Modeling Human Learning as a Cooperative Multi Agent Interaction

Mathew Davies
Department of Computer Science
Columbia University
1214 Amsterdam Avenue, Mailcode 0401
New York, NY 10027 USA
mdavies@cs.columbia.edu

Elizabeth Sklar
Department of Computer Science
Columbia University
1214 Amsterdam Avenue, Mailcode 0401
New York, NY 10027 USA
sklar@cs.columbia.edu

ABSTRACT

In the Meta-Game of Learning, fashioned after the Iterated Prisoner's Dilemma, teacher and student become agents that interact in an implicit game played within the context of formal learning. The Teacher can pose to the Student either HARD or EASY questions; the Student can respond with either RIGHT or WRONG answers. When the hard questions are answered correctly, the Student is *learning* — the result of a *cooperative* action on the part of the two participating agents. In this paper we expand on the MGL, modeling the Student's behavior and suggesting a simple basis for the complexity underlying real students' responses in formal educational settings. We show that even a very simple model can account for much of the richness of typical classroom dynamics. We consider *motivation*, *emotion* and *ability* as contributing factors and present results of preliminary experiments applying our model and varying each of these factors.

1. INTRODUCTION

Based on the assumption that learning is fundamentally interactive, we are studying the dynamics of interactions that occur in formal educational settings. Our longterm direction with this work is the exploration of optimal decision-making by pedagogical agents. In earlier work [3], we modeled the interactions between teachers and students and the manner in which these agents are rewarded as a *Meta-Game of Learning* (MGL) [2], which is fashioned after the Iterated Prisoner's Dilemma (IPD) [1]. In the MGL, we consider the Teacher and Student to be players participating in an implicit game obeying the constraints of the IPD. Each player can make one of two moves, or choices, at each iteration of the game. The Teacher goes first and presents to the Student either a HARD or an EASY question. The Student responds with either a RIGHT or WRONG answer. Figure 1 illustrates the MGL using a matrix of possible question-answer pairs, labeling the type of experience associated with the pair in each case.

Student:	RIGHT	WRONG
Teacher: HARD	*learning*	*frustration*
EASY	*verification*	*boredom*

Figure 1: The Meta-Game of Learning (MGL).

We take the overarching goal of the education system to be sustained *learning* by the Student. This entails a mix of the four possible states, with a drive towards the HARD question and RIGHT answer state. We assume that both Student and Teacher are motivated by this student-centered goal, and define individual agent behaviors (on the part of either Student or Teacher) that try to advance the Student towards this goal as *Cooperation*, while behaviors that do not seek to advance the Student are defined to be *Defection*. In other words, the Student cooperates by trying to learn, while the Teacher cooperates by enabling the Student to learn, i.e. by presenting appropriately challenging questions. This framework reflects the assumption that interactions comprising active learning in formal settings are fundamentally cooperative. As in the IPD, in order for maximum mutual payoff to occur, both agents must cooperate. We leave the actual payoffs of the interaction unspecified, although it is not difficult to formulate particular payoffs under the constraints of IPD that are consistent with the interpretation we will present.

Although many different behaviors and actions may constitute either cooperation or defection, our framework reduces all such activity to the idealized choices, or actions, available to each agent. Although we call a given move or action in the game a "choice" in this context, it should be understood that it is not a choice in the usual sense, because it is not independently made but depends on other factors. Thus we can identify cooperation with the choice of a hard question by the Teacher, or a right answer by the Student, and defection with the choice of an easy question or a wrong answer. If the Student tries to answer a question well enough, in other words, we presume the answer to be right; if the Student does not try hard enough, the answer is presumed to be wrong. While this kind of presumption may appear unfounded — a real student, for instance, may try hard yet still fail to answer a difficult question correctly — the abstraction of the MGL justifies it. Firstly, the Student and Teacher can only judge the other's action by what is visible to each agent: for the Student, whether a question was hard or easy; for the Teacher, whether a Student answered right or wrong. Secondly, what matters in the long run is not whether a particular intended action failed, but the aggregate results of what the actor was trying to do at each step. To the extent that real students and teachers can evaluate each other's actions, they can make allowances

for imperfection if the intention was clear. The chaining of result to intent permits the construction of models that are not particular to any knowledge domain, yet capture essential features of interaction dynamics.

