"Oh, dear Stacy!" Social interaction, elaboration, and learning with teachable agents

Amy Ogan[*], Samantha Finkelstein[*], Elijah Mayfield[*], Claudia D'Adamo[†], Noboru Matsuda[*], Justine Cassell[*]

[*]Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
{aeo, slfink, emayfiel, mazda, justine}@cs.cmu.edu

[†]Wheaton College
26 E. Main Street
Norton, MA 02766
claudiadadamo@gmail.com

ABSTRACT

Understanding how children perceive and interact with teachable agents (systems where children learn through teaching a synthetic character embedded in an intelligent tutoring system) can provide insight into the effects of social interaction on learning with intelligent tutoring systems. We describe results from a think-aloud study where children were instructed to narrate their experience teaching Stacy, an agent who can learn to solve linear equations with the student's help. We found treating her as a partner, primarily through aligning oneself with Stacy using pronouns like *you* or *we* rather than *she* or *it* significantly correlates with student learning, as do playful face-threatening comments such as teasing, while elaborate explanations of Stacy's behavior in the third-person and formal tutoring statements reduce learning gains. Additionally, we found that the agent's mistakes were a significant predictor for students shifting away from alignment with the agent.

Author Keywords

Teachable agents, peer tutoring, impoliteness, rapport, ECA

ACM Classification Keywords

H.5.2 [Information interfaces & presentation]: User Interfaces - Graphical user interfaces.; K.3.1 [Computers & Education]: Computer Uses in Education

General Terms

Human factors, design

INTRODUCTION

Teachable agents are a specific type of tutoring system that provide a platform for children to learn through teaching [3]. Such systems give students the opportunity to engage in peer tutoring exercises that may increase self-efficacy and motivation, and even contribute to learning [8, 17].

The success of teachable agents has been referred to as the tutor learning effect [3]. A number of theories have been proposed to explain this effect, including increased motivation to learn the material [22], increased reflection on already learned material [18], and increased effort turning tutor knowledge into coherent, communicable ideas [9, 10, 28].

Among real children, while both tutors and tutees achieve significant learning gains from peer tutoring sessions, peer tutors learn more when their tutees struggle with the material [27]. This increase in learning gains is hypothesized to relate to increased reflection, self-explanation, and necessary reworkings of the problem from multiple perspectives. This may even lead to the tutor learning additional domain material not explicitly covered in the session [22].

Unfortunately, Walker et al. [27] found that tutee errors, while helpful for tutor learning gains, generally lead to less learning for the tutee. Research into the development of successful teachable agents can address this issue, as agents may play the role of a struggling tutee without evoking concern about detrimental consequences for a child. Teachable agents also allow researchers to examine how specific tutee behaviors affect how different children tutor, and thus learn, in identical educational environments.

However, one of the notable challenges with using teachable agents is that there are many components of human peer tutoring that are still not completely understood. For example, researchers have proposed that there are substantial social aspects of peer tutoring that are responsible for evoking tutor learning effects, such as a strong feeling of accountability for ensuring the tutee is learning the proper information [23], as well as a desire to avoid the face-threat of not being able to fully respond to tutee questions [27]. While prior research has shown that children do treat virtual characters similarly to peers in both language use and nonverbal behavior [5], one of the open questions in teachable agent research is whether child tutors are capable of the social motivations described here with a virtual tutee, and

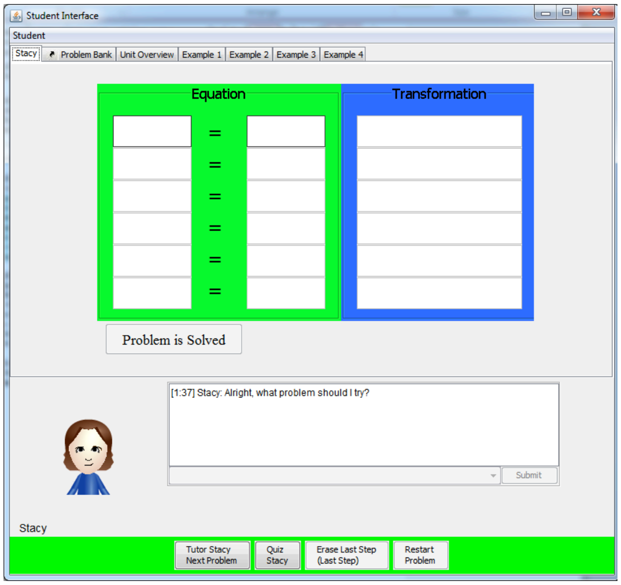
CHI 2012, May 5-10, 2012, Austin, TX, USA.

whether these social behaviors effect the same tutor learning benefits that can be seen with human peer tutoring.

While *cognitive* process data is relatively easy to collect in a technologically-enhanced learning system in which students work through problems [26], *social* process data that elucidates children's relationship with the agent is not. To our knowledge, analyses connecting the social processes that occur in either human-agent or human-human peer tutoring to learning gains have not been carried out, making it difficult to understand how social perceptions affect and change the course of these educational sessions. In this work, we therefore examine how children interact with Stacy, a teachable agent designed to learn linear equations with the help of a child tutor.

