Theoretical Foundations for Intelligent Tutoring Systems

John Self

1990

AAI/AI-ED Technical Report No.45

Computing Department
Lancaster University
Lancaster
LA1 4YR
UK

aai@comp.lancs.ac.uk
http://www.lancs.ac.uk/computing/research/aai-aied/
ftp.comp.lancs.ac.uk (148.88.8.9) /pub/aai/

Theoretical Foundations for Intelligent Tutoring Systems[1]

John Self
Computing Department
Lancaster University
Lancaster LA1 4YR, U.K.

Abstract

This paper considers the case for formalising aspects of intelligent tutoring systems in order to derive more reliable implementations, as opposed to the present use of informal theories to build experimental systems which are then studied empirically. Some recent work in theoretical AI is suggested as a possible source for the elements of a 'theory of ITS'.

Introduction

The engineering of any complex device (such as an ITS) gradually relies less on empirical experimentation and more on mathematical or scientific theory. As yet, there is no significant 'theory of ITS': all of the recent ITS texts (e.g. Wenger, 1987; Mandl and Lesgold, 1988; Polson and Richardson, 1988) are entirely discursive and attempt no kind of formalisation of their content. The aim of this paper is to suggest that it is not premature for ITS research to begin an attempt to complement a short-term emphasis on pragmatic aspects (Kearsley, 1989) by seeking theoretical foundations for its implementations.

Most AI researchers regard ITSs as peripheral applications of AI, an understandable opinion in view of the virtual absence of ITS papers from the major AI journals and conferences. But Clancey (1986) has argued that work on ITSs is not a "mere matter of putting well-known AI methods into practice" but is (or should be) "broadening the meaning of AI research".

Historically, ITS research began within AI, but AI researchers have retreated from the ITS arena as they have come to appreciate the need for more fundamental work on mental models, language understanding, knowledge representation, etc., leaving others to move into an intrinsically multi-disciplinary field. However, if there is ever to be a formal theory of (aspects of) ITS then it will be derived from elements of AI. Moreover, recent AI research begins to indicate what those elements might be.

What is an ITS?

To understand new concepts we often resort to analogy. The immediate analogy for an ITS is the only other agent able to perform a similar task, the human teacher. The *ITS as human teacher* analogy pervades the ITS literature - in fact, it is so pervasive that it is difficult to bear in mind that it is only an analogy. Like all analogies it can be over-stretched. Studies of human teachers in real settings show, for example, that "about 35% of tutors' comments had some motivational or affective content" (Lepper and Chabay, 1988) and that in 327 minutes a tutor made only 3 utterances which might be interpreted as using a student model or teaching history (Kamsteeg and Bierman, 1989). To seek to use analogical inference to draw ITS design principles from such observations is risky, especially when we compare the context of these observations to that facing a typical ITS user.

---

[1] based on an invited paper at the 4th International Conference on AI and Education, Amsterdam, May 1989.

For a school-based analogy, we might do better to move from the ordinary classroom to the gymnasium. We enter a gym expecting to find specialised, expensive equipment, to be assisted by trained coaches, and to suspend certain social conventions. We expect to leave, after strenuous exercise, physically invigorated. This, rather than the leisurely atmosphere of the ordinary classroom, gets closer to the frame of mind in which we may expect a user to come to an ITS.

To illustrate the *ITS as cognitive gymnasium* analogy, consider the Decider system (Bloch and Farrell, 1988). This system attempts to help students develop opinions about political history by repeatedly presenting case histories which cause the student to reflect on opinions just expressed. In the ordinary classroom, this relentless style of interaction would soon pall but for the student who wants a real cognitive workout, who wants his opinions probed, stretched and tuned, it might be welcome.

These two analogies are both concerned with the style and role of finished ITSs - they are not directly concerned with the process of ITS design. At base, ITS design is an engineering enterprise since the objective is to build technological systems which perform effectively. Perhaps there is something to be learned from other engineering design problems, e.g. aircraft design.

The science of aircraft design, aeronautics, is "a blend of beautiful theory and empirical fine tuning" (Shevell, 1983). Of course, it has not always been so. The first aircraft were built by enthusiasts like the Wright brothers with no formal training in engineering. The 'test flight' was an occasion of great drama and apprehension. It was also very dangerous (Orville Wright had two serious accidents before killing an observer in 1908) and likely to provide only an uninformative pile of wreckage. Very soon, the test flight was largely displaced by more specialised environments, such as wind tunnels, and by the elaboration of the theory of aeronautics.

