoretical, but it also explains in a more explicit way what is or should be meant by theoretical work and what are central as opposed to peripheral parts of it.

For example, the heuristic "redescribe" tells us that the terms in which we describe the world are as important an object of study as finding the "right" propositions using the terms we already have. Heuristic strategies of evolving more precise and powerful descriptions are thus a central set of moves in making theoretical advances, of which "do it" (redescribe) is the crudest.

A basic move I find myself rehearsing explicitly and self-consciously for students embarked on their theses might be called the "characterize, systematize, re-examine loop." Typically, one immerses oneself in data, using whatever initial predilections for analytic frameworks one has at one's disposal. Usually one comes out having found a number of critical phenomena -- happenings that can be somewhat effectively characterized in available terms and seem also to be critical in one way or another. Then, one takes the terms of description, categories, and implied or conjectured relationships among them and tries to complete and systematize the story. What could a generic characterization of such knowledge be? Why might this relation hold? Is it an example of a more general relation, or what co-requisite (but undescribed) circumstances might make the relation more comprehensible and "necessary"? With a more articulated, complete and more evidently causal story to tell, we need to return to the data. Can we see the differentiation of contexts implied? Is there, in fact, only one critical feature, or is the phenomenology of our data much more diverse than we presumed? Do the new categories developed in the second phase help make better sense of the data?

The second phase is one students especially need coaxing to do. It's not an obviously workable tactic in an empirically dominated world view. It seems rather rationalist -- how can we find ambiguity in terms, extend items to "a full list," and so on, without looking at the data? Yet, this is where theory originates or is iteratively improved. We not only can, but we must be analytic and systematic in reordering existing perceptions and observations, in sharpening the meanings of categories that define how we see things, in completing fragmentary patterns, which gives us new eyes to check the data.

Summary

Theory is a tough goal to maintain in the face of the state of the art in learning and instructionally oriented investigation. It would be easier if we could just "bail out" and think we were more like "literary critics" of practice, or artisans fabricating all-the-time better, but unprincipled artifacts. I think we should face up to the fact that it is very likely we could, if we chose to, be a science in the making, however limited our present powers. If we do not critique our work by high standards, then we will certainly delay obtaining the kind of power deep scientific understanding might bear.

I have tried to advance an image of theory building that is incremental and heuristic as much as it is a set of simple, hard standards by which we will know when we are done. In fact, I've really avoided the "standards" view for the most part, except to give a sense for why I judge we are not far along on the path to excellent theory. The heuristic view of theory building is especially important given that no one can say with much certainty how much future learning theories will look like the excellent theories we know in other domains. It is also simply more important to know how to move things forward than it is to know when you are done. So, theory-building can be hard-nosed in its goals, but at the same time generous and truly exploratory in its active parts.

As a community, I am arguing we should exercise more effort in and attention to theoretical matters. We should cultivate a critical capacity to understand modest advances at the same time we recognize the many types of limitations of existing theories. I think we should share and systematize methods to improve our frameworks. Most especially, I urge we scrutinize, articulate and refine the theoretical moves we've all intuitively developed and found powerful. We should do this for the benefit of our students, for our colleagues, and, especially, for ourselves.

Acknowledgements

I owe a debt to my mentors for teaching me what I have thought valuable enough to try to convey here. I wish to thank Steve Adams and Jack Smith for extensive and thoughtful commentary on a draft that influenced the form and content of this paper. Ehud Baron similarly stimulated my thinking about the topic with his reactions.

References

Carpenter, T. P., Moser, J. M., & Romberg, T. A. (1982). *Addition and Subtraction: A Cognitive Perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner and A. Stevens (Eds.), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum. 15 - 33.

diSessa, A. A. (1986). Models of computation. In D. A. Norman and S. W. Draper (Eds.), *User Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum, 201 - 218.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman and P. Pufall (Eds.), *Constructivism in the Computer Age*. Hillsdale, NJ: Lawrence Erlbaum, 49 - 70.

diSessa, A. A. (1989). A child's science of motion: Overview and first results. In U. Leron and N. Krumholtz (Eds.), *Proceedings of the Fourth International Conference for Logo and Mathematics Education*. Haifa, Israel: Israeli Logo Center, Technion -- Israel Institute of Technology, 211 - 231.

diSessa, A. A. (1991-a). Local sciences: Viewing the design of human-computer systems as cognitive science. In J. M. Carroll (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface*. NY: Cambridge University Press, 162 - 202.