Using a think-aloud technique, we assess how children talk to and about Stacy throughout two tutoring sessions, and how their dialogue changes based on Stacy's success, perceived competence, and the length of time the students spend working with the agent. We examine the varying levels to which students choose to suspend disbelief and talk to Stacy as a peer – applauding her successes and reassuring her after failures – and when they instead choose to align themselves with the human experimenter in the room and refer to Stacy as *it* or *she*. We explore these linguistic nuances as they relate to the participants' social behaviors, and examine how these factors, among others, affect learning. We present results that indicate that it is in fact primarily social behaviors that correlate with increased learning gains, and that an *outside-system* perspective, where perceptions of Stacy's partner status are abandoned for viewing Stacy as a *she* or an *it,* predict fewer learning gains.

By examining how children interact with teachable agents, we begin to understand the social processes of human peer tutoring. This also allows us to delve into the more general area of human-agent interaction, addressing foundational questions about how children perceive the agency and competency of virtual characters, with implications for designing better teachable agents as learning interventions, and better agents in general.



[1:37] Stacy: Alright, what problem should I try?

Figure 1. The SimStudent interface, with Stacy in the lower right corner.

RELATED WORK

The efficacy of teachable agents has received support in the literature, with several systems demonstrating the success of this intervention for user learning gains (e.g., [3, 16]). Although many more teachable agent systems have been developed than have been evaluated, it has been shown that children can achieve learning gains by tutoring a teachable agent. In fact, this learning result can be stronger than when the child is being taught by a virtual tutor, as demonstrated in an evaluation of the Betty's Brain system [3].

Investigators are also trying to understand the impact of social moves with teachable agents. Some propose that bringing off-task social conversation into educational dialogues may allow for cognitive rest, increase engagement, provide memory cues, and promote trust and rapport-building with the agent [11]. Gulz et al. [11] developed an interface where an embodied agent learns through either simply *observing* what the child is doing or requiring the child to explicitly explain the rules using a multiple choice dialogue interface. The system also allows users to engage in free open-ended chat with the agent as both a motivational tool and a way to evaluate if this type of behavior might have an effect on self-efficacy and generally improve feelings about math. They found a trend indicating that students allowed to engage in off-task chat had a more positive game experience and that their teachable agents had learned more of the material, but the evaluation did not take social process data about the type of social interaction nor how it changed over the course of the interaction into account.

A substantial amount of research also indicates that one's perceptions of the motivations for learning and of the learning partner affect how people speak and even the resulting learning gains. For example, Bargh and Schul [2] found that people primed to believe they were studying for a quiz to teach others about the material performed much better on the assessment than people who were preparing for the exam for themselves. This result is mirrored in virtual agent studies, where children achieve smaller learning gains when they believe they are teaching a virtual agent that represents themselves than when they are teaching a virtual agent that is presented as a different (virtual) student [6]. Chase et al. refer to this as the protégé effect. When interacting with the agent representing an other, students spend more time on learning activities, attribute mental states and responsibility to their agents, and are more likely to acknowledge errors by displaying negative affect and justifying and explaining why their agent has failed. However, the social mechanisms behind this effect have not been explored.

There is also research indicating that these varied learning effects change depending on the user's *belief* about whether they are interacting with a human or a computer agent. Even when agent responses are identical, students do not spontaneously offer the kinds of self-explanations to an agent that they produce when they believe they are interacting with a human. Instead they tend to answer questions with short keywords, providing no explanation [19, 20].

Okita et al. also explored whether the mere belief that a student was interacting with another real person makes a difference in learning gains [14]. Participants were given a script to use to teach a virtual character who responded identically to everyone, controlling for potential differences in dialogue. Regardless, participants who believed they were talking to a virtual character that was an avatar of a human in another room learned more from the tutoring session than participants who believed they were talking to an autonomous agent.

There are a number of potential explanations for this phenomenon. Students have reported that the social motivations of teaching, such as feeling accountability for helping another person prepare for an exam, forced them to gain deeper understanding of the materials [3]. Chase et al. [6], based on attribution statements for success and failure, hypothesize another social explanation for the learning by teaching effect. By having a second party who shares in the interaction and can take the blame for mistakes, rather than only (a representation of) oneself, the social implications could be that the student's ego is protected from the psychological ramifications of failure, which might in turn facilitate learning.

Conversely, other students interviewed by Biswas et al. [3] proposed a more cognitive explanation of the effectiveness of the tutor learning effect - that it is the need for the clear, conceptual organization of materials required by teaching that produces learning gains. Additionally, the explicit self-explanation that must occur in order to teach someone else has also been hypothesized as the main factor responsible for tutor learning [22]. While a cognitive explanation would hold regardless of who (or what) tutors believe they are interacting with, the mere belief results described above require the tutor to attribute some form of agency or social motivation to the teachable agent. Those results, then, suggest that there is some social aspect affecting the learning process, though details have not been explored until the current study.

HYPOTHESES

Cognitive hypotheses of learning by teaching suggest that tutors will engage in more mental organization of the material and perform more self-explanation as they tutor, leading to learning gains [9,10,15,19,24]. Therefore, we expected analysis of think-aloud protocols to demonstrate that (1) thinking about the state of the agent's knowledge, (2) reflecting on the agent's performance, and (3) providing extended explanations of domain material would result in improved learning gains for the participant.

On the other hand, previous literature has also hypothesized that it is social factors that motivate the tutor effect learning gains [3, 6, 10, 14]. Given conflicting prior work on whether social relationships can be formed with virtual agents [5,15,16,17] we chose to look at the type of language students used when referring to the agent as a clue to their social stance. We expected that speaking directly to the agent using pronouns such as you (e.g. "you got it right, Stacy"), which we call inside-system language, would be correlated with learning. Conversely, we expected that outside-system language, i.e. referring to the agent as she or it, would be less associated with learning gains because it demonstrates a reduced social relationship with the virtual peer. Similarly, we expected that increased use of explicit social dialogue moves in the think-aloud would be positively associated with learning gains.

