The Plausibility Problem: Human Teaching Tactics in the 'Hands' of a Machine

Benedict du Boulay, Rosemary Luckin & Teresa del Soldato†
School of Cognitive and Computing Sciences
University of Sussex, UK
Email: bend@cogs.susx.ac.uk
†*Agora Professional Services, Kingston upon Thames, UK*

Abstract: This paper explores the Plausibility Problem identified by Lepper and his colleagues, namely the question as to whether teaching tactics and strategies which work well for expert human teachers can also work for machine teachers. We offer examples of the plausibility problem and argue that the issue arises in part from the impoverished expectations of students about what intelligent learning and teaching systems can achieve and in part from expectations about who should play what role in a learning interaction with a machine. We believe the problem will be reduced, though not necessarily eradicated, if systems are properly integrated within the overall educational context in which they are to be used, though further work is needed to establish this.

1 Introduction

There are three principled methodologies for developing the teaching expertise in AIED systems. First is the empirical observation of human learners and human teachers followed by an encoding of effective examples of the teacher's expertise, typically in the form of rules. An influential early example of this methodology was "Socratic Tutoring" [6]. Socratic Tutoring provides a number of detailed teaching tactics for eliciting from and then confronting a learner with her misconceptions in some domain. A more recent example of the methodology is provided by Lepper et al. [14] who analysed the methods that human tutors use to maintain students in a positive motivational state with respect to their learning. Ohlsson [17] provides an analysis of the great variety of teaching actions in a versatile teacher's repertoire, and berates AIED for implementing only a tiny proportion of this versatility. Bloom [2] compared the effectiveness of a number of general teaching strategies in terms of learning outcomes and provided adaptive systems with the goal of increasing mean learning gains by two standard deviations compared to conventional classroom teaching.

The second methodology starts from a learning theory and derives appropriate teaching tactics and strategies from that theory. Conversation Theory [18] and its reification in various teaching systems is an example of this approach. As with Socratic Tutoring, Conversation Theory is concerned essentially with epistemology rather than with affective aspects of teaching and learning. It is concerned to ensure that the learner constructs a multifaceted understanding of a domain that allows her to describe (to herself or to others) the inter-relationships between concepts. In some ways it is echoed by the "self-explanation" view of effective learning [4]. An example of the second methodology that partially addresses some of the affective issues is Contingent Teaching [23]. Here the idea is to maintain the learner's agency in a learning interaction by providing only sufficient assistance at any point to enable her to make progress on the task. The evaluation of this strategy in the hands of non-teachers who had been deliberately taught it shows that it is effective but sometimes goes against the grain for experienced teachers who often wish to provide more help at various points than the theory permits [24].

The third methodology is an amalgam of the above two. This builds a computational model of the learner or of the learning process and derives a teaching strategy or constraints on teaching behaviour by observing the model's response to different teaching actions. For example, VanLehn et al. [22] compared two strategies for teaching subtraction to a production rule model of a subtraction learner and concluded, on the basis of the amount of processing engaged in by the mode, that the "equal additions" strategy was more effective than the more widely taught "decomposition" strategy. With a similar general methodology VanLehn [21] derived "felicity conditions" on the structure of tutorial examples, for instance that they should only contain one new subprocedure.

In addition to any problems of educational effectiveness in practice, all three of these methodologies are vulnerable to what Lepper et al. [14] call the "Plausibility Problem":

> "Even if the computer could accurately diagnose the student's affective state and even if the computer could respond to that state (in combination with its diagnosis of the learner's cognitive state) exactly as a human tutor would, there remains one final potential difficulty: the plausibility, or perhaps the acceptability, problem. The issue here is whether the same actions and the same statements that human tutors use will have the same effect if delivered instead by a computer, even a computer with a virtually human voice." [14](page 102)

In other words will human teaching tactics and strategies, or tactics derived from learning theories or learning systems work effectively in an intelligent learning or tutoring environment?

It is important to stress that this paper is not arguing for an "anti" Artificial Intelligence in Education (AIED) stance. Indeed, although there have not been very many evaluations of the educational utility of the individual adaptiveness implemented in AIED systems, those that are reported offer grounds for optimism, see e.g. [11]. For brief surveys of AIED evaluations see [9, 20].

One response to the plausibility problem has been an increasing interest in the development of learner companion systems of various kinds, see e.g. [3]. Here the idea is that the human learner has access to a (more or less) experienced fellow learner who can either provide help, act as a learning role model, or through its mistakes act as a reflective device for the human learner. Most of these systems start from the premise that learners need interactions with more than just teachers and that certain sorts of interaction are better conducted with a peer than with a teacher. Of course, computer-based companions raise their own version of the "plausibility problem" compared to their human counterparts.

