Motivationally Intelligent Educational Systems: the contribution of the Human Centred Technology Research Group

Benedict du Boulay

Human Centred Technology Research Group
University of Sussex
Brighton, BN1 9QJ UK

This paper describes educational systems built by members of the Human-Centred Technology Research Group at the University of Sussex that address different aspects of motivation. These systems have been described elsewhere, so this paper is essentially a drawing together of existing work. In particular, we have recently set out a view of motivation, distinguishing different kinds of intelligent educational system including the cognitive, the metacognitive, the affective and the meta-affective, as well as the motivational, and indeed the meta-motivational. The body of the paper gives brief descriptions of eight systems that have operated across a range of domains but always reasoning at a level beyond the cognitive. In each case the evaluation of the system is described, and not all functioned as their designers had expected. The paper concludes with questions and lines of future research.

INTRODUCTION

Most intelligent tutoring systems and intelligent learning environments have focused on the issue of developing the learner’s knowledge and skill by adjusting the educational experience so as to maximise learning. Personalised adjustments in terms of advice about which problem to tackle next, specific and focused help when errors are made and information about the degree to which learning goals are being achieved are three tactics commonly deployed. By contrast, over the last 15 years or so there has been increasing interest in developing systems that take account of the metacognitive, the affective and the motivational aspects of learning in addition to the cognitive. For example, Aleven and his colleagues, built additions on top of a geometry tutor to help learners make better use of the inbuilt help mechanism. While the main goal was still to foster better learning of geometry, the route was via considerations of the learner’s metacognitive insight into their use and misuse of help (Aleven, McLaren, Roll, & Koedinger, 2006). At the affective level Graesser, D’Mello and their colleagues have instrumented AutoTutor to gather facial, physiological and other data from learners that betray their affective state and then deploy tactics that steer the learner away from individual or sequences of affective states that are not conducive to effective learning (see e.g., D’Mello, Graesser, & Picard, 2007). At the motivational level Song and Keller (2001) built a system to maintain the student’s level of motivation by deploying dynamically various components of Keller’s (1987) ARCS model of motivation (such as the Relevance of the material).

Hierarchy of intelligent systems

In a recent paper (du Boulay, et al., 2010), we identified a partial hierarchy of intelligent educational systems ranging from Cognitively Intelligent and Affectively
Intelligent systems at the base, up to Caring systems (Self, 1999) at the apex, see Table I.

**Table I: Partial hierarchy of intelligent educational systems (adapted from du Boulay, et al., 2010)**

<table>
<thead>
<tr>
<th>Kind of system</th>
<th>Pedagogic Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caring systems</td>
<td>The growth of the learner as a person</td>
</tr>
<tr>
<td>Meta-motivationally intelligent</td>
<td>Increasing the learner’s meta-motivational capability e.g. her insight and regulation of her Motivation</td>
</tr>
<tr>
<td>Motivationally intelligent</td>
<td>Increasing the learner’s desire to learn, e.g. her willingness to expend effort on the learning process.</td>
</tr>
<tr>
<td>Metacognitively intelligent</td>
<td>Increasing the learner’s metacognitive capability, e.g. insight into what she understands and can do, and her ability to regulate her learning process effectively.</td>
</tr>
<tr>
<td>Cognitively intelligent</td>
<td>Increasing the learner’s knowledge and skill</td>
</tr>
<tr>
<td>Meta-affectively intelligent</td>
<td>Increasing the learner’s meta-affective capability, e.g. her insight and regulation of her feelings as a learner.</td>
</tr>
<tr>
<td>Affectively intelligent</td>
<td>Increasing the learner’s overall sense of well-being</td>
</tr>
</tbody>
</table>

Table I identifies the varying pedagogic foci of these different kinds of system and indicates motivationally intelligent systems as comprising two intertwined aspects of learning: the cognitive and the meta-cognitive together with the affective and the meta-affective. Here meta-affective stands in the same relation to affective as meta-cognitive does cognitive. It is about the learners knowledge of, and ability to regulate, the feelings that are associated with learning, such as elation when things go
well, frustration when things do not go as hoped, or anxiety when one is unsure of one's ability to succeed. So a motivationally intelligent system needs to reason about the cognitive and affective aspects of learning, while also taking account of the learner's metacognitive and meta-affective sophistication.

A motivationally intelligent tutoring system will have a sense not only of what the learners know and can already do, but also of why they are engaged in the new learning and of the educational goals that they hold, as well as of how they view themselves as learners. A motivationally intelligent system will thus be able to distinguish, for example, a committed, hard-working but unselfconfident learner from one who is bright and self-confident but not all that interested in the material. And in distinguishing between these different kinds of learner it will be able to act on those distinctions to try to ensure that each learner comes out of the learning experience having gained from it.