As it happens, the elements of aeronautics already existed in nineteenth century physics, within Newtonian mechanics, hydrodynamics and fluid mechanics. Unfortunately, early aeronautic theories never gave the 'right answers' (i.e. they did not accord with the empirical evidence). The effect of viscosity had been overlooked, and although only a minor modification to the theory its inclusion gave very different results. We must assume that this is always possible (especially for any putative theory of ITS!). But this is no reason for not striving for a theoretical basis for our engineering designs.

Towards a theory of ITS

Principles of ITS

Several authors have suggested informal principles which may be the beginning of a 'theory' (Burton and Brown, 1982; Collins and Stevens, 1982; Anderson, Boyle, Farrell and Reiser, 1986; Ohlsson, 1986). But these are different kinds of principle to the foundations of aeronautics (Bernoulli's equation, Helmholz Vortex theorems, and the like). The principles are not foundational but more like lemmas derived from an underlying theory. Anderson's eight principles, for example, are supposed to follow from his ACT* theory of human cognition. The principles do not follow in any formal sense, however, and indeed it is arguable whether they follow at all. And just as there is a gap from 'theory' to principle, so there is a gap from principle to implementation: the principles do not determine an implementation and it is not possible to say categorically whether an ITS has been implemented in accordance with the principles or not.

Architectures for ITS

If it is not conceivable at the moment formally to derive implementations from a theory, perhaps we may instead work back from implementations towards a theory. There is a remarkable consensus on the 'standard architecture' for ITSs: an ITS consists of components which know about the subject matter, the student, and tutoring. This is the organisational framework presented in all the recent ITS texts. For example, Polson and Richardson (1988) introduce it in the very first paragraph, as a definition of an ITS.

We are free to define an ITS as we see fit - Anderson (1988) considers that "by definition, ITSs can be built only for domains for which expert systems (criterion-defined) exist" - but an unquestioning adoption of the traditional trinity model will cause problems in ITS implementation and will restrict the scope of ITS research (Self, 1988a; Cumming and Self, 1989a). Here, however, the main concern is that the traditional trinity imposes a philosophy which is not an appropriate basis from which to develop a theory of ITS.

The philosophy of ITS

Wenger (1987) considers that ITSs are concerned with "knowledge communication" and Ohlsson (1986) that the goal of teaching is "to transmit a particular subject matter to the student". While these phrases are subsequently elaborated to provide a richer view of the educational process, they do express a common philosophical point of departure for ITS research.

Such an 'ITS philosophy' runs counter to almost everything of significance in twentieth century educational philosophy. All the major figures (Dewey, Montessori, Piaget, Rogers, even Skinner) have rejected the 'education as transmission' model which had dominated the previous three centuries in favour of an 'education as growth' model. So our ITS philosophy is in danger of being regarded as obsolete.

'Knowledge' is not the kind of commodity which can be transmitted. It cannot be simply received by students but must be constructed anew by them. While most ITS researchers declare such a view, it is belied by the systems we build. The emphasis is on designing 'correct' representations of a domain to serve as a basis for overlay models and bug catalogues and to promote learning through the accretion of knowledge and the remediation of what are deemed to be errors with respect to our correct representations.

This ITS philosophy derives from a commonsense theory of knowledge, which holds that items of knowledge exist in an objective sense in the external world and that we can acquire knowledge from the world, via our senses or teachers or ITSs. But according to most contemporary epistemologists, knowledge, even in the natural sciences, is conjectural. Knowledge grows mainly through criticism. Students should therefore be encouraged to view knowledge not as received wisdom but as the fallible creation of human beings.

If we take 'knowledge' as 'justified, true belief' (the opening gambit in philosophers' discussions of 'knowledge') then it is contradictory to regard the student model as describing the student's 'knowledge' when it includes descriptions of misconceptions. It would be more sound to regard the student model as describing a student's beliefs, not knowledge, and, if we are honest, to regard the domain model as describing our own beliefs. The 'education as growth' model may then become partly describable in terms of a process of 'belief growth' or 'belief revision'. With this philosophical basis, we may turn to the concerns of ITS research to seek an appropriate theory.