In the work presented here, we utilize the MGL framework to model complexities in Student behavior. We posit that, within the four simple experiences shown in figure 1, the response a Student chooses depends on a number of internal attributes with definite values.

- *ability (A)* — determines the relative ease with which the Student learns a new concept.
- *motivation (M)* — determines, in part, the likelihood of cooperation; in general, a high value for motivation is commensurate with cooperation (i.e., responding with the RIGHT answer, as the Student is trying) and the converse for defection (i.e., responding with the WRONG answer, as the Student is not trying).
- *emotion (E)* — represents the Student's emotional state, or relative happiness, contentment, etc.; we take emotion to be a variable distinct from, but associated with, motivation.

Although conceivably these attributes should be considered continuous and possibly vector-valued functions of time, we employ discrete (binary or n-ary) scalar values that change through the course of interaction according to simple rules. These values range between a minimum (zero) and some maximum. The specification of a model must determine how an agent's attribute values map to the agent's intent to cooperate or defect.

We have defined a set of simple rules determining changes in each of these attributes in response to the game state selected by cooperating and defecting behaviors on the part of the interacting agents. These attributes in turn determine the next choice, or action, of the Student, and so on through a sequence of questions posed by the Teacher. (In the discussion that follows, we use the terms choice and action interchangeably when describing metagame interactions). We define *progress (P)* as a value indexing an ordered abstract knowledge domain composed of questions, each with an absolute level of difficulty, or *hardness (H)*, and define *learning* as a positive change in progress. Regardless of the hardness of a question, a Student that answers a question correctly is taken to have learned the associated concept, and thus has made progress. Note that a Student can only make progress if she cooperates, regardless of what the Teacher chooses to do.

The next section describes two foundational rules of response, one each for motivation and for emotion. We then describe a model of an idealized lecture scenario based upon these rules, which particularizes interactions between one Teacher and a number of Students. Section 4 shows preliminary results of a mathematical simulation of this model. We then supplement the basic lecture model with an element of tutoring and show the change in simulation results. We close with a discussion of current and future work.

2. RULES OF RESPONSE

Our foundational rule for motivation can be stated very simply: a reward for a given move to cooperate or defect encourages pursuing the same choice next time; a punishment for a given move encourages making the opposite choice next time. However, there is some complication in that the reward or punishment derived by a Student depends on the Teacher's choice. We need not specify actual payoff values to determine which case applies: since this is a type of Prisoner's Dilemma, if the Teacher cooperates, the Student is rewarded and will tend to repeat her last action, whereas if the Teacher defects, the Student will tend to follow a different course of action the next time. The rule for emotion is even simpler: answering right improves the Student's emotional state, and answering wrong detracts from it. These rules are expressed using the matrix shown in Figure 2.

Student :	Cooperate (C)	Defect (D)
Teacher : Cooperate (C)	$M+, E+$	$M-, E-$
Defect (D)	$M-, E+$	$M+, E-$

Figure 2:
Changes in Student's Motivation (M) and Emotion (E). The positive indicator (+) means that the value of the attribute increases; negative (-) indicates that the attribute decreases.

The effect of these rules may be seen as follows. Firstly, note that the Student's emotion E will invariably go up if she answers correctly, or down if she answers incorrectly; the change in emotion depends only on the Student's action. Conversely, the change in motivation depends only on the Teacher's choice in the current round. These changes may be loosely interpreted as corresponding to real situations, although we stress here that no fixed interpretation is necessary (or even desirable), as long as there is some consistent interpretation within the MGL framework. If the Teacher and Student both cooperate — for instance, a teacher presents a challenging question that a student correctly answers, affirming the student's effort — then the Student's motivation M and emotion E both increase (or, possibly, remain high) and the Student makes progress, i.e., P increases. If the Teacher defects while the Student cooperates, although the Student's emotion E goes up — she did, after all, answer the question correctly — the Student's motivation M goes down (or remains low), since the easy question required little effort. The Student becomes less likely to cooperate the next time, although she still made progress in the current round.