This paper explores the plausibility issue by reference to two examples from the work of the authors. The first example offers an account of students finding certain human teacher-like behaviours unacceptable when exhibited by a machine. The second example is used to argue that this sense of what is acceptable and what is not acceptable is strongly conditioned by the rather narrow range of machine behaviours that students have actually experienced. Finally it suggests that as intelligent learning environments and teaching systems find their way into the mainstream of education the plausibility problem is likely to diminish, especially if the systems are properly embedded in the educational context in which they are to operate. However even careful embedding may not be enough to undermine deep preconceptions about what a “reasonable” role for a machine teacher should be.

2 Denial of Help by the System

Students are used to being observed by a teacher while they struggle with some problem, for example a mathematics problem, and yet do not receive help. It may be frustrating for them to think that if only the teacher proffered some assistance the intellectual struggle could be terminated sooner, but most will accept that in many circumstances there is value to be had in trying to solve a problem for themselves.

Del Soldato [7, 8] implemented various of the motivational tactics, e.g. derived by [10, 12, 13, 14] in a prototype tutor to teach rudimentary debugging of Prolog programs. Her system had three sets of teaching rules. The first set of (problem domain) rules were concerned with helping the student move through the curriculum of debugging problems, from easy to the more difficult, respecting prerequisite links.

A second set of (motivational) rules was concerned to maintain the students’ sense of confidence and control. Sometimes these two sets of rules would make similar suggestions to the tutoring system about the difficulty of the next problem to be given to the student or about the level of specificity of help that should be provided in response to a request for help from the student. But there were situations where the problem domain and the motivational rules offered opposite advice. In order to reconcile such occasional conflicts of advice within the system, there was a third set of rules whose job it was to try to meld the suggestions from the other two sets of rules into a coherent single strategy — in fact, giving priority to the motivational if there was an irreconcilable clash.

The system (MORE) was evaluated by comparing a version with the motivational rules switched on with one where they were disabled. The version using motivational rules was generally liked by students but two negative reactions from students are noteworthy. One of the rules in the system was designed to prevent the student prematurely abandoning a problem and moving on to the next one, if the system believed that the student was not exhibiting enough “effort”, as measured by the number of actions the student had taken in the partial solution.

> “One subject was showing signs of boredom from the start of the interaction. . . . After a little effort trying to solve a problem, the subject gave up and the tutor encouraged him to continue and offered help. The subject kept working, grumbling that the tutor was not letting him leave. When comparing the two versions of the tutor he recalled precisely this event, complaining that he had not been allowed to quit the interaction.” [7](page 77)

Further rules were concerned with deciding how specific a help message should be delivered in response to a help request — not dissimilar to the rules in Sherlock, see e.g. [15], or indeed to the Contingent Teaching strategy [23]. However in some circumstances the help system refused to offer any help at all in response to a request from the student, in the belief that such students needed to build up their sense of control and that they were becoming too dependent on the system.

> “The subjects who were refused a requested hint, on the contrary, reacted strongly against the tutor’s decision to skip helping (ironically exclaiming “Thank

> you" was a common reaction). Two subjects tried the giving-up option immediately after having had their help requests not satisfied. One case resulted in the desired help delivery (the confidence model value was low), but the other subject, who happened to be very confident and skilled, was offered another problem to solve, and later commented that he was actually seeking help."
>
> "One of the subjects annoyed by having his help request rejected by the tutor commented: "I want to feel I am in control of the machine, and if I ask for help I want the machine to give me help". When asked whether human teachers can skip help, the answer was: "But a human teacher knows when to skip help. I interact with the human teacher but I want to be in control of the machine". It is interesting to note that the subject used to work as a system manager." [7](pages 76–77)

In both these cases the student was surprised that the system behaved in the way that it did — not we believe because the system's response was thought to be educationally unwarranted, but because it was "merely" a machine and it was not for it, *as a machine*, to frustrate the human learner's wishes.

3 Refusal of Help by the Users

Learner's expectations are an important factor of the plausibility problem. Increasingly learners are exposed to computers in their learning and in other aspects of their lives. They absorb the cultural computation conventions and facilities for giving help. These build up expectations of the degree of focussed assistance that they might reasonably expect.