The elements of a theory of ITS

Socratic tutoring

Let us first consider the now rather hackneyed dialogues analysed by Collins and Stevens (1982):

Can you grow rice in Missouri?
Yes
Why?
Because it's hot.
What about Arizona?
...

This dialogue relates to various propositions, e.g.

```
p1: Wet(Missouri)        p5: Hot(x) & Wet(x) -> Rice(x)
p2: Hot(Missouri)        p6: Hot(x) -> Rice(x)
p3: Dry(Arizona)         p7: Rice(Missouri)
p4: Hot(Arizona)         p8: Rice(Arizona)
      ...
```

At the beginning of the dialogue let us assume that the student believes the meteorological propositions: `B(s,{p1,p2,p3,p4})`. However, students do not reason with *all* their beliefs but only with that subset which they are 'aware' is relevant. The first question may make the student aware of those beliefs containing terms mentioned in the question, which may perhaps be expressed as an axiom:

```
Ask(t,q) & Contains(q,term) & B(s,p) & Contains(p,term)
        then A(s,p)                                    ..(R1)
```

where `A(s,p)` denotes that the student `s` is aware of the proposition `p`. So after the first question we have `A(s,{p1,p2})`. We will use `P` to denote the set of propositions of which the student is aware.

We may assume that if the tutor asks about a proposition `q` which is implied by `P` then the student will answer "yes":

```
Ask(t,q) & A(s,P) & P->q  then Answer(s,"yes")    ..(R2)
```

'Implied' in this context may be resource-bounded, since we cannot assume that students will make all the possible implications from `P`. However, here it is not the case that `P->q`. So the tutor may seek other propositions of which he is aware which might explain the student's answer:

```
A(s,P) & ~(P->q) & A(t,r) & {P,r}->q
        then it is possible that A(s,r)        ..(R3)
```

If we assume that the tutor is aware of the propositions `p5` and `p6` (that is, he is aware that they are potentially relevant, not that he necessarily believes either), then the above rule will suggest that `A(s,p5)` or `A(s,p6)`. Hence the question "why?" may be intended to resolve a possible ambiguity. (There are other possible reasons for the "why?" - the tutor may have no or only one explanation for the "yes", in which cases the tutor may be asking the student to state a new proposition or to confirm the one the tutor has.)

The "because it's hot" leads (eventually, through some analysis of the content of the propositions, which is beside the point here) to `p6` being added to the set `P`. The tutor's objective may be to guide the student into situations which might promote refinements of the set `P`. For example, he may hope that the set `P` will be refined if it becomes inconsistent:

```
A(s,P) & P->r & P->~r  then Refine(s,P)            ..(R4)
```

(The definition of "`Refine`" is touched on below.) So the tutor may try to find a proposition `q` which may be asked of the student:

```
A(s,P) & P->r & A(t,q) & {P,q}->~r  then Ask(t,q)  ..(R5)
```

For example, if `A(t,~p8)` then `p8` may form the subject of the question.

There are two distinct rationalisations for the "Arizona" question, depending on whether or not the tutor believes that the student believes `p8`. If we do not have `B(s,p8)` then the question is presumably the start of a sequence which we will eventually lead to the tutor asserting `p8`, so that student becomes aware of it and (through `R4`) refines `P`:

```
Assert(t,p)  then  A(s,p)                          ..(R6)
```

If, however, we do have `B(s,p8)`, then just asking the question may lead the student (through `R1`) to become aware of `p8` and hence (through `R4`) to refine `P`

- in other words, the question is intended to make the student immediately aware of the inconsistency of his beliefs.

This analysis is admittedly sketchy, but it may be a beginning. The main theoretical construct used is the distinction between implicit and explicit belief (Levesque, 1984; Fagin and Halpern, 1987), which is central to the active area of logics of belief in theoretical AI. The process of Socratic tutoring is formalised in terms of the drawing out of a student's implicit beliefs (or "thoughts which need to be awakened into knowledge", as Socrates put it). The Collins and Stevens' principles can then be interpreted not as generalisations of empirical observations, but as the outcome of reasoning based on descriptions of the participants' beliefs.