The interpretation of defection for the Student is less intuitive. If the Teacher cooperates with a challenging question while the Student defects, the Student's motivation will go down, together with emotion — a student may still learn from incorrectly answered questions, but if the questions are consistently too difficult, failure is certain, and there is little incentive to apply much effort to them. If both Student and Teacher defect, then the Student's emotion still decreases, but her motivation increases — failing to give a right answer that one was capable of giving has no learning value at all, and a student in that situation tends to feel renewed incentive to work.

Although it is possible to interpret the changes in motivation and emotion in each case in terms of explicit IPD payoffs, distinct from our definition of progress, we refrain from doing so. Our response rules have no direct relationship to either payoffs or progress in the MGL framework; rather, they attempt to capture indirectly the dynamic rela-

tionship between agents engaged in cooperative interactions. Why don't students and teachers simply cooperate all of the time? Embedding our response rules in a particular model suggests that, while individual students may tend toward an equilibrium state, a heterogeneous collection of students will exhibit complex behaviors in a formal setting precisely because of their heterogeneity.

3. THE LECTURE MODEL: A SIMPLE MULTI AGENT MGL

In our prototypical Lecture model, one pedagogical agent (the Teacher) interacts simultaneously with many (n) Student agents. The Teacher presents a finite series of related concepts from the knowledge domain in some particular order, asks the students questions about each concept in turn, and identifies the answer as right or wrong. "Asking" in this case may be either in person (explicitly), through some indirect vehicle (e.g. a test or homework), or via rhetorical questions (implicitly). We assume continuity across the series of interactions, disregarding the effect of breaks in time between iterations of the interaction. We then partition the simultaneous interactions between the teacher and the students into n two-agent meta-games, each one involving the Teacher and one Student interacting over the series of concepts. We assume that the Teacher's questions are the same for each Student, but the Students' responses are taken to be independent from those of other Students.

As mentioned earlier, agents are not omniscient; no agent truly knows what action another agent intended to take, but can only judge from what the agent appears to have done. Since the Teacher's effective action (cooperation or defection) depends on the Student's perception, we introduce a rule to determine the action taken by the Teacher from the point of view of the Student:

```
if (Student.ability > Concept.difficulty)
  Teacher.action ← Defect
else
  Teacher.action ← Cooperate
```

In other words, if the student's ability is sufficiently high, the concept seems easy, and the teacher appears not to be challenging the student, i.e., the teacher appears to be defecting. Conversely, if the concept is difficult to the student, the teacher appears to be cooperating. In practice, the mapping from difficulty to perceived action can be either deterministic or probabilistic. Our implementation of the lecture model, given in the next section, uses the latter method; this approximates both errors in judgment, and the possibility that students of equal ability may not all perceive a question the same way.

Similarly, to determine the Student's effective choice on a move, as perceived by the Teacher, we introduce a rule that relates a Student's motivation and emotion levels to the Student's likelihood of answering a real or hypothetical evaluation question correctly. This rule is central to the dynamic behavior of the lecture model, because it feeds the changes in attributes for one iteration, as determined by the response rules, into the next iteration. If the Student is both emotionally positive and highly motivated (E high, M high), then she is presumed to have learned the concept well enough to answer correctly no matter the difficulty of the concept, which is implies cooperation. If the Student is both unhappy and unmotivated (E low, M low), then she is presumed inattentive and will answer incorrectly no matter the (lack of) difficulty, which implies defection.