In summary, we are interested in three primary questions: how do (1) increased cognitive reflection moves (2) inside-system vs. outside-system language and (3) increased social moves correlate with learning? We expected that both cognitive and social moves would improve learning gains, while outside-system language would hurt learning gains. Additionally, to support the creation of future teachable agent systems that can re-engage the child with the agent right as they began to slip away, we investigate what factors may affect shifts in alignment throughout the dialogue. We predicted that Stacy's competency would predict alignment-shifting, with students tending to use inside-system language when Stacy performs well, and outside-system language when Stacy begins to make mistakes.

SIMSTUDENT DESCRIPTION

Our study was carried out using the SimStudent platform [12]. SimStudent is a Learning by Teaching environment in which students interact with a virtual tutee named Stacy which inductively learns procedural rules in various domains. SimStudent starts off the interaction with a knowledge base of production rules that relate to a specific domain. In our work, the domain is linear equations, and the tutee's knowledge base includes the four basic math operations. SimStudent modifies and adds production rules to this knowledge base as students demonstrate problems.

As shown in Figure 1, the SimStudent interface consists of a set of domain overview materials, a set of worked-out example problems, a problem bank sorted by problem difficulty, and an interface for completing problems. While working, students create a linear equation and enter it into the interface for Stacy to try. Stacy completes steps based on her current production rules. After each step, Stacy asks the tutor if the rule she just applied was a good move. The response reinforces her learning algorithm, and allows the tutor to recognize and correct errors.

At times, Stacy may not have an appropriate rule to apply, and then will ask the tutor for help entering the next step. When a student demonstrates a step for Stacy, she creates a generalized rule by checking which operators can result in the input the student provides. If the example is divide by 3 for 3x=6, she might generalize to "divide by the first number." If the student tells Stacy that a step is incorrect, she uses inductive logic to determine constraints to only use the rule in appropriate situations - e.g., if a new negative example says that it is incorrect to divide by 2 for 2x+3=5, she might add the constraint "when the left hand side does not have a constant term".

Throughout the interaction Stacy also asks the tutor other questions, such as, “Why should I do this problem?”, or “I did [x] before, why can’t I do that here?” These questions are intended to provoke reflection and self-explanation in the tutor. Students can select an answer from a drop-down box or can type in their own explanation.

At any point in time, tutors can have Stacy take a quiz on the material. Tutors can use this quiz both to test that she has acquired the knowledge they have taught, as well as to understand where her misconceptions lie. As she passes sections of the quiz, new problem types appear that give tutors an indication of the domain rules they should be working on next with Stacy.

While Stacy is embodied, her image is not articulated and has a cartoon-like appearance. She is modeled using an agent creation system that mimics characters on the Nintendo Wii system, and has three poses: a standard pose, a questioning pose in which she appears to be thinking, and a happy pose that is seen when the tutor marks a step correct.

STUDY

Participants

12 students (2 girls and 10 boys), ranging from entering 7th grade to entering 10th grade, were recruited from an e-mail list of parents who had previously indicated interest in research participation. All students reported experience with algebra. Students came for two 90-minute sessions, and were compensated $40 at the completion of both sessions.

Equipment

Students sat at a desktop computer, with the SimStudent interface pre-loaded. A chart showing twelve classrooms labeled with different school grades was taped to the wall on their left. A digital video camera recorded participants on their right side, and captured their position in the chair, the grade chart behind them, and part of the screen. They were provided scratch paper and were invited to use it, though only some did.

Procedure

In the first session, students first took a pre-test which consisted of algebra problems. Once completed, students were asked to look at an image of Stacy, and place a post-it on the grade chart beside them to indicate what grade they thought she was in. They were told they could move the post-it to update their choice at any time. Next, students watched an 8-minute video describing Stacy, and were given instructions on how to think-out-loud during a study. Once finished, students began working with Stacy, and were reminded to speak out loud whenever they became quiet. Students were told that their goal was to help Stacy learn how to solve equations with variables on both sides to help her pass four sections of a quiz.

In the second session, students immediately began working with Stacy and continued the think-aloud protocol. They worked until either Stacy passed all four quiz sections, or 45 minutes had passed. They then completed the post-test, which took anywhere from 10-35 minutes. They were asked some final interview questions, and were compensated for their time.

LEARNING GAINS

Before we investigated students’ behaviors in the tutoring sessions, we calculated pretest and posttest scores to assess their learning gains over the course of the intervention, which were significant ($t = 2.84$, $p < .02$, effect size 0.56σ). Significance was calculated using a student’s paired t-test across each student’s pretest and posttest scores. We also collected the following demographic data for each participant: school grade, gender, and previous tutoring experience; however, none of these variables were significantly correlated with learning gains.

We then computed normalized learning gains using the standard formulation to account for differences in children’s prior knowledge:

$$normalized\ gain = \frac{posttest - pretest}{1 - pretest}$$

Normalized gain is used in our subsequent results sections to explore relations to learning.

OBSERVATIONS FROM THE VERBAL PROTOCOL

Observationally, students made comments during the think-aloud in three primary categories: social moves, tutoring strategies, and cognitive evaluations of Stacy’s knowledge. In this section, we discuss notable examples of social and tutoring utterances.