In the second example the plausibility problem may be responsible for results which confounded expectations. There are a number of differences between this system and that of del Soldato, described above. It was aimed at school children, specifically designed to be similar to other educational systems they had used and was evaluated in the children's everyday class. It also explored a topic — simple ecology — that the children were learning at school and, in the versions that decided how helpful to be, was designed to ensure that the child succeeded as far as possible, even if this meant that the system did most of the work.

Three versions of a tutorial assistant which aimed to help learners aged 10 - 11 years explore food webs and chains were implemented within a simulated microworld called the Ecolab [16]. The system was developed to explore the way in which Vygotsky's Zone of Proximal Development might be used to inform software design. The child can add different organisms to her simulated Ecolab world and the complexity of the feeding relationships and the abstractness of the terminology presented to the learner can be varied. The simulated Ecolab world can be viewed differently, for example in the style of a food web diagram, as a bar chart of each organism's energy level or as a picture of the organisms in their simulated habitat. The activities the learner was required to complete could be "differentiated" (i.e. made easier) if necessary and different levels (i.e. qualities) of help were available.

One version of the system: VIS maintained a sophisticated learner model and took control of almost all decisions for the learner. It selected the nature and content of the activity, the level of complexity, level of terminology abstraction, differentiation of the activity and the level of help. The only option left within the learner's control was the choice of which view to use to look at her Ecolab. A second version of the assistant: WIS, offered learners suggestions about activities and differentiation levels. They were offered help, the level of which was decided on a contingently calculated basis [23]. They could choose to reject the help offered or select the "more help" option. The third system variation was called NIS. It offered 2 levels of help to learners as they tried to complete a particular task. The first level consisted of feedback and an offer of further help. The second level which was made available if the child accepted this offer involved the assisting computer completing the task in which the child was currently embroiled. Of the three systems NIS offered the smallest number of different levels of help and allowed the greatest freedom of choice to the child. She could select what she wanted to learn about, what sort of activity she wanted to try, how difficult she wanted it to be and then accept help if she wanted it. The choices were completely up to the individual child, with not even a suggestion of what might be tried being offered by the system.

Three groups of 10 children (matched for ability) worked with the three systems. Outcomes were evaluated both through pre/post-test scores on a test of understanding of various aspects of food webs and chains, and via an analysis of what activities the children engaged in and how much help they sought and received. Pre/post-test comparisons showed that VIS produced greater learning gains than WIS and NIS, see [16, 19] for details. Our focus here is not on the learning gains but on the help seeking behaviour of the students.

3.1 Help Seeking

It is clear from the the records logged by the systems of each child's interactions none of the NIS users accepted the option of seeking more help when offered feedback. There is a clear and typical pattern within the interactions of NIS users: actions are attempted, feedback is given with the offer of help, help is not accepted. The action is re-attempted and once completed successfully it is repeated, interspersed with view changes and further organism additions at differing rates of frequency. Only one of the NIS users asked for a differentiated activity and only two attempted to interact at anything other than the simplest level of complexity or terminology abstraction. The child who tried the differentiated activities chose the highest level of differentiation and once the activities were done he returned to the typical NIS pattern. The help seeking or lack of it is particularly marked in the two children who opted to try the most advanced level of interaction. Both made errors in their initial attempts at completing the food web building action selected, but neither opted to take more help when offered. Few activities were attempted and those that were chosen were accessed with the lowest level of differentiation. The same food web building activity was repeated in both sessions of computer use and in both sessions errors were made. The presence of these errors and the apparent desire to tackle more complex concepts would suggest that the children were willing to move beyond what they already understood. However, the lack of collaborative support restricted their opportunities for success and their progress was limited. What could have been a challenging interaction became a repetitive experience of limited scope.

Unlike the NIS users all the WIS users accepted help above the basic level and the majority used help of the highest level and then remained at this level. A typical WIS approach would be to try an action, take as much help as needed to succeed with this action and then repeat it before trying another different action. Activities were requested with differentiation. In the majority of cases this differentiation was at the highest level. Without question the WIS users were more willing to attempt actions with which they were going to need help. There were members of this group who progressed through the curriculum both in terms of complexity and terminology abstraction. This is a direct contrast to the NIS user group.

The clear difference between one group's willingness to use help over and above simple feedback (WIS) and the other group's complete lack of help seeking is interesting. The help instances for the NIS users were either simple feedback or a demonstration of the particular action being attempted: equivalent to the highest level of help in WIS or VIS. All but one of the NIS users made mistakes and were given feedback, but none of them accepted the offer of further help. It is difficult to explain this startling lack of help seeking behaviour and any attempts are clearly speculative.