If E and M are not binary values, we can define a numerical threshold that determines what values are "low" and "high". When the E and M states are dissimilar — E high and M low, or the converse — the Student's disposition is less than optimal, giving a probability of answering correctly that is in proportion to the relative difficulty of the concept. We compute this probability `delta` as:

```
delta ← (Student.ability/Concept.difficulty)
```

If `delta` < 1, the Student may answer incorrectly, (since `Concept.difficulty < Student.ability`, the question seemed hard), whereas if `delta` > 1, the Student will answer correctly (the question seemed easy). We define the number `chance` to be a volatile random number between 0 and 1, and express the Student action rule as follows:

```
if (E low & M low)
    Student.action ← Defect
elseif (E high & M high)
    Student.action ← Cooperate
else
    if chance > delta
       Student.action ← Defect
    else
       Student.action ← Cooperate
```

This model constitutes not one, but n simultaneous, non-interacting metagames. We describe one particular implementation of this model below.

4. EXPERIMENTS

We have simulated the Lecture model using two different modes of Teacher behavior. In both modes, each Student gets the same series of questions (i.e., questions of the same difficulty at a given level of progress). In the impersonal mode, the Student gets no help upon missing a question, but must continue trying to answer until getting it right. This has a negative impact on progress because the harder the question, in the absence of any other changes, the more likely the Student is to become stuck in constant defection (when both E and M are low). In the personalized mode, the Teacher is able to address this problem by backtracking the Student to successively easier questions until she starts answering correctly, then returning the Student to where she left off.

At the beginning of each simulation, the student agents are assigned random values for ability, emotion and motivation. These values range from 0 to 20 for ability, and from 0 to one less than the degree of discreteness (`degree`) for emotion E and motivation M — 1 for binary values, 3 for 4-ary values, etc. At each iteration, all students at a given point of progress are presented with the same question (i.e., of the same difficulty level). The difficulty level increases uniformly for all students, one level per cooperation, although cooperation itself is not uniform. Each Student's response to a question is a function of the Student's ability, emotion and motivation values, as well as the difficulty of the question, according to the rules described in previous sections. After

determining whether the Student will *cooperate* or *defect*, the Student's emotion and motivation values are updated.

The first set of figures, Figures 3 through 8, show the results of simulating the Lecture model in the impersonal teacher mode. The simulation was run for combinations of the number of students (`nstudents`), the number of steps, or iterations (`nsteps`), and the degree of discreteness. The number of students was set to either 1000 or 10000, not to mean gargantuan class size, but for statistical smoothing and to verify consistency. The number of steps was either 20 or 100, showing results for both short and long series of interactions. The degree of discreteness was either binary (values 0,1) or 4-ary (values 0,1,2,3).

Each figure contains two sets of graphs. The upper graph is a range plot of progress against ability for all students at the end of the interaction series; this shows the range but not the distribution of ending states. Since difficulty increases by one with each progress step, the horizontal axis represents both maximum progress and difficulty of the final question; for example, a student ending up in the 8th horizontal position progressed to the 8th concept (e.g., out of 20 in figure 3). Note that if multiple students end up at the same position, then only a single mark is indicated on the plot. The corresponding lower graphs illustrate the normalized distribution of students as a function of maximum progress at the end of the simulation, and the average final value of motivation, emotion and ability (respectively) for the students at each value of progress. For example, in the same figure, about .5 of the students got to level 20; the final motivation value has little correlation to progress, but final emotion value does clearly correlate to progress (as does ability). These distributions may be seen more clearly in the corresponding graphs with 10 times the number of students (equivalently, averaged over 10 identical trials), in figure 4. The vertical axes for these lower graphs follow the range of possible values for each variable.