**Pedagogic strategy**

We also classified the data that the different kinds of system use to undertake their pedagogical reasoning about the likely consequences of a pedagogical tactic (such as setting harder problem) as well as the reasoning needed to identify what tactics should be deployed in order to help the learner attain a particular desired state (du Boulay, et al., 2010). For example, If the learner is bored and the system would wish that learner was engaged, then what are the options available to the system to shift the learner from the first state to the second. The data, see Table II, were classified into four broad categories. We noted that the system might react to data in one category (e.g. physiological: posture indicative of boredom) with a pedagogic tactic operating within a different category (e.g. context: getting the student to work with another student).

In respect of pedagogical reasoning, in a different paper (du Boulay, in press), we argued that simply knowing that the student was in a particular state (frustrated, say) was not enough to immediately identify what might be a good pedagogic tactic to deploy. It is important to know what the cause of the undesirable state is in order to try to counter it. For example, a student may be frustrated because he or she is not making progress on solving a problem. But they also may be frustrated because they would rather be doing something else altogether and their frustration stems from the misuse of their time rather than their lack of progress. In this analysis we borrowed from the work of Pintrich (2003) to identify two broad areas (i.e. values and expectancies) within which the learner's reasoning about their learning experience may give rise to motivational states with associated negative feelings such as anxiety, frustration and boredom. So frustration arising from a desire to do something else is largely a values issue, whereas frustration arising from a failure to progress is more an expectancies issue.

This paper does not seek to re-describe theoretical work already published (du Boulay, in press; du Boulay, et al., 2010). Rather it offers a more personal account of our activity in this general area by describing systems that we have been involved in developing and evaluating.
Table II: Categories of diagnostic input and feedback reaction (adapted from Benedict du Boulay, et al., 2010)

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>DIAGNOSTIC INPUTS</th>
<th>FEEDBACK REACTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTEXT</td>
<td>The spatial, physical, social and temporal milieu within which the student is learning.</td>
<td>Change of location, use of available peers and others, Change of ambience e.g. light levels.</td>
</tr>
<tr>
<td>META-CONTEXT</td>
<td>What the learner knows and can articulate and regulate about the context in which she is learning.</td>
<td>Conversations about the nature of the context and how different contexts affect learning.</td>
</tr>
<tr>
<td>MOTIVATION</td>
<td>What drives the learner to learn or not to learn, what they think they are going to achieve, why they are learning at all and the social and temporal milieu within which the learner is learning.</td>
<td>Flaggng of goals and outcomes, intrinsic and extrinsic rewards. Use of available peers and others, change of educational milieu.</td>
</tr>
<tr>
<td>META-MOTIVATION</td>
<td>What the learner knows and can articulate and regulate about her motivation.</td>
<td>Conversations about the nature of the development of motivation.</td>
</tr>
<tr>
<td>AFFECTIVE</td>
<td>How the learner feels about the learning activity.</td>
<td>Praise, encouragement, criticism, politeness, teacher’s demeanour.</td>
</tr>
<tr>
<td>META-AFFECTIVE</td>
<td>What the learner knows, can articulate and regulate about her actual and expected feelings.</td>
<td>Conversations about expectations of feelings, state of motivation, engagement.</td>
</tr>
<tr>
<td>PHYSIOLOGICAL</td>
<td>Bodily aspects such as heart and breathing rate, skin conductance, facial expression, body language and posture.</td>
<td>Breathing exercises, mantras, pauses. Changes in the physical context e.g. warmth, light, ambient noise.</td>
</tr>
<tr>
<td>META-PHYSIOLOGICAL</td>
<td>What the learner knows and can articulate and regulate about her physiological responses.</td>
<td>Conversations about physiological response.</td>
</tr>
<tr>
<td>COGNITIVE</td>
<td>Knowledge and skills of the learner.</td>
<td>Activity choice, pace or order of work, provision of help.</td>
</tr>
<tr>
<td>META-COGNITIVE</td>
<td>What the learner knows, can articulate and regulate about her knowledge and skills.</td>
<td>Conversation about performance, degree of challenge, use of help, narrative framework.</td>
</tr>
</tbody>
</table>

**Evaluation**

Evaluating the effectiveness of systems that operate at the cognitive and affective levels is fairly straightforward. In the cognitive case we look for evidence that the learners engaged in the kinds of behaviour that we would wish – problems-solving in a particular manner, for example – and that there were knowledge and/or skill gains in the chosen domain. In the affective case we look for evidence, in addition to that for the cognitive case, that the affective states and trajectories traversed by the learners...
were broadly positive (while allowing for the fact that learning necessarily involves some transitory negative states e.g. frustration).

Evaluating metacognitively intelligent systems is more tricky. We hope to find evidence that the learners have engaged in the desired reflective and regulatory behaviours in their own right (e.g. choosing a study strategy), and hope that they may also have learned something at the cognitive level too. Ideally we might also seek some evidence of pre/post gains and/or transfer at the metacognitive level, though this is hard to measure and hard to achieve. Similar considerations apply in the evaluation of meta-affectively intelligent systems.

Evaluating motivationally intelligent systems is also hard and would involve finding evidence that the learners had been better at overcoming obstacles (including lack of interest in the domain), had shown more perseverance, or had worked harder in addition to gains made at the cognitive, metacognitive and affective levels. We might also hope to find evidence of some more general pre/post gains and/or transfer at the motivational level, though this is also hard to measure and hard to achieve.