diSessa, A. A. (1991-b). Epistemological micromodels: The case of coordination and quantities. In J. Montangero and A. Tryphon (Eds.), *Psychologie genetique et sciences cognitives*. Volume from the Eleventh Advanced Course. Geneva: Archives Jean Piaget.

diSessa, A. A. (in press-a). Toward an epistemology of physics. *Cognition and Instruction*.

diSessa, A. A. (in press-b). Speculations on the foundations of knowledge and intelligence. In D. Tirosh (Ed.), *Implicit and Explicit Knowledge: An Educational Approach*. Norwood, NJ: Ablex.

Feynman, R. (1965). *The Character of Physical Law*. Cambridge, MA: MIT Press.

Gruber, H. E. (1981). *Darwin on Man: A Psychological Study of Scientific Creativity*. Chicago: Chicago University Press.

Karmiloff-Smith, A., & Inhelder, B. (1974/75). If you want to get ahead, get a theory. *Cognition*, 3(3), 195 - 212.

Newell, A. (1973). You can't play twenty questions with nature and win. In Chase, W. G. (Ed.), *Visual Information Processing*. NY: Academic Press.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Smith, J. P., diSessa, A., & Roschelle, J. (in preparation). Misconceptions reconceived.

Appendix: A Theoretical Orientation

I deliberately avoided discussing particular theories or theoretical orientations, for the most part, in the body of this essay. This was to avoid contentious detailed issues that could easily obscure the main points. But, theory-building is not a "meta" exercise. Scientists must take on or develop an orientation toward theory, and find the classes of theorizing they believe appropriate to the subject matter they are investigating. I am advocating that we articulate and advocate particular lines. I wish to do a little of that here.

Every researcher develops a particular "sense of mechanism" about what the basic principles operating in a domain are like. I believe this is a precious personal and community resource that guides observation and generalization, but it needs explicit consideration. If you think theories will look like prescriptions, that's what you will develop. If you think "thick descriptions" are explanatory, you won't develop other kinds of explanations. If you believe that a particular social relationship can define learning, or that no description of knowledge "in the head" is relevant to learning, you won't pay attention to the structure of content domains.

My instincts are that we must develop mathematical-computational theories of mind and learning. I am drawn to current attempts to do this on several accounts. First, there are at least languages of analytic precision in play. This also builds in some strong mechanisms for testing the ambiguity or sufficiency of the ideas involved, and for surpassing reliance on intuitively attractive, but "magical" ideas about the way things may work that common sense provides us in abundance. There is plenty to criticize about most present computationally formulated theories, but I don't see the sense in denying the ways in which they are attractive.

On the other hand, I don't yet insistently couch my own ideas in these terms. This is a judgment that we haven't got the mathematical-computational foundations quite right yet. Most directly, the best developed theories in this area (and they are better formed by many standards than "theories" belonging to many other traditions) just don't, in my judgment, reach the issues or touch the empirical phenomena I am most interested in pursuing, mainly those dealing with conceptual change and long-term conceptual and intellectual development.

The crux of this lack of contact, I believe, is that current theories just do not get to the heart and power of knowledge. More specifically, I believe there is a tremendous diversity to the kinds of knowledge and systems of knowledge that one can find. Essentially all computational theories are much too "flat" and uniform, to my taste, suggesting much more uniformity than I believe exists. I believe I perceive many different subsystems of human knowledge that have very different properties, which properties I don't know how to describe in the terms of these theories. (Or better, I don't see how the precision of the theories improves the apparently looser descriptions I make outside of them.)

This leads directly to a general program for studying thinking and knowing. It is roughly at the level of knowledge itself, though one needs to have at least a minimal sense of computational mechanism in order to see how pieces of knowledge relate to one another, and how the system functions dynamically. The basic plan is, roughly, to develop a sense for the grain size of knowledge elements and of their rough individual properties, but then the real business is to describe the system properties of these elements. How "densely" are the elements interrelated? Are they tightly interconnected and used almost always in contexts of the same other elements, such as elements of a skill that are activated only in patterned sequences of that skill deployment? Or are they very loosely interconnected and fluid in their composition in particular thought contexts? Can we describe the functions of the particular system at issue and how they join with other systems to perform more complex functions? Are there mechanisms that produce levels of systematicity other than those that have to do with performance? For example, do some core set of ideas in some sense derive the rest, though derivation is not the usual mode of operation of the system?