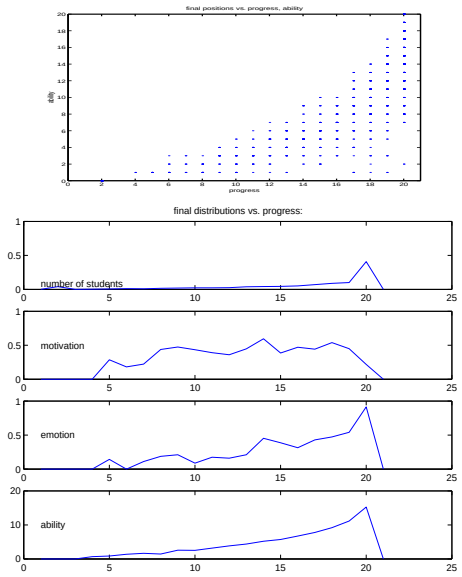

Figure 3:
degree=binary, nsteps=20,
nstudents=1,000, maximum ability=20

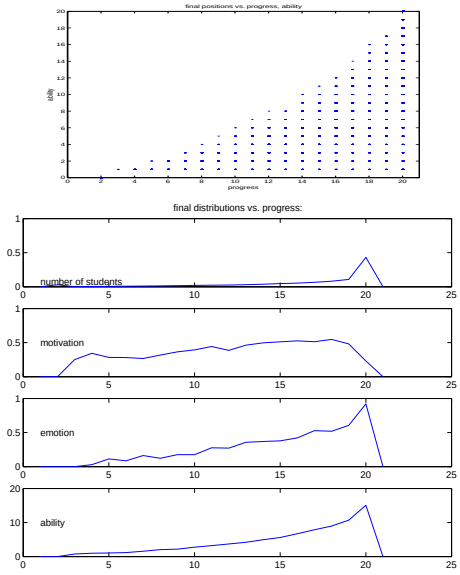

Figure 4:
degree=binary, nsteps=20,
nstudents=10,000, maximum ability=20

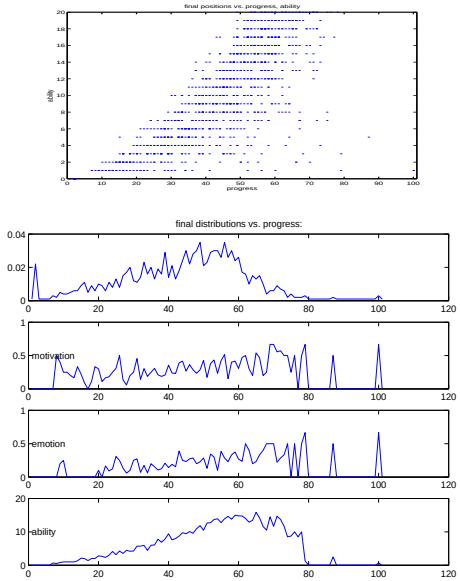

Figure 5:
degree=binary, nsteps=100,
nstudents=1,000, maximum ability=20

In varying not just the degree of discreteness, but also the range of abilities, starting level of difficulty, progression of difficulties (linearly increasing, flat, random), number of iterations and number of students, the behavior of the model proved to be stable, as expected, with the relative progress of students tending to increase with ability and the smoothness of the distribution graphs increasing as expected with the number of students simulated. However, within this broad stability we found a number of significant features that are logical, but not immediately obvious, results of the model.