This theoretical baggage is rather heavy for so little gain (after all, we already have the Collins and Stevens' principles for Socratic tutoring). But perhaps the theory enables us better to understand the principles, to appreciate the assumptions on which they are based, and, in the longer term, to derive further principles. Perhaps also we may proceed to apply related areas of theoretical AI to aspects of ITS design.

Points of view, analogy and metaphor

The idea that there is a single, correct domain representation to be transmitted to a student is, of course, a fiction. Any body of knowledge can be represented in a number of ways, corresponding to different viewpoints. Often, we encourage students, through analogy and metaphor, to develop intermediate 'incorrect' viewpoints to serve as stepping stones to a desired viewpoint. Sometimes, students bring to bear viewpoints which we do not intend. For example, studies of the learning of physics (Osborne and Freyberg, 1985) and second language learning (Ellis, 1986) have demonstrated that we cannot simply communicate the knowledge to be learned but must take account of what the student already believes.

Osborne and Freyberg describe four common viewpoints that students hold about the currents in the two wires from a battery to a bulb. Students can give some kind of rationalisation of their own viewpoint and can also sympathise with alternative viewpoints when they are described. In solving a more complex problem, students are liable to switch from one viewpoint to another. If a student does not already have a viewpoint, he may create one by analogy, e.g. to the flow of gas to a cooker. Before we hasten, in the ITS style, to label all but one of these viewpoints as incorrect and in need of remediation, we should reflect on the fact that Ampere himself believed that an equal current passed along both wires from the battery to the bulb. It may be more productive to seek out the justifications for viewpoints and to devise experiments which may falsify them.

Whereas the previous section considered the incompleteness of belief sets (not all consequences of beliefs being drawn), here we are concerned partly with the problem of inconsistency: under different viewpoints it is possible for an agent to believe both a proposition and its negation. The provocation of such conflicts and their subsequent resolution is a key educational strategy, but it is one which if it is to be described formally must be done so with great care - for inconsistent systems can lead to technical difficulties. Kaufmann and Grumbach (1986), Fagin and Halpern (1987), Wilks and Bassim (1988) and others have all presented formal treatments of the problem of viewpoints.

Reflection

On the whole, we reason with our beliefs not about them. But recently it has been argued that learning is essentially the result of a reflective process: we learn not so much from doing but from reflecting on what we are doing or have done (Collins, Brown and Newman, 1987; Brazdil, 1988). Significantly,

this emphasis on reflection in ITS is paralleled by discussions of metacognition in education (Schoenfeld, 1985) and meta-level architectures in computing (Maes and Nardi, 1988).

Reasoning about one's own beliefs is a special case of reasoning about any agent's beliefs, with the difference that one's own belief system may be more accessible. The process of reasoning about one's own belief system and its relation to the world (the problem being solved) has been formalised, to some extent, as 'introspection' by Konolige (1986) and others.

The distinction made in computational reflection between the 'object part' (which solves problems and returns information about an external domain) and a 'reflective part' (which solves problems and returns information about the object computation) corresponds to our description of ITSs in terms of two levels, a 'task level' and 'discussion level' (Cumming and Self, 1989b). Similarly, our concerns about the extent to which the two levels may profitably be decoupled and about the kinds of access the discussion level needs to the task level are echoed in the technical literature by the distinction between procedural and declarative reflection and by techniques such as semantic attachment. For ITSs, the significance of any theoretical analysis of reflection lies not in its application to the dynamic modelling of reflective processes but in its consequent clarification of the nature of reflection, the circumstances under which it may occur, and the possible outcomes of reflection.

Belief revision

Most of the work done on belief logics has been concerned with how an agent reasons with beliefs. Since we are interested in designing systems to promote learning, it is equally important to consider the question of belief revision, that is, the situations which may cause a change of beliefs and the conditions under which particular changes occur.

The topic of belief revision is in fact a very active research area in both AI (e.g. Vardi, 1988) and philosophy (e.g. Harman, 1986). Harman, for example, offers some informal principles of belief revision, such as, that one "should make minimal changes to one's view that increase its coherence as much as possible while promising suitable satisfaction of one's ends". Which beliefs one discards depends on the justifications one has for those beliefs and the consequent effect upon the overall consistency of the belief set - questions which have been studied technically under the heading of 'reason maintenance' in AI (Smith and Kelleher, 1988).