Social Moves

Students made social comments to Stacy with varying frequency, with one student never saying anything social at all, and some making a social move every ten utterances. Positive social moves were common, particularly on day one, including compliments:

Stacy: Would this be a good move?
P12: Yes! You’re a smart person.

Congratulatory praise: *P12: You got it, Stacy. Congratulations! Let's try the quiz again.*

Reassurance: *P10: You're almost there Stacy! Oh, Stacy. You were so close.*

Empathy: *P10: Negative 8? Oh dear. You didn't like that one, did you?*

Not all social moves were so clearly positive, however. We observed many comments that could be characterized as *face-threatening*, by which is meant dialogue moves that threaten the other person's identity management, or positive sense of him or herself [4]. Examples were students playfully insulting and teasing Stacy throughout the conversation, particularly during the second session. This included students minimizing her successes:

P7: You got lucky Stacy.
P8: Problem is solved, no thanks to you.
P7: Yes Stacy listened for once.

Being overtly face-threatening:
P7: Oh God. You fail Stacy. Oh God.
P8: That's terribly not right.

Using sarcasm:
Stacy: I think the problem is solved.
P10: Really? Well I don't.

Stacy: I'm stuck and I don't know what to do next. Can you show me what to do?
P10: Hmm, really? Yeah, I've heard that one before.

Expressing frustration:
P12: Argh, you annoy me so much.
P2: Stacy, what are you doing?!

These utterances were typically said with a playful tone, and it's important to note that, qualitatively, the students did not seem to be harboring actual frustration, annoyance, anger, or any strong negative emotion. Instead, they seemed to engage in the teasing one would observe among friends.

Tutoring

Most students, for at least part of their session, took their role as tutor seriously. All students made comments at some point about what they should do to make sure Stacy was learning, with various degrees of serious analysis about her knowledge. Participant 4 made many of these formal tutoring moves, though very few social moves:

P4: All right, so Stacy got the first problem right, and didn't have a clue how to do the next problem.
P4: I don't think she knows how to deal with parentheses.
P4: All right, so I'm just going to use their example problem now.
Stacy: Why should I do this problem?
P4: You should do this problem because you got this problem wrong, and I want to see what you did not understand about it.

This utterance also depicts a common phenomenon in our data, which we call face-saving alignment. Students who made many formal tutoring moves and few social moves often used outside-aligned speech to discuss what Stacy did and did not know, which we hypothesize is because it would be face-threatening to discuss her incompetencies with her in detail, along the lines described by Reeves and Nass [15]. When Stacy prompts the participant back into inside-aligned speech with a question, the child does provide an answer, though it has less specific elaborations about what exactly they thought Stacy was doing wrong. The students who made fewer social moves such as teasing may not have felt the same sense of rapport that the social students felt, and might not have been comfortable being face-threatening with Stacy the way the social students did.

DATA ANNOTATIONS

In order to adduce evidence for the hypotheses we lay out above, we analyzed the linguistic behaviors of the children in our corpus, based on annotation of the think-aloud protocols in the way described below.

Coding Scheme

Each think-aloud session in our data set was divided into utterances by a human annotator, based on pauses in speech and thought completeness. Our data consists of 3,433 utterances from 12 participants over 20 sessions (four participants finished the tutoring task in their first session, returning only for the posttest on day two).

These utterances were then coded in five categories developed to evaluate our hypotheses. The coding was carried out by two independent coders who first evaluated interrater reliability by independently coding a random child's full dialogue. Reliability is given for each coding category below in a Cohen's K [7]. Every utterance was given a code from every category, with *none* always as an option in the case that the utterance contained no features for that category. Our five coding categories and their sub-categories are described below.

1. A *social* utterance, either *positive* that represents feelings including hope, encouragement, or excitement (e.g. "yay, dear Stacy, you can do it!") or *negative* expressing face-threat or frustration (e.g. "got this one right, no thanks to you.") (interrater reliability Cohen's K = .773)
2. A *tutoring* move that included conceptualizations of Stacy's knowledge and informed decisions about how to proceed (e.g. "now I'll give you an example fractions problem because you got that one wrong last time"), and elaborations about domain material (e.g. "now you need to divide to make sure the variable is alone on that side of the equation.") (Cohen's K = .686)
3. An *alignment* based on pronoun use, including inside-system alignments such as *you* and *inclusive-we* (e.g. "you got this one wrong, Stacy, we should do a new one

now"), and outside-system alignments such as *she, Stacy*, and *exclusive-we* (e.g. "Stacy doesn't know fractions, we'll see if she can do this one now.") (Cohen's K=.823)

4. A *cognitive* assessment about Stacy's knowledge that was either *simple* (e.g. "She gets this one") or *elaborated* (e.g. "Okay, Stacy doesn't understand the distributive part."). These elaborations are a hypothesized mechanism for learning, as described earlier in the paper. (Cohen's K=.823)
5. A *correctness* evaluation of Stacy's knowledge as being either *correct* or *incorrect.* (Cohen's K = .707)

Analysis Methods

Our quantitative experiments are framed around exploring the effects and interactions of coded language behaviors in the five categories described above. In the upcoming sections, we report our results from several quantitative experiments that explore:

1. Correlations between language behaviors, learning gains, and participant attributes to examine how participant features and behaviors are associated with agent interaction and learning.
2. Shifts in behavior between sessions 1 and 2 to better understand how increased exposure to the system affects language behaviors.
3. How specific linguistic behaviors in the child affect upcoming child alignment on a turn-by-turn level using a novel machine learning approach.