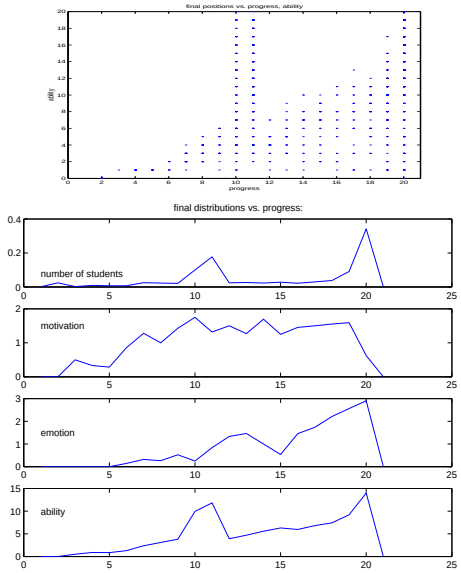

Figure 6:
degree=4-ary, nsteps=20,
nstudents=1,000, maximum ability=20

In the beginning stages of iteration, when the ability of any student may be greater than the difficulty of the question, there is a significant difference between binary and n-ary degrees of discreteness for emotion and motivation. Careful inspection of the model shows that, for the binary case, if a student defects on an easy question, the student is likely to cooperate on the next question[1]. However, in the n-ary case, if a student with low emotion level and borderline motivation cooperates on an easy question, the student is doomed to defect at least once before cooperating. This is the case because while emotion does not cross the threshold to "high", motivation does cross the threshold to "low". With both low emotion and low motivation, the student will defect. This pattern may then repeat. In short, students with high emotion levels will progress on virtually every iteration, while those with low emotion will progress on at most every other iteration. This is seen in the simulation as a band of students of every ability moves (over time) half as quickly as the students with uniformly high emotion level (see in particular figures 6 and 7). As degree of discreteness increases, the width of this band increases. The size of possible changes in emotion or motivation per iteration relative to the degree of discreteness thus has a significant but subtle effect on the behavior of the model that is difficult to isolate.

[1]unless the size of delta is very small

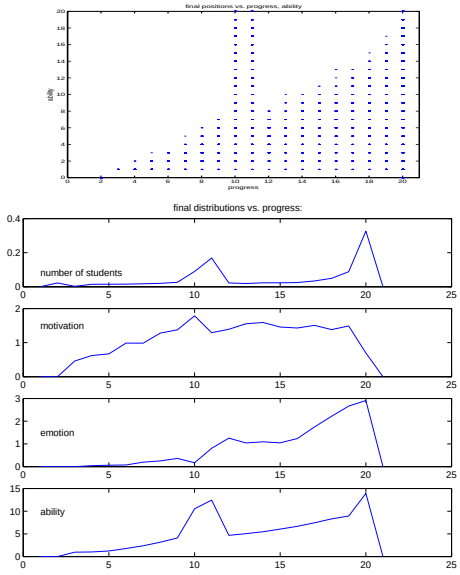

Figure 7:
degree=4-ary, nsteps=20,
nstudents=10,000, maximum ability=20

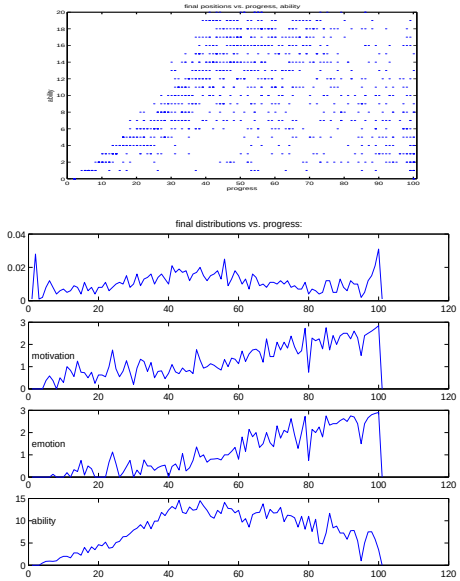

Figure 8:
degree=4-ary, nsteps=100,
nstudents=1,000, maximum ability=20

Although a student's progress is obviously proportional to the number of times the student cooperated, the student's final motivation level is not a strong correlate with progress until questions become very hard. For longer series of iterations (e.g. figures 5 and 8), the most consistent correlate of progress tends to be not ability, but rather emotion level! Although ability is obviously a strong determinant of progress, we found that students with low ability values may, under the right circumstances, actually perform consistently better than those with high ability. This happens because students with low ability find more of the questions challenging, and are thus more likely to enter a state of high motivation and high emotion early, in which they will remain as long as the difficulty of questions continues to increase. Such students may be seen as the rightmost outliers in the bottom right-hand corner of the range graphs, and account for the unexpected tail of the ability distribution in the 100-step series, where the average ability of students at the highest progress position is actually lower than that of students at many lower positions.

The second set of figures (9 through 14) show the results of simulation with the same parameter combinations described above, in the same order, but in the personalized teacher mode. Here, the teacher may respond differentially to each student, preventing students from getting stuck in states of needless defection. A "personal best" in terms of progress P is recorded for each student; if the student had to backtrack, progress (when it happens) resumes at the personal best. In addition, if a student finds questions too easy, the student is allowed to skip ahead, until reaching questions that are difficult for that student, preventing states of low motivation associated with boredom. The second set of figures shows a dramatic rightward shift of the distribution of students with progress.