Cooperative dialogues

Since, we have argued, the construction of new beliefs occurs through criticism, the role of dialogue is crucial in ITSs. The main technical requirement is a means of handling nested beliefs (that is, beliefs about another agent's beliefs).

For example, Wilks and Ballim (1987) present a formalisation of an interaction which may be re-interpreted as one between a tutor (a doctor) and two students, one of whom believes that 'thalassemia is a genetic disorder', and the other of whom does not. The derivation of an appropriate tutor response to a remark from the first student (such as, "Have you told her children that she has thalassemia?") the significance of which may be lost on the second is considered to depend on reasoning about what one agent believes the second believes the third believes, and to require some kind of default reasoning (again, an active AI research area), such as that Y believes that X believes what Y does provided that it does not contradict anything which Y believes X already believes.

Cohen and Levesque (1987) present a theory of communication derived from a formal theory of rational interaction. Informal principles, such as Grice's, are no longer primitive but follow (or not, as the case may be) from the formal theory - just as we have suggested that informal principles of tutoring may, eventually, follow from a formal theory of learning and teaching. The speech acts of requesting, informing and questioning are not primitive, as in some analyses, but derived from first principles. Such derivations allow us, for example, to distinguish between real questions, rhetorical questions and tutor/student questions.

It is interesting that the formalisations are simpler for a 'collaborative dialogue' than for a 'tutoring dialogue'. The former is more straightforward in both the formal and informal sense. In a collaboration, a question is not asked if the answer is already known by the asker; in a tutorial, the tutor is often devious about his own beliefs. The latter is, as a consequence, that much more difficult to formalise. This theoretical justification for a bias we have expressed previously is reassuring. We have argued elsewhere that there are potential computational and educational benefits from designing ITSs which adopt a genuinely collaborative style rather than a conventional tutorial one (Gilmore and Self, 1988; Self, 1988b; Cumming and Self, 1989a).

Summary

ITS-building is often disparaged as mere engineering. If only it were so: ITS design would be a very different activity it were based on the kind of "beautiful theory" which aeronautical engineers use. Unfortunately, unlike the case of aeronautics at the beginning of this century, the elements of our would-be theory are not already in place. However, theoretical AI of the last ten years has begun to lay some foundations - belief logics, belief revision, computational reflection, reason maintenance, dialogue theories - which may eventually provide the elements of a theory of ITS. It may not be premature to anticipate the application of these theories within ITS research and development. It has often been remarked how the field of ITS forms a good testing ground for theories in psychology and education: it forms an even better testing ground for theories in AI.

Evaluation

It is de rigueur for ITS papers to have a penultimate section on 'evaluation'. The demand for empirical evaluations of ITSs arises from an inappropriate methodology borrowed from another tradition. With the present state-of-the-theory, the need for empirical evaluations may seem inescapable, but the long-term goal of ITS research should be to eliminate empirical evaluations, not to embrace them within the design process.

With aircraft, the maiden flight is a demonstration not an evaluation. All the essential properties of the aircraft have been theoretically determined, rendering the 'test flight' redundant. The role of empiricism has been reduced to "fine-tuning" the theory. Moreover, the empirical tests are carried out not in the real world but in environments specifically designed to test particular aspects of the theory, e.g. wind tunnels to test the aerodynamics. To further ITS design, we need 'ITS wind tunnels' to enable us to carry out focussed studies of components of a theory of ITS.

Of course, we are a long way from an ITS theory from which designs may be formally derived. But there is an intermediate strategy possible for ITS designers which was not available to aircraft designers: ITSs can function as their own simulators. So, instead of deriving the properties of an ITS by some mathematical theory, we derive them by observing the (simulated) ITS in a simulated world.

The basic process is as follows. Assume we have a description of a student model `S`, perhaps as a set of beliefs. Assume there is a set of teaching actions `A` available. Assume we have a theory of learning `L`, which maps a teaching action and a student model into a new student model: `L(A`$_i$`,S)->S'`. (Such a theory is implicit in any ITS design, and for the ITS to function as a simulator it may stay implicit. The process being discussed does not itself depend on the theory of learning being a good one.) We may now run `L` over the set `A` to predict the effect of possible teaching actions.