This paper is organised into four sections. The next three deal with metacognitively, affectively and motivationally intelligent systems. Each section describes one or more systems designed and evaluated at the University of Sussex. The final section draws some conclusions from this work and suggests future directions.

**METACOGNITIVELY INTELLIGENT**

This section describes three systems that have operated at the metacognitive level. The first, MIST, aimed to teach the skill of academic learning from texts, for example distinguishing skim reading from careful study. The system did not have any specific domain of academic expertise but it did know about how to study from texts and about how different kinds of reading had different kinds of consequences for learning.

The second system, HabiPro, was concerned to teach programming concepts at the domain level, but at the metacognitive level it aimed to foster effecting cooperative problem-solving. This system enabled several students working remotely to jointly work on a programming problem using a shared space and a chat-type interface. Unknown to the human students, one of the “students” was in fact a system and its contributions to the public chat were designed to help the other students both in programming but also in managing their interactions more effectively.

The third system, MIRA, was designed to assist students develop reflective awareness when solving simple algebra word problems. In particular it focused on the degree to which the students were able to monitor their own problem-solving knowledge (Tobias & Everson, 2002) and whether they exhibited any biases in so doing; for example, optimistically believing that they understood more than they actually did.

**Tutoring learning from texts**

MIST was designed to help secondary school level students practise, regulate and reflect on the processes involved in learning from texts (Puntambekar & du Boulay, 1997, 1999). In terms of Table I and Table II it operated wholly at the metacognitive level and provided an interface for students both to understand the skill of learning from texts as well as to describe their educational goals, the kind of reading that they
The system knew nothing of the texts that the students wished to read, other than via the descriptions offered by the students; and these descriptions were at a high level so the system was not in any way capable of testing the students on the content of those texts. But it did help students with strategies through which they could test each other and test themselves, e.g. by inviting them to predict the kinds of questions that they might be asked. Crucially the system was designed for use by pairs of students, with the idea that the two learners would collaborate spontaneously over the choices that each was making, as well as having the system prompt the one who seemed to be more effective as some sub-task to help the other.

The educational theory behind the system design was that the interface and the tasks undertaken by the students should make them reflectively aware of processes that would otherwise remain covert (Brown, 1981). This was another reason for designing the system for use by pairs. The system had two parts. One was a hypertext-based browsing system to help students passively learn about learning from texts and different study strategies. The other was an interactive system with which the pair of students could actively plan their separate regimes of reading, could undertake that reading, could reflect on how well that went and employ various tools to assist the understanding process. Note that the system had no knowledge of the content of the readings chosen by the students, and that reading was undertaken “offline”, though with the help of the system operating a timer set to the value chosen by the student.

In planning for the reading, the student used menus to set a goal, allocate study time, decide on an appropriate method and sort out prior knowledge. There were tools for note-taking during and post reading such as tables, spider diagrams, conceptual mapping and also a tool to prompt the student to predict the kinds of question that they might be asked in any future assessment of their understanding of the text. Part of the planning interface to the MIST system is shown in Figure 1.

As the reading was undertaken offline, the system relied on the students being honest about what they had done matched what they planned. The fact that the students worked in pairs in such a situation helped warrant the requirement that they explain or clarify to each other the choices that they had made. The system kept a record of the planned activities and invited the students at the end of a session to reflect on their own and each other’s choices and adjust them if necessary. If the students did not ask for a summary, the system automatically presented one to invite their reflection. Across sessions the system kept a record of the activities undertaken overall and encouraged them to try as many techniques as possible.
The system was partially evaluated in a school with 10 pairs of students aged between 14 and 15 years. The pairs were based on friendship and each pair contained students of the same gender. Overall there were equal numbers of boys and girls. These pairs were divided into two sets of 5. One set (group HI) had scored well on the Learning and Study Skills Inventory (LASSI, Weinstein, Zimmermann, & Palmer, 1988) and on their academic performance and the other (group LO) had not. The evaluation method involved looking at behavioural differences using the system and pre/post changes between the two sets of pairs. There was mixed success with the system.

“Group LO students had a lot of difficulty understanding the planning options . . . [whereas] group HI students used the interface more “actively” – i.e., they used the planning options to elicit prior knowledge, they were more attentive toward features in the text such as diagrams, and they engaged in active self-testing and also tested each other. Students of group LO, on the other hand, used the options more rigidly and provided answers in monosyllables most of the time . . . Students of group HI discussed their learning extensively during the training. On the contrary, group LO students did not converse very much.” (Puntambekar & du Boulay, 1997, page 30, 31)

**Tutoring cooperative working**

As argued in the previous section, getting students to work together can bring many benefits. However things can also go wrong. For example, one student can dominate the other, or they can spend their time concentrating on activities or chatting that have nothing to do with the task in hand. Operating effectively as a learner in a group setting is a metacognitive skill that needs to be mastered just like any other skill. This was one of the tasks tackled by HabiPro (Vizcaíno, 2005; Vizcaíno & du Boulay, 2002). The domain of HabiPro was simple programming and was aimed at supporting a small group of students working remotely but using a system that offered
a common workspace and a chat interface. Students were posed a programming problem and could chat about co-constructing and submitting an answer, see Figure 2.