DIALOGUE BEHAVIORS AND LEARNING GAINS

We first investigated children's learning gains as they related to the relative frequencies of their dialogue moves.

For each child *c* in the set of children *C*, we calculate probabilities that an utterance will have a label *l* for label category *L* (where the labels represent the language behaviors we coded). This results in a value $P(L = l \mid C = c)$ for each possible pairing of category label and child. For each category, these probabilities sum to 1 for each student. For each variable individually, we perform a linear regression to fit normalized learning gain and evaluate significance of the regression using a one-way ANOVA test. Labels with a statistically significant relationship to learning gain are summarized in Table 1.

Annotation Label	Gain r^2	Sig.
Alignment: Outside	-.510	**
Tutoring	-.314	*
Cognitive-elaborated	-.316	*
Social-negative	.646	***

Table 1. Behaviors which explain significant variance in normalized learning gain. Significance marked as * (p < .05), ** (p < .01), * (p < .001)**

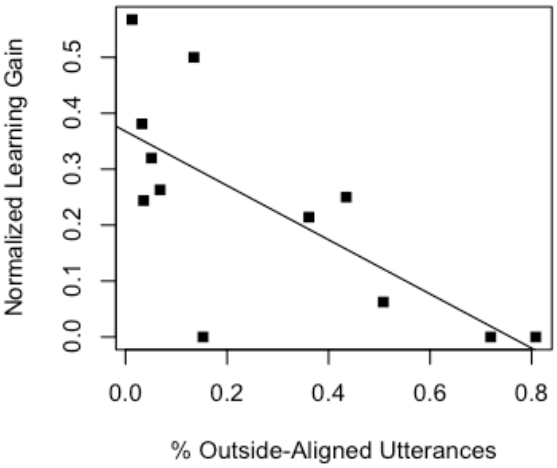


Figure 2. Correlation of relative percentage of outside-aligned utterances to normalized learning gains

We found a significant positive correlation between negative social moves, such as face-threatening or teasing comments, and student learning gains.

We also found significant negative correlations between learning gains and three specific student behaviors: (1) aligning outside of the system by talking *about* Stacy rather than *to* Stacy, (2) describing very formal tutoring moves such as stating what they planned to do, why they were doing it, and what they hoped it would achieve, and (3) giving elaborate cognitive assessments about Stacy's understanding of the material, such as explaining what exactly it is, in detail, that Stacy knows or doesn't know.

None of the other dialogue behaviors were significantly correlated to learning, nor did they demonstrate any significant correlation to any of the demographics collected.

Cognitive elaboration and outside alignment

Because previous human-human peer tutoring research reports that increased elaboration is associated with learning gains [22], we conducted further analysis to understand why elaborations were negatively correlated with learning in our study. We found that there was a significant correlation between cognitive elaboration and outside-system alignment, and that it is the outside-system alignment that's driving the significance (see the following section.) When we controlled for alignment, the effects of cognitive elaboration on learning gains were not significant, though they still trended negatively, p<.4. Closer examination found that 39% of utterances that involved elaborations were preceded by inside-aligned student moves, indicating that students were breaking away from alignment with Stacy to explain her cognitive state in the third person.

ALIGNMENT SHIFT PREDICTION

Overall, we found that shifting away from direct communication with the agent and instead talking *about* Stacy in the third person was more negatively correlated with learning than any other annotated student behavior.

In the next experiment, we attempt to *predict shifts* in child alignment within a single session based on our coding of the think-aloud utterances. This gives us insight into what behaviors are most likely to indicate a child's upcoming break in rapport (shifting to outside-alignment) so we know how to address this issue in the design of future learning interventions.

We treat this as a three-way classification task. Machine learning was performed using the SIDE text mining toolkit [13]. We built a Naïve Bayes model and evaluated accuracy through leave-one-child-out cross validation, where eleven students are used to train each model, and that model is tested against the sessions from the held-out child. This is done twelve times and the accuracy is averaged.

In a given model, we choose a window size of *n* utterances Then we define two features for each category label: a numeric feature marking how many times this feature occured in the previous *n* utterances, and a boolean feature marking whether the label was observed at all in the window. We then add one additional feature marking whether Stacy made an incorrect move since the most recent utterance.

Through cross-validation, we find window size *n*=6 to be the most predictive. This model has 70.9% accuracy in predicting upcoming alignment overall (predicting outside alignment with precision = .611, recall = .745, κ=.554).

We observe several interesting characteristics of the model. Relatively little adjustment is made based on observing a behavior once in the window. Observing a behavior two or more times, however, has a more drastic effect. For instance, one negative social move in the recent window drops the probability only slightly, while two or more drop the probability near zero. On the other hand, elaborated cognitive moves have a similar pattern in the opposite direction, with a slight increase in likelihood when only one is observed, but virtual certainty of outside alignment if two or more elaborated moves have been observed in the last six utterances. What this suggests is that individual moves in any category are possible across a broad range of interactional styles. However, repeating the same style of moves (such as elaborated cognitive moves, or negative social moves) more than once in quick succession gives the model the evidence it needs that recent moves fit into an outside-aligned or inside-aligned pattern.