4 Educational Context

The only difference between the WIS and NIS system with regard to differentiation or the presentation of help is in the way that WIS suggests that the user try a particular level of differentiation for an activity or ask for help. This policy of offering suggestions was not universally successful. WIS users received suggestions about which activities they should try. These were however accepted less often than the suggestions about the differentiation of an activity. If a suggestion was enough to allow the child to accept an easier activity then it seems reasonable to consider the possibility that without the suggestions, the NIS users viewed choosing a more difficult activity as being somehow better and therefore what they should be attempting.

As part of the design of the experiment, note was taken of the computer programs the children had experienced previously. One tentative explanation of the different behaviours is that children did not believe that either asking for more help or for an easier activity would be successful. The WIS users received suggestions and once the higher levels of help were experienced they were taken up and used prolifically. In this sense the WIS system demonstrated its plausibility as a useful source of assistance in a way that the children never gave the NIS system a chance to show.

A further factor which is consistent with this help seeking behaviour is found in the observation that none of the children accessed the system help menu or system help buttons. These were available to explain the purpose of the various interface buttons and the way that action command dialogues could be completed. The children had all used a demo of the system, which allowed them to determine the nature of the interface and none reported problems at the post-test interview. However, when observing the children using the system it was clear that there were occasions when they were unsure about a button or a box and yet they did not use the help button provided. This may well be an interface issue which needs attention in any further implementations of VIS. However, it may also be part of the same plausibility problem.

There is another facet to the Plausibility Problem, besides the violation of expectation about what a machine *may* do (MORE) or what a machine *can* do (Ecolab). This is related to the way that the intelligent system is used within the overall educational context. The following example from Anderson's work illustrates the issue.

One of Anderson's most recent evaluations concerns a system designed to be used in Pittsburgh High Schools [11]. The Practical Algebra Tutor (PAT) is designed to teach a novel applications-orientated mathematics curriculum (PUMP — Pittsburgh Urban Mathematics Project) through a series of realistic problems. The system provides support for problem-solving and for the use of a number of tools such as a spreadsheet, grapher and symbolic calculator.

Of special note here, apart from the positive evaluation of the system, is the way that attention was paid to the use of the Tutor within the classroom. The system was used not on a one-to-one basis but by teams of students who were also expected to carry out activities related to the use of PAT, but not involving PAT, such as making presentations to their peers. In this situation the educational interactions involved the system almost as a third party, or even as a "conversation piece", so students were not so starkly faced with the problem of dealing with the machine as the *sole* provider or withholder of help.

5 Conclusions

We have argued that Lepper and his colleagues were correct to raise the issue of the Plausibility Problem and that in our work we have encountered examples of it. However we believe that one aspect of the plausibility problem derives from students' expectations of what intelligent teaching systems can actually achieve. Such a conclusion must necessarily be very tentative as few examples of such systems have found their way to the classroom and so most students' beliefs about the degree of insight and adaptability will be based on some mixture of computer games and CAL programs as well as on science-fiction. Such mixed models of what computer teachers might and can be like can only be confusing.

Once students have experienced a number of adaptive systems, their surprise that such systems will exercise a similar degree of agency to human teachers should diminish, so long as the use of those systems in the classroom is properly thought through. Even in this situation there may still be some resentment if the machine is seen to be usurping its authority. Further work is needed to establish whether this will be the case. But there is still a further issue to be wary of and that concerns students expectations of themselves when working with an intelligent system.

Barnard & Sandberg [1] built a learning environment for the domain of tides to help students understand why, where and how tides occur in relation to the movement of the earth, moon and sun. Despite encouraging their students to engage in self-explanation so as to reveal areas of the tidal process which they did not understand, students were loathe to do this and in general they had little insight into how partial their knowledge of these processes actually was. It may be that this problem can be reduced by providing a more effective interface, rather than encouragement, to make reflective insight more likely [5]. They describe another facet of the Plausibility Problem, namely that strategies that can be adapted by a human teacher to provoke reflection and self-explanation may not work when the teacher is known to be a machine. In other words, is the very methodology so carefully nurtured by AI-ED systems to track the learning of the student effectively a message to students that they do not need to do this for themselves? To the extent that the system can track students at all, the student will reasonably believe in the high quality, patient record-keeping of the machine, *as a machine*. While the human teacher may or may not build a detailed model of a student that she is interacting with, it will be clear to the student, especially as one among a group of students, that a human teacher really will not be able to track their work in detail, and if anyone is to do it it will have to be the student herself.