Figure 2: HabiPro interface showing the problem on the left, the chat interface on the right, and the workspaces at the bottom (adapted from Vizcaino, 2005).

The system had an inbuilt Tutor that set problems, responded to the students’ solutions and offered help. But unknown to the students, one of the “student” participants in the chat was in fact a simulated student. This simulated student adjusted its contributions largely around keeping the cooperation between the human students on track. So for example, if they started to chat about non-programming topics, it intervened suggesting they get on with the problem. Or if one student was very quiet, it invited him or her to contribute, or if another student was dominating the chat it found a gentle way to get the others to contribute more. At the domain level the simulated student might have sudden inspirations if the others were all stuck, or if things seemed to easy, it might ask a provocative question. Table III shows the pedagogic tactics of the simulated student (SS).

The evaluation of HabiPro was based on comparing the experience of two sets of 11 pairs of undergraduate computer science students working over two sessions. In the first session, one set used HabiPro without the simulated student and the other set used the system with the simulated student.
Table III. The pedagogic tactics of the simulated student in HabiPro (adapted from Vizcaino & du Boulay, 2002)

<table>
<thead>
<tr>
<th>Situation</th>
<th>Role</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem-solving</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students do not have enough knowledge so they don't know how to work.</td>
<td>The SS gives hints or explains the exercises.</td>
<td>Proposing clues, or solutions but always with the goal of fostering students' reflection.</td>
</tr>
<tr>
<td>Students always try wrong solutions, (perhaps they are trying to guess the solution)</td>
<td>The SS explains why that solution cannot work. The SS tries to motivate the students (if it occurs that students are bored or tired).</td>
<td>To accustom the students to think about the advantages and disadvantages of a proposal.</td>
</tr>
<tr>
<td>Students have different points of view about the solution, and they propose different or even opposing answers.</td>
<td>The SS helps the students to reflect on the different proposals. The SS encourages the student who proposes the solution to explain it.</td>
<td>To teach respect for different ideas and to think about their advantages or disadvantages. Learning by listening and learning by teaching.</td>
</tr>
<tr>
<td>Students propose correct solutions.</td>
<td>The SS checks that students really understand the solutions and that they did not arrive at it by chance. The SS proposes a wrong solution to create doubt.</td>
<td>Checking gain of knowledge</td>
</tr>
<tr>
<td><strong>Off-topic conversations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students talk about other topics for a long time.</td>
<td>The SS suggests continuing with the problems and asks questions or proposes solutions.</td>
<td>Drawing students' attention back towards the problems.</td>
</tr>
<tr>
<td><strong>Passive students</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student with deficient knowledge.</td>
<td>The SS asks other students to explain the exercises. It can check if a gain of knowledge has arisen. The SS investigates what topics the student demonstrates more knowledge about and invites her/him to explain these topics. The SS checks if it is appropriate to lower the level of difficulty of the exercises.</td>
<td>Learning by teaching. Learning from an explanation. Adaptation of the level of difficulty of the exercises to students' knowledge level.</td>
</tr>
<tr>
<td>Student with adequate knowledge.</td>
<td>The SS motivates and invites the passive student to intervene. The SS suggest turn taking protocols.</td>
<td>To motivate students to participate. Reinforce self-confidence.</td>
</tr>
<tr>
<td>Hyperactive student.</td>
<td>The SS moderates the hyperactive participation and encourages the rest of the students to participate. The SS suggests using turn protocols.</td>
<td>To guarantee equitable participation.</td>
</tr>
</tbody>
</table>
In the second session the conditions were reversed, so that the pairs who had worked with the simulated student then worked without it, and vice versa. The students were allocated into their pairs randomly and each member of a pair worked remotely from the other member. Those pairs who worked with a simulated student were not told this: as far as they were concerned the third participant was another human student. The evaluation looked at the nature of, and the behavioural consequences of, the simulated student’s interventions as well as at the human students’ reactions to their experience. Data about the latter was collected via questionnaires at the end of each of the two sessions.

At the domain level, the students solved more problems when the simulated student was present and intervening when things seemed to be going wrong than when the simulated student was absent. In terms of metacognitive intervention, HabiPro detected all eight instances of passive student behaviour and in only one case was this behaviour repeated after an intervention by the simulated student. In only one case did the simulated student mistake non-passive behaviour for passive behaviour and intervened unnecessarily. This was because the student’s comments were briefer than the threshold set for passivity, though examining the logs showed that they were engaged and not in fact passive. HabiPro detected 12 out of the 14 instances of off-topic conversations and was successful in getting the students back on track in 11 of these instances. In one case HabiPro intervened mistakenly with respect to an apparent off-topic conversation because its dictionary was inadequate.