Assume we have a theory of teaching `T`, which selects a sequence of teaching actions. For example, `T` might select that sequence which maximises `E(S`$_f$`)`, where `S`$_f$ is the student model at the end of the sequence and `E` is some evaluation function. `E` might be defined with respect to some pre-specified objective or, better, in terms of the intrinsic properties of `S`$_f$. We may now run `T` over the possible sequences to determine, for example, how much the optimum sequence improves upon alternatives, such as a random sequence.

For a given `S`, `A`, `L` and `T` this process will always determine the same optimum sequence - it takes no account of any student action while the teaching sequence is being carried out. If `S`, `A`, `L` and `T` were entirely accurate then this would not matter, for the student's actions would be predictable. But of course they cannot be. It would be better, therefore, to use some kind of analysis, perhaps based on Bayes' theorem, to progressively refine `S` in the light of (simulated) student actions. For this reason and to avoid a combinatorial explosion it may be better to determine teaching actions one at a time, not as a sequence, and so to give the opportunistic planning that is regarded as characteristic of human teachers.

This analysis was applied to two learning tasks, vocabulary learning and concept learning, using then current stochastic learning models (Self, 1977a, 1977b). It was shown that, under the assumptions of those models, the optimum sequences gave, respectively, 28% and 172% improvements over a random sequence. These figures do not in any way validate the models but they do give some basis for assessing the benefits of implementation before expensive empirical evaluations. After all, if the designers' own in-built assumptions predict negligible improvements then the case for full-scale experimentation is rather weak!

This "pinch of `S`, `A`, `L`, `T`" theory of ITS is naive and the above evaluations were premature because the computational models were too simple. Nonetheless, the methodology may be worth considering as a means of escaping from the uninformative empiricism of current ITS research. It requires that we develop theories of learning and teaching in which we have sufficient confidence that we would consider the results of such a simulation useful: a worthwhile objective in its own right.

Conclusions

This paper has argued for the transition indicated in Figure 1: from the present situation, where we have repeated iterations through a loop in which informal theories lead to experimental implementations which are subject to empirical study, through an intermediate stage, where semi-formal theories are partially evaluated through simulations before complete implementations are empirically studied, to an 'ultimate methodology' where implementations are derived from formal theories of learning and teaching and then subject to "empirical fine-tuning".

The open questions are:

(1) is ITS design, in principle, sufficiently amenable to theoretical analysis to permit useful derivations?, and

(2) if so, is it premature to begin seeking to develop the field of 'theoretical ITS'?

<insert Figure 1 about here>

R e f e r e n c e s

Anderson, J.R. (1988). The expert module. In M.C. Polson and J.J. Richardson (eds.), *Foundations of Intelligent Tutoring Systems*. Hillsdale, N.J.: Lawrence Erlbaum.

Anderson, J.R., Boyle, C.F., Farrell, R. and Reiser, B.J. (1986). Cognitive principles in the design of computer tutors. In P.E. Morris (ed.), *Modelling Cognition*. London: Wiley.

Bloch, G. and Farrell, R. (1988). Promoting creativity through argumentation. *Proceedings of Intelligent Tutoring Systems Conference*, Montreal.

Brazdil, P. (1988). Meta-level extensions of logic and machine learning. In P. Maes and D. Nardi (eds.), *Meta-Level Architectures and Reflection*. Amsterdam: Elsevier.

Burton, R.R. and Brown, J.S. (1982). An investigation of computer coaching for informal learning activities. In D.H. Sleeman and J.S. Brown (eds.), *Intelligent Tutoring Systems*. New York: Academic Press.

Clancey, W.J. (1986). Qualitative student models. *Annual Review of Computer Science*, 1, 381-450.

Cohen, P.R. and Levesque, H.J. (1987). Speech acts and rationality. *Proceedings of the International Joint Conference on Artificial Intelligence*, Milan.

Collins, A., Brown, J.S. and Newman, S. (1987). Cognitive apprenticsehip: teaching the craft of reading, writing and mathematics. In L.B. Resnick (ed.), *Cognition and Instruction*. Hillsdale, N.J.: Lawrence Erlbaum.

