Socially-capable Conversational Tutors can be Effective in Collaborative-Learning situations

Rohit Kumar[†], Hua Ai[†], Jack L. Beuth[‡], Carolyn P. Rosé[†]

[†]Language Technologies Institute, [‡]Department of Mechanical Engineering,
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213
{ rohitk , huaai , beuth , cp3a } @ andrew . cmu . edu

Abstract. Tutorial Dialog has been shown to be effective in supporting both individual as well as group learners. However, unlike the case with individual learners, teams of learners often ignore and abuse the automated tutors. Both theory and empirical work in the area of small-group communication argue that group participants display both task as well as socio-emotional behaviors during interactions. However, in connection with automated conversational agents, the effects of socio-emotional behaviors are much less well understood, especially in the case of multi-party interactions. In this paper, we will describe a socially-capable conversational tutor that supports teams of three (or more) learners in a design task. Further, this tutor is evaluated in comparison with a socially-neutral baseline agent and human capability “gold standard” tutors. Results show that our socially-capable tutors can be effective support, even though there is still room for improvement to match the human gold standard.

Keywords: social interaction, tutorial dialog, conversational agents, collaborative learning, small-group communication

1 Introduction

Conversational Agents are autonomous interfaces that interact with users via spoken or written conversation. Automated tutoring is a widely studied application of such agents. Various research groups have developed conversational tutors for a variety of educational domains including algebra, calculus, computer literacy, engineering, foreign languages, geometry, physics, reading and research methods. Many evaluations show that these tutors can be effective support for learners [1][2][3].

While most of the work on conversational tutors has focused on one-on-one tutoring involving only one learner, use of such tutors in collaborative learning situations involving two or more human students has been investigated. Our previous work [2] has shown that tutors in a collaborative learning situation can lead to over one grade improvement. Other work [4][5][6][7] has explored a variety of interaction pattern / tactics that could be used in multi-party educational situations.

However, despite the effective support that automated tutors offers to students learning in groups, it has been reported that groups of students often ignore and abuse the tutor, unlike the case where students are individually tutored [2][8]. We reason that the presence of other students in collaborative learning scenarios causes the tutors to compete for the attention of the students. Since the tutors are not capable of initiation or participating in social interaction which makes up the bulk of formative interaction in the group, they are pushed to the periphery of the learning group.

Research in the area of small group communication has shown that humans employ both task-related as well as social interaction strategies while interacting in groups. However, research on conversational tutors has focused on presenting only task-related information, i.e., lessons and instructions in case of tutors. In this paper we report the first study in our investigation on the effects that conversational agents in general can achieve if they are equipped with social conversational skills.

The rest of the paper is organized as follows: In the next section, we motivate social interaction strategies for agents based on relevant literature from small group communication research. Section 3 describes our flexible architecture and implementation details for a tutor with social conversational skills. Results from the evaluation of the tutor against a baseline and a human gold standard are presented in section 4 before conclusion.

2 Small Group Communication

Theoretical and empirical study of group interaction processes has been of interest in sociology and communications research communities since the 1950's. McGrath [9] reviews various theories that address the functions of group interaction processes. Of particular interest among these are the theories proposed by Robert F. Bales [10] and Wilfred R. Bion [11]. Both of these theories propose that two fundamental processes operate within groups i.e. instrumental (task-related) vs. expressive (social-emotional) in the case of Bales and work vs. emotion in the case of Bion. Over attention on one of these processes causes lapses on the other. Hence, interaction shifts between these two in order to keep the group functional.

In the case of conversational tutors, the task (or work) related interaction include aspects like instructing students about the task, delivering appropriate interventions in suitable form (e.g. socratic dialog, hints), providing feedback and other such tactics [12]. Some studies [13] [14] have evaluated the effect of these task-related conversational behavior in tutorial dialog scenarios. Work in the area of affective computing and its application to tutorial dialog has focused on identification of student's emotional states [15] and using those to improve choice of task-related behavior by tutors. However, there has been only limited study of expressive (social-emotional) aspects of the tutor's conversations with learning groups. Besides focusing on expressive behavior of the tutor, the novelty of this work lies in the use of small group communication research in designing tutor behavior.

2.1 Social Behavior for Conversational Tutors

As discussed earlier, current state-of-the-art conversational tutors are incapable of performing the social-emotional function of interaction which is known to be a fundamental aspect of group interaction. Hence, we hypothesize that socially-capable tutors will be able to perform better in collaborative learning scenarios. In order to further specify social capability, we use the interaction process analysis (IPA) schema developed by Bales [19]. Beside the influence and popularity of IPA over the last five decades, our choice is based on the unit of analysis at which IPA is applied which is individual utterances compared to Bion's units of analysis (sessions) which are typically much larger (10-50 utterances).

IPA identifies three positive social-emotional interaction categories: showing solidarity, showing tension release and agreeing. We have mapped these categories to practically implementable conversational strategies, which are distinguishable from each other and are relevant to interactive situation employed in our experiment. This mapping is shown in Table 1 below.