**Tutoring knowledge monitoring**

The MIRA system was based on a “metacognition reflection model, named the Reflection Assistant (RA), that focuses on the following metacognitive skills: (1) problem understanding and knowledge monitoring, (2) selection of metacognitive strategies, and (3) evaluation of the the learning experience.” (abstract, Gama, 2004).

Like HabiPro and unlike MIST (both described earlier), MIRA taught metacognitive skills in the context of a particular domain which it also tutored. So it operated both at the cognitive and metacognitive levels and the domain was solving algebra word problems (Gama, 2004). In order to minimise the extra cognitive load on the students of both thinking about problem-solving as well as doing problem-solving, MIRA concentrated pre- and post-task reflection.

“The pre-task reflection and familiarization assistant aims at preparing the student for the problem-solving activity, promoting reflection on knowledge monitoring, assessment of the understanding of the problem to be attempted, and awareness of useful metacognitive strategies. The post-task reflection assistant presents activities related to the evaluation of problem solving and takes place just after the student finishes a problem”. (Page 672, Gama, 2004)

The system employed five metacognitive reflective tactics:

1. It invited the student to view a graphical and textual account of their proclaimed assessment of how well they understood and thought they would be able to tackle previous problems against their actual performance (see Figure 3).
2. It invited the student to view any tendency in past problems towards optimism or pessimism in the kinds of knowledge monitoring judgements that they had been making.
3. It invited the student to make assessment of her understanding of the current problem and her ability to solve it.
4. It invited the student to choose which metacognitive strategies to apply in the current problem e.g. monitoring progress and controlling errors or revisiting solution paths.
5. It invited the student to review progress on the most recent problem including use of resources, outcome, time spent and so on.

![Figure 3](image)

**Figure 3**: MIRA: Reflection on predicted and demonstrated knowledge skills (adapted from Gama, 2004)

The evaluation of MIRA was conducted over three one-hour sessions. There were two groups of undergraduate students (N=25). One group (experimental) used MIRA as described above and the other group (control) worked on the same set of problems using a version of MIRA having the reflective activities removed. As expected, the extra time needed to undertake the reflective activities meant that the experimental group solved fewer problems than the control group, but performed better in terms of the number of correct and nearly-correct answers produced. The same Knowledge Monitoring Accuracy (KMA) test (Tobias & Everson, 2002) was used pre and post the experiment, but there was not a significant difference in the change in KMA score between the two groups.
AFFECTIVELY INTELLIGENT

There is increasing interest in building systems that attempt to gauge the affective states of students using a variety of different sources of data: from sensors attached to the students and to their chairs or mice, from cameras pointed at their faces and from an analysis of their interactions with the system. For a contemporary example of such an approach, in this case to try to understand the pre-cursors to students experiencing the surge of pleasure and satisfaction that comes with coming to understand a topic or solving a problem, see Muldner, Burleson & VanLehn (2010). It is, however, just as important to think about the “output” aspect of systems: what should they do to try to change the affective state of the student for the better, if that state has deteriorated.

On the output side, Yussof has designed and evaluated an Affective Tutoring System to teach programming that involved the system in reasoning about the learner’s general sense of well-being and prescribing a regime of neck and shoulder exercises as part of the interaction (Yussof & du Boulay, 2009, 2010). The system was designed to teach data structures at university level. It set problems and provided access to help materials and indicated which parts of the student’s answer needed further work, just like many other cognitively intelligent systems. It also invited students to say how their sense of well-being had changed (if at all) after each topic lesson and slightly adjusted the affective tone of its feedback as a result of their answers as well as on the basis of its own reasoning about their likely sense of well-being.
Figure 4: The Affective Tutoring System interface (adapted from Yusoff, 2010)

Two versions of the system were compared as part of the evaluation with 68 undergraduate computer science students. One version provided feedback about data structures (i.e. domain dependant, problem-focused feedback – see Figure 4). The other version did the same, but also offered feedback and a relaxation activity that was domain independent and emotion-focused (Lazarus, 1991). The group that undertook relaxation exercises in addition to getting problem-focused feedback learned more as measured by pre/post comparisons, particularly those of lower ability with the cohort. The evidence about changes in well-being was equivocal. The two groups showed the same pre/post decline in overall well-being (as measured by the PANAS questionnaire: Watson, Clark, & Auke, 1988) but within each lesson, students who had undertaken the relaxation exercises said that they felt better for them.

MOTIVATIONALLY INTELLIGENT
This section briefly describes three motivationally intelligent systems that have been developed at Sussex. They each exploit “thermostat” models. In terms of the underlying model of the student’s motivational state they each use subsets of the variables identified by Keller in his ARCS model (Keller, 1987). These variables are
updated by both student actions and by system reactions. When any of these variables reach a critical threshold value it triggers system reactions.

**Motivational planning**
MORE was one of the first motivationally intelligent systems to be developed (del Soldato, 1994; del Soldato & du Boulay, 1995). The domain was Prolog debugging and the system was designed for one-to-one undergraduate use. The basic system offered a sequence of increasingly difficult Prolog debugging problems to the learner each involving a range of debugging issues. Depending on the individual student’s performance the system could move the student faster or slower through the curriculum.