Qualitatively, it appears that students' alignment often reflects their perceptions of Stacy's ability. At one point in the interaction, participant P4 is inside-aligned, talking to Stacy and telling her what to do:

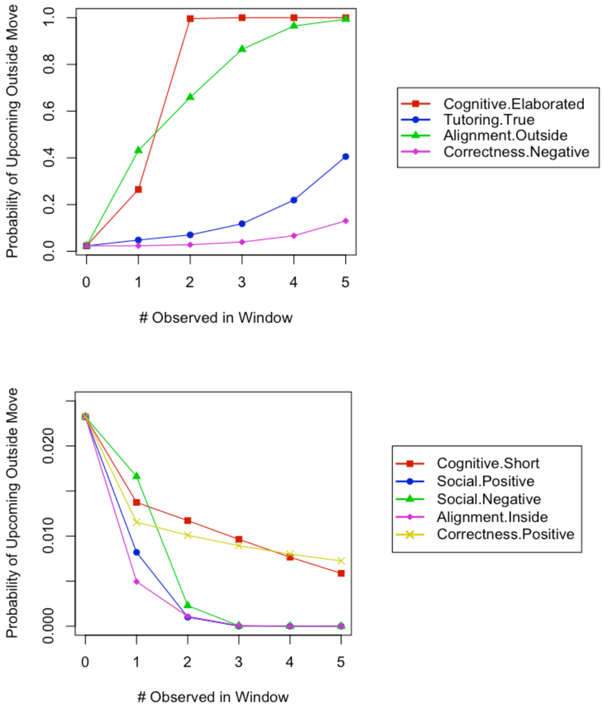


Figure 3. Impact of each feature individually on prediction of upcoming outside-aligned moves (holding all other features constant at 0)

P4: Ok so 10c+3, and your 10c... So you have to div.. If 10c=3 you have to divide by 10.

When he then evaluates her subsequent performance as poor, he switches alignment to outside the system, talking to the experimenter (and demoting her in grade level):

P4: And I'm going to move her back to about 7th grade cause she can't solve this.

Although we did not make a distinction between calling Stacy 'she' and 'it' in our coding scheme due to the infrequent occurrences of 'it', we believe that a shift between the use of one of these pronouns to the other might, in a larger data set, have just as much significance as a shift from inside to outside alignment. Where the shifts from "she" to "it" occur, they seem to follow a similar pattern to switches between inside ("you") and outside alignment ("she"). For example, participant P2 at one point in the interaction is aligned outside the system, talking about Stacy:

P2: She didn't do problem number 8.

This is followed by two steps which were correct, but the participant wrongly evaluated as incorrect:

P2: No that's not right...No.

Stacy then takes five correct steps followed by two incorrect steps, at which point the participant dehumanizes her in his speech:

P2: Uh…. Um… I don't think ***it*** *knows how to distribute things.*

In addition to predicting shifts in alignment, we were also interested in discovering what behaviors were most predictive of upcoming social moves, given that these moves were such a strong predictive feature for alignment, and were also associated with learning gains. We attempted to replicate this experiment to predict upcoming social moves; however, accuracy above chance was much weaker (K < .2). This suggests that the factors that are responsible for influencing children's use of social moves are likely outside of our coding scheme, and further research is necessary to determine what these features are and how they play a role.

BEHAVIOR SHIFTS ACROSS SESSIONS

As tutoring interventions ideally happen over an extended period of time, and we know that language changes among human interlocutors as they become more familiar with one another, we examined differences in the children's linguistic behavior between sessions 1 and 2. In our study eight participants returned to engage with the tutoring intervention on a second day. Investigating only these participants, we had 8 sessions comprised of 1,164 utterances that represent the first interaction a participant has with the agent, and eight sessions comprised of 1,387 utterances representing the second session on a later day. Quantitative measures of shifts in each annotation are given in Table 2. Significance is calculated using unpaired student's t-tests between utterance distributions in each day.

On day 2 of the intervention, we saw that while positive social moves decreased, negative social moves significantly increased. Additionally, simple cognitive evaluations increased, as did both negative and positive statements about Stacy's correctness on the task.

DISCUSSION

A number of authors have posited that human peer tutoring is successful because of the increased elaboration of material that is necessitated by interaction between tutor and tutee [21, 22]. However our results demonstrated the opposite effect – that increased elaboration and reflection resulted in fewer learning gains. Further exploration found that the driving force behind this effect is that elaborations were strongly correlated with outside-aligned speech, which in turn was strongly associated with negative learning gains. We also saw that 39% of the time, students switched from inside to outside-aligned speech for the purpose of this elaboration, taking themselves out of a role as Stacy's partner and assuming the role of Stacy's observer.

We hypothesize that this switch to outside-aligned speech occurs because students who don't feel comfortable with Stacy – those with fewer social behaviors – don't want to offend Stacy by directing their detailed assessment of her incompetencies to her. We believe these students are doing face-saving alignment – an unconscious switch from inside to outside-aligned speech when they want to elaborate about Stacy's abilities. We also hypothesize that this switch away from partnership and alignment with Stacy removes some of the social motivations of peer tutoring, and that making the kind of useful tutoring moves that consist of elaborating information about domain material, but without the scaffolded support of the system, make it difficult for the child to maintain an effective peer relationship with the agent.