Collins, A. and Stevens, A.L. (1982). Goals and strategies of inquiry teachers. In R. Glaser (ed.), *Advances in Instructional Psychology, Vol. 2*. Hillsdale, N.J.: Lawrence Erlbaum.

Cumming, G. and Self, J.A. (1989a). Learner models in intelligent collaborative educational systems. In P. Goodyear (ed.), *Teaching Knowledge and Intelligent Tutoring*. Palo Alto: Ablex.

Cumming, G. and Self, J.A. (1989b). Intelligent educational systems: identifying and decoupling the conversational levels, in preparation.

Ellis, R. (1986). *Understanding Second Language Acquisition*. Oxford: Oxford University Press.

Fagin, R. and Halpern, J.Y. (1987). Belief, awareness and limited reasoning. *Artificial Intelligence*, 34, 39-76.

Gilmore, D.J. and Self, J.A. (1988). The application of machine learning to intelligent tutoring systems. In J.A. Self (ed.), *Artificial Intelligence and Human Learning*. London: Chapman and Hall.

Harman, G. (1986). *Change in View*. Cambridge, Mass.: MIT Press.

Kamsteeg, P.A. and Bierman, D.J. (1989). Teacher thinking, student thinking. In P. Goodyear (ed.), *Teaching Knowledge and Intelligent Tutoring*. Palo Alto: Ablex.

Kaufmann, H. and Grumbach, A. (1986). Multilog: multiple worlds in logic programming. *Proceedings of 7th European Conference on Artificial Intelligence*, Brighton.

Kearsley, G. (1989). Can artificial intelligence save educational computing? *Journal of Artificial Intelligence in Education*, 1, 5-9.

Konolige, K. (1986). *A Deduction Model of Belief*. London: Pitman.

Lepper, M.R. and Chabay, R.W. (1988). Socializing the intelligent tutor: bringing empathy to computer tutors. In H. Mandl and A. Lesgold (eds.), *Learning Issues for Intelligent Tutoring Systems*. New York: Springer-Verlag.

Levesque, H.J. (1984). A logic of implicit and explicit belief. *Proceedings of the National Conference of the American Association for Artificial Intelligence,* Austin, Texas.

Mandl, H. and Lesgold, A., eds. (1988). *Learning Issues for Intelligent Tutoring Systems*. New York: Springer-Verlag.

Ohlsson, S. (1986). Some principles of intelligent tutoring. *Instructional Science*, 14, 293-326.

Osborne, R. and Freyberg, P. (1985). *Learning in Science*. Auckland: Heinemann.

Polson, M.C. and Richardson, J.J., eds. (1988). *Foundations of Intelligent Tutoring Systems*. Hillsdale, N.J.: Lawrence Erlbaum.

Schoenfeld, A.H. (1985). *Mathematical Problem Solving*. New York: Academic Press.

Self, J.A. (1977a). Concept teaching. *Artificial Intelligence*, 9, 197-221.

Self, J.A. (1977b). A state-space approach for automatic instruction. *Computers and Education*, 1, 199-205.

Self, J.A. (1988a). The use of belief systems for student modelling. *Proceedings of the 1st European Congress on Artificial Intelligence and Training*, Lille.

Self, J.A. (1988b). Bypassing the intractable problem of student modelling, *Proceedings of Intelligent Tutoring Systems Conference*, Montreal.

Shevell, R.S. (1983). *Fundamentals of Flight*. Englewood Cliffs, N.J.: Prentice-Hall.

Smith, B. and Kelleher, G., eds. (1988). *Reason Maintenance Systems and their Applications*. Chichester: Ellis Horwood.

Vardi, M., ed. (1988). *Proceedings of the 2nd International Conference on Theoretical Aspects of Reasoning about Knowledge*. Los Altos; Morgan Kaufmann.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems*. Los Altos: Morgan Kaufmann.

Wilks, Y. and Ballim, A. (1987). Multiple agents and the heuristic ascription of belief. *Proceedings of the International Joint Conference on Artificial Intelligence*, Milan.