Table IV: The MORE system (adapted from del Soldato & du Boulay, 1995)

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>INPUTS</th>
<th>REACTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>COGNITIVE</td>
<td>Performance, effort</td>
<td>Activity choice, provision of help</td>
</tr>
<tr>
<td>META-COGNITIVE</td>
<td>Difficulty of work chosen, use of available help</td>
<td>Comments about performance, degree of challenge, (ab)use of help.</td>
</tr>
<tr>
<td>AFFECTIVE</td>
<td></td>
<td>Praise and encouragement or criticism about performance or effort</td>
</tr>
<tr>
<td>META-AFFECTIVE</td>
<td>Comments from student about willingness to tackle difficult problems</td>
<td></td>
</tr>
</tbody>
</table>

Alongside the domain planner there was a motivational system that used a variety of inputs from the student to tailor its motivational reactions to the student, see Table IV. The motivational pedagogical theory was expressed in terms of sets of production rules for key events in the ongoing interaction such as failing to solve a problem. The condition parts of these rules were matched to the current values of the variables. These rules were used in a data-driven manner i.e. IF confidence high THEN react as follows. Such a rule might change the way that help was offered or provided in future. To that extent the system could execute its explicit motivational pedagogic theory but not reason about it. For the latter to happen the system would have needed, for example, to be able to work backwards from a desired state of the student to find, coordinate and then execute those pedagogic rules that were predicted to move the student from the state that the system thought she was currently in to a desired, better future state. In other words MORE’s theory was essentially reactive and all about how the system should behave at particular points in the interaction. A more sophisticated theory might have tried to shape the very nature of the interaction dynamically to optimize the motivational (and cognitive) state of the student.

The system built a motivational model of the student based on three variables: the student’s effort, their confidence and their independence. The degree of persistence the student showed in solving problems affected the system’s view of the value of the effort variable. Asking for help, or responding positively to unsolicited offers of help from the student affected the value of the independence variable. At the start of a new
problem the system invited the student to make a quick assessment about how they felt about solving it, thus getting some sense of their degree of confidence.

Essentially the system’s reactions were based on rules which expressed a motivational theory about what to do at critical junctures e.g. when a student failed in solving a problem, when they were successful or when they asked for help. The important issue was not so much whether the rules implemented within the system were optimal, but more the fact that the rules were explicit within the system and were open to inspection and adjustment by designers. An example of the rules associated with when the student asked for help is shown in Table V. Note that in this system two of the critical variables were confidence and independence, and for each of these variables there were two value ranges: low and OK on which the condition part of the rules matched.

Table V: What MORE did when the user asked for help (adapted from del Soldato & du Boulay, 1995).

<table>
<thead>
<tr>
<th>Independence</th>
<th>Confidence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LOW</td>
<td>Goal: Facilitate success so provide specific help</td>
<td>Goal: Increase independence so comment on independence and skip providing help</td>
</tr>
<tr>
<td>OK</td>
<td>Normal situation so provide generic help</td>
<td></td>
</tr>
</tbody>
</table>

The motivational theory enshrined in the rules in Table V suggests that when the system detects that confidence of the student it low its main goal is to try to facilitate the student’s success so it provides specific help irrespective of the student’s degree of independence. By contrast when the student’s confidence is within normal bounds, the system differentiates its behaviour depending on the degree of independence. Note that in the case where the independence is OK, the system comments on this and skips providing the requested help. We will return to this issue later.

A limited evaluation of MORE was conducted by comparing reactions of a small number of students who used a version with the motivational rules enabled and when they used the identical system but with the motivational rules disabled. When they used the motivational version they said that they liked using it but did comment on two particular rules that they found problematic. One was the fact that under some circumstances, the system refused to provide help when asked (see Table V), and in some other cases it refused to allow a student to quit from a problem when it thought they ought to try a bit harder. Now it was not that these reactions of the system would have been implausible if produced by a human teacher. Rather, the students’ dismay was that it had been merely a machine that had acted in this way, beyond its status as it were. This issue of “plausibility” was noted by Lepper et al. (1993) and is explored in more detail in du Boulay & Luckin (2001).
Motivation and goal orientation
The next two systems to be described are both extensions of a system, Ecolab II, designed to teach the concepts of food webs and chains to 10-11 year olds (Luckin, 1998; Luckin & du Boulay, 1999; Luckin & Hammerton, 2002). The system offers an on screen laboratory which can be viewed from different perspectives and into which children can introduce various creatures and populations of creatures to see what eats what and how the sizes of population vary over time. A special feature of Ecolab II is that it operates at both the domain level and the metacognitive level. One part of the interface shows which parts of the curriculum have been covered (domain level) but it also shows how much help the child has requested in so doing and also the degree of challenge inherent in the activities that the child has chosen to explore the curriculum (see Figure 5). At points the system comments to the child about her strategy e.g. to advise him or her to seek more challenging tasks or to make better use of the help facilities.