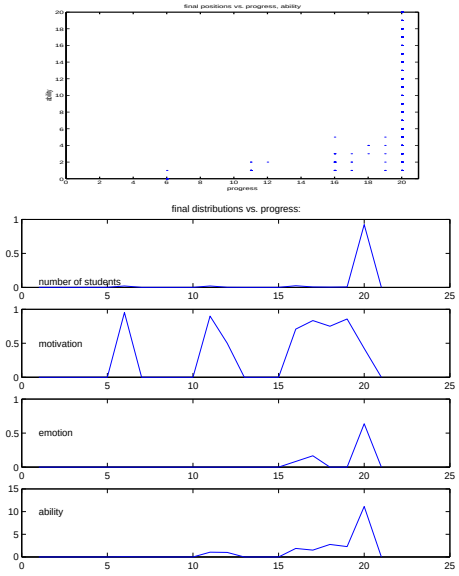

Figure 9:
degree=binary, nsteps=20,
nstudents=1,000, maximum ability=20

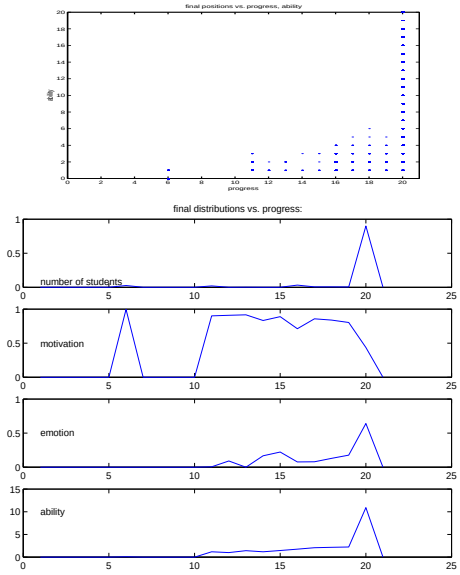

Figure 10:
degree=binary, nsteps=20,
nstudents=10,000, maximum ability=20

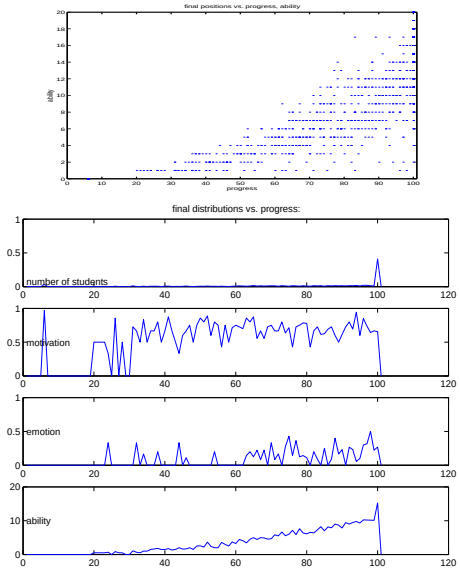

Figure 11:
degree=binary, nsteps=100,
nstudents=1,000, maximum ability=20

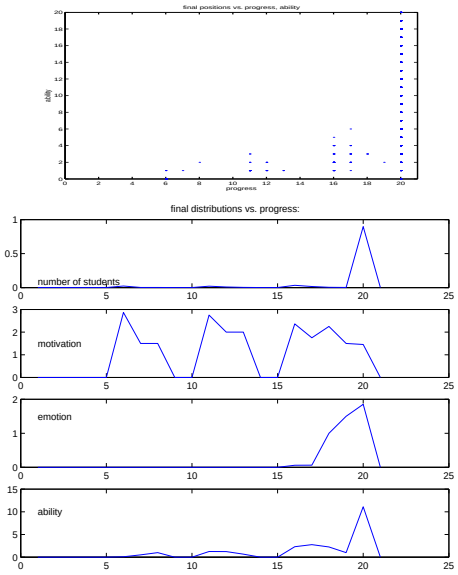

Figure 12:
degree=4-ary, nsteps=20,
nstudents=1,000, maximum ability=20

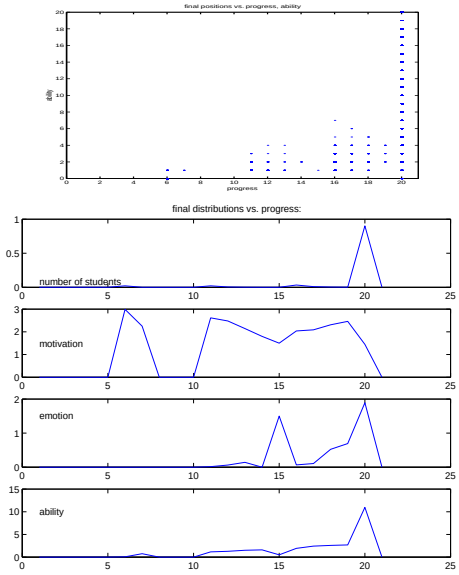

Figure 13:
degree=4-ary, nsteps=20,
nstudents=10,000, maximum ability=20

5. DISCUSSION

The Lecture Model presented here demonstrates that a set of simple rules operating in the framework of the Meta-Game of Learning may capture many features of complex cooperative interactions in real settings. We believe this approach holds some promise for understanding the features of such interactions in real educational settings and applying that understanding to the construction of optimal pedagogical agents. While we will continue to clarify and analyze the Lecture Model, we also plan to model additional scenarios corresponding to other learning modalities, and refine these new models according to current research on the psychological factors affecting the performance and behavior of students in real learning environments.