References

[1] Y. F. Barnard and J. A. C. Sandberg. Self-explanations, do we get them from our students? In P. Brna, A. Paiva, and J. Self, editors, *Euroaied: European Conference on Artificial Intelligence in Education*, pages 115–121, Lisbon, 1996. Edicoes Colibri.

[2] B. S. Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6):4–16, 1984.

[3] T.-W. Chan. Learning companion systems, social learning systems, and the global learning club. *Journal of Artificial Intelligence in Education*, 7(2):125–159, 1996.

[4] M. Chi, M. Bassok, M. Lewis, P. Reimann, and R. Glaser. Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13:145–182, 1989.

[5] A. Collins and J. S. Brown. The computer as a tool for learning through reflection. In H. Mandl and A. Lesgold, editors, *Learning Issues for Intelligent Tutoring Systems*, pages 1–18. Springer-Verlag, New York, 1988.

[6] A. Collins, E. H. Warnock, N. Aiello, and M. L. Miller. Reasoning from incomplete knowledge. In D. G. Bobrow and A. Collins, editors, *Representation and Understanding*, pages 383–415. Academic Press, New York, 1975.

[7] T. del Soldato. Motivation in tutoring systems. Technical Report CSRP 303, School of Cognitive and Computing Sciences, University of Sussex, 1994.

[8] T. del Soldato and B. du Boulay. Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education*, 6(4):337–378, 1995.

[9] B. du Boulay. What does the AI in AIED buy? In *Colloquium on Artificial Intelligence in Educational Software*. IEE Digest No: 98/313, 1998.

[10] J. M. Keller. Motivational design of instruction. In C. M. Reigeluth, editor, *Instructional-design Theories and Models: An Overview of their Current Status*. Lawrence Erlbaum, 1983.

[11] K. R. Koedinger, J. R. Anderson, W. H. Hadley, and M. A. Mark. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8(1):30–43, 1997.

[12] M. R. Lepper. Motivational considerations in the study of instruction. *Cognition and Instruction*, 5(4):289–309, 1988.

[13] M. R. Lepper and R. Chabay. Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl and A. Lesgold, editors, *Learning Issues for Intelligent Tutoring Systems*, pages 242–257. Springer-Verlag, New York, 1988.

[14] M. R. Lepper, M. Woolverton, D. L. Mumme, and J.-L. Gurtner. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie and S. J. Derry, editors, *Computers as Cognitive Tools*, pages 75–105. Lawrence Erlbaum, Hillsdale, New Jersey, 1993.

[15] A. Lesgold, S. Lajoie, M. Bunzo, and G. Eggan. Sherlock: A coached practice environment for an electronics troubleshooting job. In J. H. Larkin and R. W. Chabay, editors, *Computer-Assisted Instruction and Intelligent Tutoring Systems*, pages 289–317. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1992.

[16] R. Luckin. 'ECOLAB': Explorations in the Zone of Proximal Development. Technical Report CSRP 386, School of Cognitive and Computing Sciences, University of Sussex, 1998.

[17] S. Ohlsson. Some principles of intelligent tutoring. In R. W. Lawler and M. Yazdani, editors, *Learning Environments and Tutoring Systems: Learning Environments and Tutoring Systems*, volume 1, pages 203–237. Ablex Publishing, Norwood, New Jersey, 1987.

[18] G. Pask and B. Scott. CASTE: A system for exhibiting learning strategies and regulating uncertainties. *International Journal of Man-Machine Studies*, 5:17–52, 1975.

[19] S. Puntambekar and B. du Boulay. Design and development of MIST — a system to help students develop metacognition. *Journal of Educational Computing Research*, 16(1):1–35, 1997. Reprinted as Puntambekar and du Boulay (1999).

[20] V. J. Shute. Rose garden promises of intelligent tutoring systems: Blossom or thorn? In *Space Operations, Applications and Research (SOAR) Symposium*, Albuquerque, New Mexico, 1990.

[21] K. VanLehn. Learning one subprocedure per lesson. *Artificial Intelligence*, 31(1):1–40, 1987.

[22] K. VanLehn, S. Ohlsson, and R. Nason. Applications of simulated students. *Journal of Artificial Intelligence in Education*, 5(2):135–175, 1994.

[23] D. J. Wood and D. J. Middleton. A study of assisted problem solving. *British Journal of Psychology*, 66:181–191, 1975.

[24] D. J. Wood, H. A. Wood, and D. Middleton. An experimental evaluation of four face-to-face teaching strategies. *International Journal of Behavioural Development*, 1:131–147, 1978.