Figure 5: Ecolab II interface showing student which parts of the curriculum have been visited (only the blue node in the network), as well as the amount of help requested (high – the green dice) and the degree of challenge attempted (low – the green bar chart) (adapted from Luckin & Hammerton, 2002).

Ecolab II formed the basis for two new separate variants. The first, Motivational-Ecolab, was concerned to model the student’s degree of motivation and react to that by offering a motivating activity as well as by encasing the interactions with the system in terms of an overall narrative involving two onscreen characters (Rebolledo Mendez, du Boulay, & Luckin, 2005). Figure 6 shows the two onscreen characters, on the left a teacher and on the right a fellow student, whose demeanour in terms of what they say, their tone of voice and their facial expression are adjustable.
The second system, Goal Orientated-Ecolab, was concerned to model the goal orientation of the student and offer an interface and reactions that adjust to (broadly) whether the student was mastery orientated or performance orientated (Martinez-Miron, Harris, du Boulay, Luckin, & Yuill, 2005).

Figure 4: Goal Orientated Ecolab. Upper picture shows kinds of variation between the goal orientated and performance orientated versions. The lower picture shows specific differences interaction.
Because the two systems have similarities and both are extensions of Ecolab II, Table VI shows the categories within which all three systems operate. Motivational Ecolab and Goal orientated Ecolab inherited the underlying mechanisms of Ecolab II.

**Table VI**: Ecolab II (black font), Motivational Ecolab (blue underlined font), Goal orientated Ecolab (pink italic font)

<table>
<thead>
<tr>
<th>Category</th>
<th>Inputs</th>
<th>Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive</strong></td>
<td>Performance, effort, choice of activity</td>
<td>Suggestions about activity choice; <em>use of a quiz, extra domain information</em></td>
</tr>
<tr>
<td><strong>Meta-Cognitive</strong></td>
<td>Difficulty of work chosen, use of available help, in <em>initial degree of motivation</em>, mastery or performance goal orientation</td>
<td>Comments about performance, <em>access to performance information about self or classmates</em>, Open model of degree of challenge taken, use of help, <em>Manner of those comments, Overall narrative with agents</em></td>
</tr>
<tr>
<td><strong>Affective</strong></td>
<td></td>
<td>Praise and encouragement about performance and effort, <em>demeanour of embodied agents (tone of voice, facial expression)</em></td>
</tr>
</tbody>
</table>

Luckin & Hammerton compared learning outcomes of a cohort of 9-11 year old students (N=26) using Ecolab II with an earlier cohort who had used Ecolab (N=26). The difference between the two systems was that Ecolab II had the additional metacognitive scaffolding described earlier. Students of all abilities made learning gains in both cohorts, but it was the low ability students who made the greatest gains in both cohorts. In Ecolab II there was also evidence that the low ability students engaged in more effective learning behaviour, including using the help system more systematically.

Motivational-Ecolab was evaluated by comparing pre/post test outcomes against a version without the motivational scaffolding (Rebolledo Mendez, et al., 2005). Again 9-11 year old children took part (N=29), and children using Motivational-Ecolab scored better in a post-test and a delayed post-test compared to the unenhanced version. They also showed a greater propensity to use the help system both in terms of the quality and quantity of help requested. There was also a positive relation between use of the help system and learning gain. The evaluation of Goal Oriented-Ecolab was more complex, not least because of uncertainties around the notion of goal orientation itself (Martinez-Miron, et al., 2005). In addition, the high variances in the data meant that it was hard to draw firm conclusions about the effectiveness of
the system matching the way that it supported goal orientation to the goal orientation of the students themselves (Martinez Miron, 2008).

DISCUSSION
The purpose of this paper was to introduce systems designed at Sussex operating at the metacognitive, affective and motivational levels. These have each been reported in the literature but not brought together in one place. Table VII summarises the outcomes of the evaluations of the systems described. Note that the outcomes were based on comparisons between two versions of the system, except in the case of MIST.

Table VII: Outcomes of evaluations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MIST</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HabiPro</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIRA</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcolabII</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MORE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>M-Ecolab</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>GO-Ecolab</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Key:
- empty cell means outcome not measured.
- ✓ means outcome measured and found
- ∼ means outcome measured and result equivocal
- × means outcome measured and not found

Where this was measured all of the systems described produced pre/post learning gains, and most enabled the learners to practice particular metacognitive skills while using the system. Only MIRA attempted to measure changes in metacognitive skill as a result of using the system and did not find evidence of this. ATS found some evidence of changes in affect while using the system but no differences between the experimental and control groups on a pre/post basis. No system tried to measure overall changes in motivation as a result of using the system.

Different evaluations used different ways of looking for individual difference effects. So Ecolab II and M-Ecolab looked at the differential effects of ability, GO-Ecolab looked at differential effects of goal orientation, MIST looked at different degrees of study skill sophistication, and MIRA looked at knowledge monitoring ability.