I've developed two exemplars of knowledge system analysis. The first is my analysis of intuitive ideas in physics.[12] Roughly, my claims are the following. Intuitions about causal mechanism reside in a large system of fairly simple elements that are only loosely connected. The function of the system is to provide judgments of how adequate a description explains why one should expect a particular thing to happen. The elements are configurations of circumstances that "just happen" and need no further explanation. Trying to figure out how a physical system works or what will happen is trying to find an optimal description of the situation in terms of these causally primitive elements, and one that best matches the conditions under which each of the elements is understood to apply.

This knowledge system does "judgment." It does not solve problems per se, or even specify very much about how an individual improves his current best decomposition of a problem situation into causal primitives. As for levels of systematicity, the system is mostly ad hoc, consisting of individual abstractions that are particular to some class of situations and just don't apply to others. Typically, only a few primitives apply to a problem situation, and connections of the elements are also mostly ad hoc, determined by the situation instead of general patterns of use of multiple elements.

On the other hand, there are some higher level systematicities that are useful to know. There are a few families of primitives that share a "base vocabulary" of descriptive terms. In some cases, a family of causal primitives share a central common abstraction, for example, one abstracted from agentive interaction: a "willful" (in some sense) agent, a patient, and a legitimized, but always directed "influence type." Pushes and pulls are canonical examples. Some of these families are important in identifying problems in learning, such as the need to undermine an entire class of primitives and support a new class.

This knowledge system analysis has educational implications. The principal one is that conceptual change is a system issue. It is hopeless to believe you have found the core of intuitive "misconceptions" and can argue the core away for students, leaving the conceptual field free for new conceptions. Instead, the whole problem must be conceived as an elaborate reorganization (not replacement). One must attend to system issues in learning, not just "one-at-a-time concept learning." In addition, knowing the existing intuitive primitives constitutes knowing the basic resources that must be reorganized, and establishes particular targets of difficulty, but also opportunities to build on some particularly apt corners of the naive system. "Engineering" is an appropriate metaphor for instructional design, since the richness, generativity and diversity of the naive system means there will likely be many opportunities and possibilities, no one "right way to construct the new system."

The knowledge system of causal judgments I have described is really a system of problematic descriptions. They are problematic because they prescribe the "deep causal structure" of a situation, which may frequently not be immediately evident. On the other hand, people also have "strong and reliable" descriptive capabilities, for example, in the area of spatial organization, and possibly dynamic spatial configurations. This is a different kind of system that may be the intuitive base of more mathematical ideas rather than physical ones. It is one I intend to study in future work.[13]

The second area in which I have developed a knowledge system analysis concerns understanding complex computational artifacts -- programming languages. In this context I claim to have

----

12. A quick sketch of these ideas is available in diSessa, (1983) or diSessa (1988). A thorough treatment will appear in diSessa (in press-a).

13. See diSessa (1989) for some very preliminary results concerning dynamic spatial reasoning.

developed a short taxonomy of systems, which I describe as types of mental models, that have complementary structures, strengths and weaknesses, different learning trajectories, and to some extent also complementary functions. Learning a programming language is viewed as building and articulating properly all of these systems. Designing a comprehensible system is creating one that has good properties with respect to all of these systems.

In this area of mental models, I believe it is important to understand not only the structure of the systems involved, but their properties in several different modes in which they may be used. That is to say, the system may be complicated enough that it may configure itself in several rather different patterns.[14]

Most recently, I have tried to extend knowledge system analysis into a general view of the *evolution* of knowledge systems. I've tried to define a general scheme of causality by which one system may transform into a different one. This work is, at present, very speculative. While it might prove to be very general and possibly powerful theoretically, connections to empirical work are weak. In contrast to the work with intuitive physics and mental models of computational systems where the knowledge system analysis followed as a systematizing phase of a "characterize, systematize, re-examine loop" (see text, in the section on Cultivating a Theoretical Turn of Mind), I am attempting this work more top down. Thus, I've tried to "build the theoretical system" first, to some extent, rather than doing a more bottom up first pass through data relating to an approachable example.[15]

----

14. diSessa (1986) gives a brief introduction to this kind of analysis. diSessa (1991-b) gives a thorough treatment.

15. diSessa (in press-b) presents the program describe briefly here. diSessa (1991-b) tries to bring it a step closer to empirical development and test.