Annotation	Day 1	Day 2	Shift %	Sig.
- Social	.024	.050	108.3	***
+ Social	.056	.037	-34.5	**
- Correctness	.109	.173	58.9	***
+ Correctness	.226	.291	28.7	***
Cog Simple	.305	.439	43.9	***
Cog Elab	.024	.025	4.2	NS
Inside Align	.092	.112	21.7	NS
Outside Align	.215	.241	+12.3%	NS
Tutoring	.137	.119	-13.4	NS

Table 2. Distributions of behavior in separate days, and normalized shift in those behaviors. Significance marked as * (p < .05), ** (p < .01), * (p < .001), or Not Significant.**

Because previous research has shown that elaboration and self-explanation is so beneficial to learning, we propose that learning by teaching interventions should have provisions to allow students to elaborate on knowledge as part of a joint activity with the agent, to discourage them from disaligning with the system. Stacy has very limited social moves, and is only able to ask a handful of open-ended questions designed to prompt the student to elaborate. However, Stacy could not respond to these elaborations, nor could she encourage the child to elaborate while remaining *inside* the system. It's possible that children picked up on Stacy's inability to respond, which is what lead them to disalign. It is also possible that if Stacy were able to scaffold how one could reflect and elaborate while continuing to co-construct knowledge with the child, the participants would have been able to follow suit. Future work will examine different ways of encouraging inside-aligned speech during reflection and elaboration.

In addition, we found that tutoring moves were negatively correlated with learning. Based on qualitative observation of students' verbalizations, these moves were also highly formal speech like their elaborations, and may also indicate that in these turns, students were playing the role of tutor and not socially engaging with the tutee.

These findings highlight the importance of role in tutoring, with students who speak to Stacy as a peer rather than a socially-distant tutee achieving the highest learning gains. We found that negative social moves, such as teasing and face-threat, were the most predictive of learning gains. These moves are also indicative of rapport between interlocutors, and are thought to mostly take place between intimates [24]. In fact, the literature on social moves among middle-schoolers makes it clear that alliance building is not confined to supportive behaviors. Episodes of playful confrontation, name-calling, and insulting sequences are prominent in middle-school communication [1]. It is notable that students who produce many of these utterances achieve the highest learning gains, indicating that in human-agent peer tutoring, the social role of the child is vital, and that a strict tutor-role division may not be the most beneficial. We suggest that systems should support rapport-building dialogue – including what may appear at first to be agent abuse.

We found that the frequency of these intimate teasing comments increased on students' second session, indicating that students may naturally gain more rapport with the agent over time. In fact, according to the theory of rapport proposed by [25], the importance of *positivity* in a relationship decreases over time, which may indicate that students felt more comfortable with the agent and thus had less need for positive statements. Upon resuming the second day, several students made explicit social opening statements:

P12: Ok. Let's try it again, Stacy.
P7: Stacy, come back for you!

However, our investigation found that it was not possible to predict when these social moves would occur from the features in our coding scheme. In future work we will examine how we can support critical rapport-building in this context by looking at what behaviors lead children to tease and in other ways ally themselves with a teachable agent or real human tutee.

We also saw that agent mistakes are an important factor in the likelihood that children will shift into outside-aligned speech. Making errors is realistic, and there is evidence that in human-human peer tutoring, these are the places where tutors do the most learning [27]. Instead of trying to create perfectly competent agents, we propose designing agents that are able to acknowledge their errors socially, in particular after committing several contiguous errors. Additionally, we found that the child's assessment of the agent's correctness was a greater predictor of alignment-shift than actual correctness. We propose that agents should keep the participant immersed in the experience by making teasing or joking face-threatening moves of their own following an incorrect assessment of their ability. In future work we will observe human-human pairs and their strategies for defusing these situations, and extract social moves following errors that lead to more learning.

We hypothesize that systems designed with agents that can interact socially with the child could prevent children from facing the identity crises we saw children experience during the course of our study. Stacy's limited social interactions but realistic learning patterns may have confused students, with some indicating they weren't sure about how to interact with Stacy – as an agent, a machine, a peer, a tutee. One participant's utterance sums up this conflict precisely:

P13: I'm mad at the computer because it's not – I'm mad at Stacy because she's not understanding what I'm saying. [pause] But I'm holding it in 'cause it's not nice to be mean to your students… even though this isn't really a student.

Finally, we do see that significant learning gains were achieved with our program, which adds to the evidence that a learning-by-teaching paradigm is successful with agents.

Of course, it's important to note that our data were derived from the kind of think-aloud protocol that would likely not occur outside of a lab study. In a paradigm where students typed directly to the agent rather than speaking aloud, such alignment shifts may not have occurred. While we believe our results are insightful, we acknowledge they may be different from how students would speak to an agent during full dialogues via chat. Think-alouds also encourage the child to verbalize what they might have otherwise kept to themselves, perhaps artificially encouraging less social children to resort to face-saving alignment that wouldn't have been necessary with a different methodology.

Additionally, we emphasize that these results are not causal. It may be the case that feeling self-efficacious about learning leads to gaining more rapport with the agent, rather than the increased rapport driving learning. It also may be the case that it is not the switch to outside-aligned speech that is causing negative learning gains, but a third factor such as frustration that is causing the students to dis-align as well as learn less. It is also important to note that this work only describes interactions with the agent in a procedural domain. Students may use different behaviors in other creativity-based or declarative learning domains. Our future work aims to address these limitations.

CONCLUSION

This work presents the most thorough analysis to date of social interaction with teachable agents. This is the first work that looks at the association between learning gains, social moves, and tutoring/elaboration moves with an embodied virtual peer, and it is thus notable that the human-human results are not mirrored in this work.