In the companion paper (du Boulay, et al., 2010) we raised a number of theoretical questions:
“These kinds of system raise a number of theoretical and practical questions: What kinds of data are available on which to make inferences about the motivational, metacognitive and affective states of the learner? What is the nature of the theory that links such data to inferred motivational, metacognitive and affective states? What kinds of motivational states are to be distinguished, one from another? What are the relationships between learning and either relatively stable personality traits or relatively transient motivational states and feelings, or less transient affective states such as moods? What are the predictable trajectories between affective states over the duration of a lesson or of a course? And what is the nature of the theory that determines how a caring system might best assist the learner to move away from trajectories or states that might inhibit learning towards those that might enhance it? In other words, what is the nature of the theory that helps the learner follow a trajectory of states that enhances and opens new possibilities (see e.g., D'Mello, Person, & Lehman, 2009; Kort & Reilly, 2002), even if there are individual negative episodes along the way, as opposed to a trajectory that limits possibilities and is dysfunctional or maladaptive, even if there are individual positive states along the way?” (du Boulay, et al., 2010)

Various further questions and suggestions for future directions suggest themselves:

1. **The relation between cognitive knowledge and skill and its associated metacognitive knowledge and skill:** All the systems concerned with metacognition attempted to teach that knowledge and assisted learners to practice those skills in the context of the basic cognitive task, such as reading from texts, solving algebra word problems or learning programming concepts. Both MIST and MIRA were at pains to get the learners to undertake metacognitive activities prior to and then after the cognitive ones. In other words they were concerned to help the learners develop their planning ability as well as their reflective capability. In the various versions of Ecolab, the metacognitive interaction was more directly intertwined with the cognitive. A special feature of MIST was that the system had no capability to monitor directly the cognitive activity of the learner (reading the text) but relied wholly on the learner’s description of that activity. It is an interesting question as to whether that separation of the two levels brings learning benefits.

2. **Scaffolding at the meta level and transfer:** One of the difficult problems of both system design and pedagogic strategy is to find ways that the metacognitive skills practiced are sufficiently understood and learned that they can (a) be recalled and reused in later sessions with the same system, (b) recalled and transferred to other systems that might support them differently, and (c) recalled and transferred to other contexts (including learning contexts without explicit metacognitive support). Part of the solution may lie in getting learners to talk with each other about such skills (see next point), and part of the solution may lie in taking learners through a sequence of contexts, some with fadeable metacognitive scaffolding and some without scaffolding at all in order to help learners experience and articulate the benefits of operating at the metacognitive level.

3. **The benefit of working with other learners:** Both MIST and HabiPro were designed to be used by more than a single learner. MIST specifically asked one learner to ask for help from the other or offer guidance to the other as the situation demanded. They were invited to review each other’s answers to planning questions about reading intentions in the hope that the offline conversations at the metacognitive level would bring extra benefit. In the case of HabiPro, the
interventions by the simulated student were viewed by the two human students also working thereby creating a kind of public forum. In Motivational-Ecolab, part of the motivational narrative involved input from a simulated other student too. So there is the question as to the degree that interactions with another student bring benefits at the metacognitive level.

4. The role of the meta level: One of the important strategies at the metacognitive level is to make knowledge and skill at that level explicit and open to interactions through the system interface. When it comes to building systems that attempt to operate at the affective or motivational level, this has been much less in evidence. The system may make a move to improve the affective or motivational state of the student (e.g. introducing a quiz in Motivational-Ecolab, suggesting a relaxation exercise in ATS, or offering praise for effort in MORE), but there is no interaction with the student about this at the meta level. Did it work, would you like something different, what works for you etc? One could imagine a system that asked learners questions about their affective or motivational states and invited planning and reflection on how best to deal with them, just as MIST did at the metacognitive level. In terms of bringing about transfer, one may need to apply similar regimes of reducing scaffolding and taking students through a variety of learning contexts as was suggested above in the case of metacognition.

5. The goal of meta-motivationally intelligent systems: A meta-motivationally intelligent system would be able to reason about the motivational dynamics of the learners it was working with, understanding, for example, what might be the best ways of dealing with a learner’s lack of motivation (du Boulay, in press). But it would be able to go beyond that by helping the learners themselves to understand their own motivational dynamics better and helping them practice the regulation of their motivation, in a not dissimilar way that sportsmen and sportswomen learn how to overcome the tedium of practice and in competitive sports learn how to “manage themselves” through situations where they are starting to lose, or indeed starting to win.

6. Finally there is the issue of plausibility (du Boulay & Luckin, 2001; du Boulay, Luckin, & del Soldato, 1999): The agenda implicit in the work described on the paper is that we should seek to embody in systems the skills of the most expert human teachers. Putting aside the difficulty of doing that in its own right there remains the issue of whether human students are always willing to react favourably to tactics that might be acceptable from a human teacher but not so from a machine teacher. Two examples from our work include annoyance from learners when MORE refused a help request (when it judged that the help was not really needed del Soldato & du Boulay, 1995), and when the peer learner in LECOBA (Uresti & du Boulay, 2004) “forgot” some of the ideas that the human student had taught it.

ACKNOWLEDGEMENTS
The author thanks his many co-workers in the IDEAs Laboratory and the Human Centred Technology Research Group for their huge contributions to this paper.
REFERENCES