Wilks, Y. and Ballim, A. (1988). Shifting the belief engine into higher gear. In T. O'Shea and V. Sgurev (eds.), *Artificial Intelligence: Methodology, Systems, Applications*, Amsterdam: North-Holland.

many iterations

Informal theories	→	Informal principles	→	Experimental systems	→	Empirical evaluation

fewer iterations

Informal theories	→	Informal principles	→	Experimental systems	→	Empirical evaluation

Semi-formal theories	→	Evaluate by simulation

empirical fine-tuning

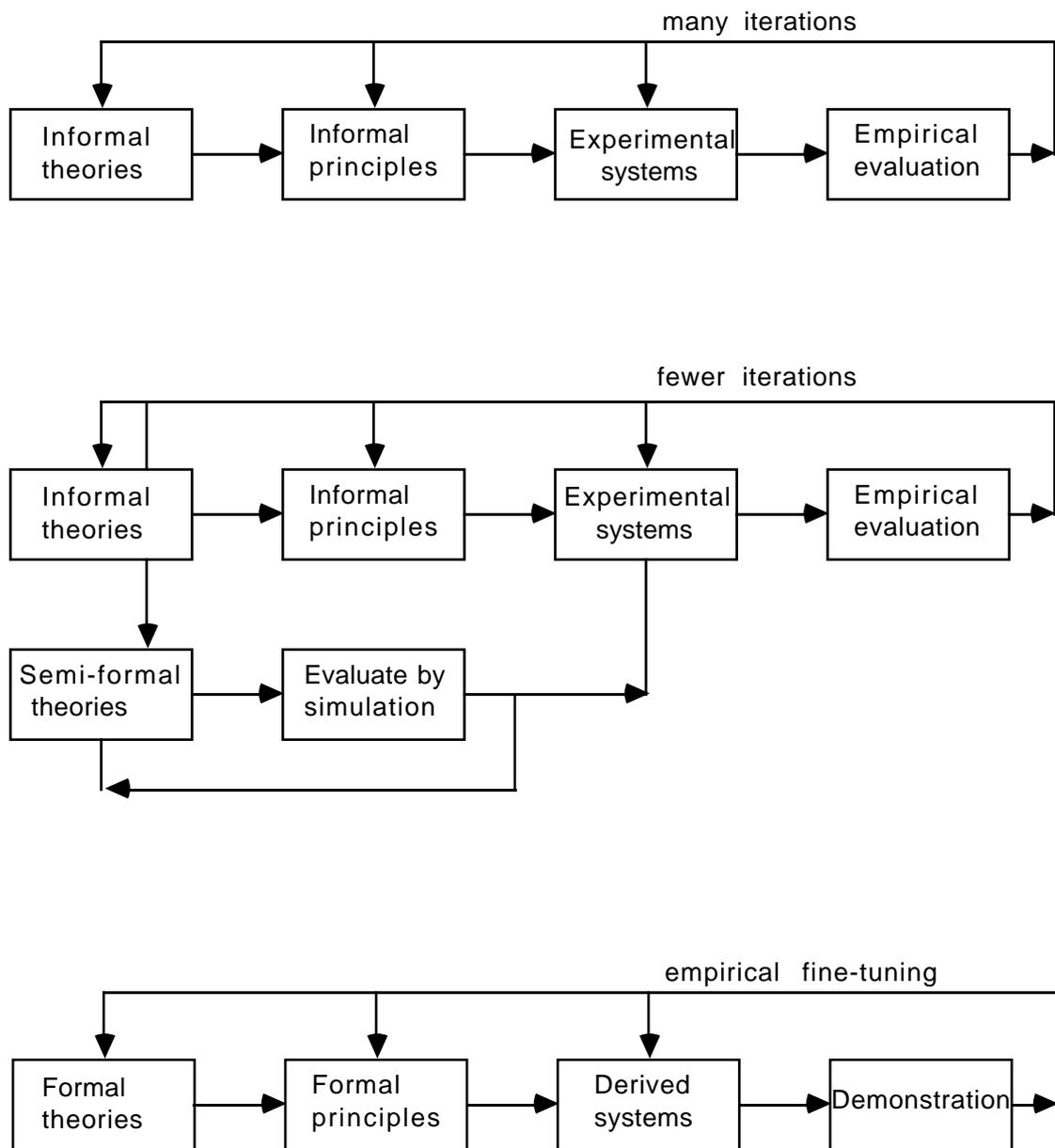| Formal theories	→	Formal principles	→	Derived systems	→	Demonstration

Figure 1.   The progression from informal empiricism to formal demonstration