Children who acted as though the teachable agent was in the room, who spoke directly to her and engaged her in conversation (even though she rarely replied!) were more successful in the learning task. This was particularly true for students who produced the most teasing and face-threatening utterances.

Based on this research, we recommend designing systems that are able to provide better social support for (1) scaffolding inside-aligned elaboration, (2) modeling appropriate elaborations within a peer-tutoring context, and (3) encouraging inside-alignment in response to agent errors through increased social dialogue.

ACKNOWLEDGEMENTS

The research reported here was supported by National Science Foundation Award No. DRL-0910176 and the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090519 to Carnegie Mellon University. This work is also supported in part by the Pittsburgh Science of Learning Center, which is funded by the National Science Foundation Award No. SBE-0836012. The opinions expressed are those of the authors and do not represent views of the funders.

REFERENCES

1. Ardington, A. Playfully negotiated activity in girls' talk, Journal of Pragmatics, Volume 38, Issue 1, January 2006, 73-95.
2. Bargh, H., Schul, Y. On the cognitive benefits of teaching. *Journal of Educational Psychology*, 72(5), 593-604. 1980.
3. Biswas, G., Leelawong, K., Schwartz, D., Vye, N., TAG-V. (2005). Learning By Teaching: A New Agent Paradigm for Educational Software. Applied Artificial Intelligence, 19, 363-392.
4. Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. New York: Cambridge University Press.
5. Cassell, J. (2004). Towards a Model of Technology and Literacy Development: Story Listening Systems. Journal of Applied Developmental Psychology, 25(1), 75-105.
6. Chase, C., Chin, D., Oppezzo, M., Schwartz, D.: Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. J. Sci. Educ. Technol. 18, 334--352 (2009)
7. Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 20(1). 1960.
8. Cohen, P., Kulik, J., Kulik, C. Educational Outcomes of Tutoring: A Meta-analysis of Findings. Journal of Educational Research. 19(2). 237-248. 1982.
9. Coleman, E., Brown, A., Rivkin, A. The effect of instructional explanations on learning from science texts. Journal of the Learning Science, 6(4), 347-365. 1997
10. Fantuzzo, J. W., King, J., Heller, L.R. Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. Journal of Educational Psychology, 84(3), 331-339. 1992.
11. Gulz, A., Silvervarg, A. & Sjoden, B. Design for off-task interaction – Rethinking pedagogy in technology enhanced learning. Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies. 2010.
12. Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G., Cohen, W. W., et al. (2011). Learning by Teaching SimStudent – An Initial Classroom Baseline Study comparing with Cognitive Tutor. Proc AIED (213-221): Springer.
13. Mayfield, E. and Rosé, C.P. An interactive tool for supporting error analysis for text mining. In Demo Session for the North American Association for Computational Linguistics. 2010.
14. Okita, S., Bailenson, J., and Schwartz, D.L. 2008. Mere belief in social action improves complex learning. In Proceedings of the 8th international conference on International conference for the learning sciences. 2. 132-139.
15. Reeves, B., & Nass, C. (1996). The Media Equation: how people treat computers, televisions and new media like real people and places. Cambridge: Cambridge University Press.
16. Reif, F. and Scott, L.A. Teaching scientific thinking skills: Students and computers coaching each other. American Journal of Physics. 67(9), 819-831. 1999.
17. Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., Miller, T. R. Peer-assisted learning interventions with elementary school students: A meta-analytic review. Journal of Educational Society. 95(2), 240-257. 2003.
18. Roscoe, R.D., Chi, M. Understanding Tutor Learning: Knowledge-Building and Knowledge-Telling in Peer Tutors' Explanations and Questions. Review of Educational Research. 77(4), 534-574. 2007.
19. Rose, C.P., Bhembe, D., Siler, S., Srivastava, R., VanLehn, K. The Role of Why Questions in Effective Human Tutoring. In Proceedings of AI-ED, IOS Press. 2003.
20. Rose, C.P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., & Weinstein, A. Interactive conceptual tutoring in Atlas-Andes. In Proceedings of AI-ED. IOS Press, 256-266. 2001.
21. Rosé, C. P., & Torrey, C. (2005). Interactivity versus Expectation: Eliciting Learning Oriented Behavior with Tutorial Dialogue Systems, Proc. Interact '05
22. Sharpley, A., Irvine, J., Sharpley, C. An Examination of the Effectiveness of a Cross-age Tutoring Program in Mathematics for Elementary School Children. American Educational Research Journal. 20(1), 103-111. 1983.
23. Slavin, R. E. Research on cooperative learning and achievement: What we know, what we need to know. Contemporary Educational Psychology, 21, 43-69. 1996.
24. Straehle, C. A. (1993) '"Samuel?" "Yes dear?" Teasing and conversational rapport', in D. Tannen, ed., Framing in Discourse. New York: Open University Press.
25. Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. Psychological Inquiry, 1, 285-293.

26. VanLehn, K. The behavior of tutoring systems. International Journal of Artificial Intelligence in Education. 16(3), 227-265. 2006.
27. Walker, E., Rummel, N., & Koedinger, K. R. Integrating collaboration and intelligent tutoring data in the evaluation of a reciprocal peer tutoring environment. Research and Practice in Technology Enhanced Learning, 4(3), 221-251. 2009.
28. Webb, N. M. Peer interaction and learning in small groups. Journal of Educational Psychology. 13(1), 21-39. 1989.