We also have many theoretical questions to consider. Consider a classroom with one teacher and a number of students, all of whom may interact with each other. Is there any tractable way to represent the $O(n^2)$ interactions among the n students? What happens if there is more than one type of metagame to consider, i.e. each agent is pursuing more than just a goal of optimal learning? Ideally, each agent can perceive whether other agents cooperated or defected on a given interaction, but in practice — especially in a group setting — different individuals may perceive the same behavior differently. The problem of perception is compounded in any of these more complicated cases. One way of handling this problem was adopted in the Lecture Model, but that entailed many simplifications. Can we account for differences in perception more expressively? We plan to investigate these and other fundamental questions concerning the MGL, and seek applications to this approach in the domain of other classes of cooperative interactions.

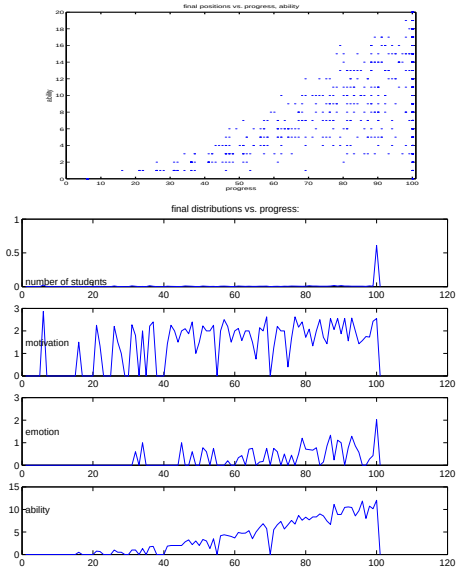

Figure 14:
degree=4-ary, nsteps=100,
nstudents=1,000, maximum ability=20

6. REFERENCES

[1] R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

[2] J. B. Pollack and A. D. Blair. Co-evolution in the successful learning of backgammon strategy. *Machine Learning*, 32:225–240, 1998.

[3] E. Sklar, A. D.Blair, and J. B. Pollack. Co-evolutionary learning: Machines and humans schooling together. In *Workshop on Current Trends and Applications of Artificial Intelligence in Education: 4th World Congress on Expert Systems*, 1998.