Table 1. Social Interaction Strategies based on three of Bales' Socio-Emotional Interaction Categories

1. Showing Solidarity: *Raises other's status, gives help, reward*
1a. Do Introductions: *Introduce and ask names of all participants*
1b. Be Protective & Nurturing: *Discourage teasing*
1c. Give Re-assurance: *When student is discontent, asking for help*
1d. Complement / Praise: *To acknowledge student contributions*
1e. Encourage: *When group or members are inactive*
1f. Conclude Socially
2. Showing Tension Release: *Jokes, laughs, shows satisfaction*
2a. Expression of feeling better: *After periods of tension, work pressure*
2b. Be cheerful
2c. Express enthusiasm, elation, satisfaction: *On completing significant task steps*
3. Agreeing: *Shows passive acceptance, understands, concurs, complies*
3a. Show attention: *To student ideas as encouragement*
3b. Show comprehension / approval: *To student opinions and orientations*

Each strategy is implemented as an instantiation of a conversational behavior. Most of these strategies are realized as prompts, triggered by rules based on agent plan, discourse and context features. For example, strategy 1e is triggered when one or more students in the group are found to be inactive for over 5 minutes. In this event, the tutor chooses to raise the status of the inactive students by eliciting contributions from them through a prompt like: *Do you have any suggestions Mike?* We did a pilot evaluation with 6 subjects to verify if the strategies were perceived as we expected.

3 Implementation of Socially Capable Tutors

The interaction between the students and tutor in the experiment presented in this paper is situated in a freshmen engineering course. In this course college students learn about basic mechanical engineering concepts like force, moment, stress, etc. The students interact with the tutor as part of a computer-aided design competition where the students are asked to design a better wrench with consideration to ease of use, safety and material cost. Students could interact with each other and the tutor using a text-based chat room which includes a shared whiteboard [16].

Table 2. Excerpt of a tutor providing a lesson to a team of four students

	Speaker	Contribution
64	Tutor	Intuitively, if you wanted to make the wrench easier to use, would you make it longer or shorter?
67	S5	Longer
70	Tutor	That's right. A longer wrench is better.
72	Tutor	Why is a longer wrench easier to use though! Let's look at the concept of Moment.
73	Tutor	When you use a wrench to turn a bolt, do you want a higher or lower moment?
74	S16	Higher

The task of the tutor is to provide lessons on the underlying theoretical concepts while the students work through a worksheet to explore various design choices. An excerpt of a lesson on the concept about the relationship between the length of a wrench and its ease of use is shown in Table 2 above. Besides performing its task related functions, the tutor also employs the social interaction strategies listed in the previous section. We have implemented this tutor using the Basilica architecture [17].

3.1 Basilica

Using the Basilica architecture, conversational agents are modeled as a network of behavioral components. Each component implements a behavior which could be a combination of perception, thought and action. There are three types of components: actors (actuators / performers), filters (perceptors / annotators / cordinators) and memories. Each component can generate events carrying signals and data. Connected components can receive and process the events and generate further events.

This architecture allows a developer to build agents by adding behavioral components incrementally. Since each component is not tightly coupled to all others, it provides the flexibility to easily change a single behavior. Also, it allows components to be reused between agent/tutor implementations for different tasks. Further, each component is fully programmable and not restricted to a small set of acts and operators, as is the case with most other dialog/conversational system architectures. This makes Basilica a suitable choice for an architecture to build agents with novel rich behavior like the social behavior we investigate here.

3.2 Avis: Tutor Implementation using Basilica

The tutor agent developed for the freshmen mechanical engineering learning domain is comprised of 21 Basilica components: four actors, fifteen filters and two memories. The tutor uses a a gender neutral name Avis. Figure 1 below shows a simplified depiction of component network of Avis. Actor components are shown as circles and filter components are shown as rectangles. Arrows depict connections and possible directions of event flow.

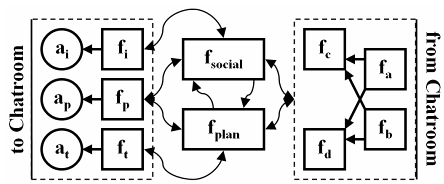

Fig. 1. Basilica component network for Avis, the freshmen mechanical engineering tutor

Three of the actor and filter components correspond to three observable behaviors of the tutor, i.e., Introducing (a_i, f_i), Prompting/Hinting (a_p, f_p) and Tutoring about concepts (a_t, f_t). Most of the other filters (depicted as f_a, f_b, f_c, f_d here) form a sub-network that processes student turns from the chatroom, as shown by filters f_a through f_d in Figure 1. This sub-network annotates turns with applicable semantic categories, accumulates them to identify inactive students and generates events which regulate the cordinators (f_{social} and f_{plan}).

The plan controller (f_{plan}) is responsible for executing the tutor's task-related interaction plan, which is comprised of 37 steps. The plan is executed largely sequentially, however the plan controller can choose to skip some steps in the interest of time. In the experiment described in the next section, the same plan controller is used in all three conditions. The social controller (f_{social}) implements the eleven social interaction strategies listed earlier. The strategies are triggered by rules based on combinations of three features: the most recent plan step, semantic category labels associated with the most recent student turns and the percentage of tutor turns generated by f_{social} (Social Ratio). The first two features attempt to ensure that social behavior is situable in the current conversational context and the third feature regulates the amount of social behavior by the tutor. Once the controllers determine a step or a strategy that is to be generated, the actors generate them. For example, strategy 1a is triggered by the social controller after the first step of the plan. The introduction components (a_i, f_i) greet the students and ask them their names.

The plan and social controllers are connected so that they regulate each other. For instance, when the plan controller is working, it blocks f_{social}. Upon completion of a step, f_{social} is given control, which can then choose to perform a social behavior by blocking f_{plan}. After the social interaction strategy is executed (if any), control is returned to f_{plan} which can then move on to the next step. Instant reflex strategies like 1b and 2b are not blocked. Next, we illustrate how the social and task behavior is intertwined using an example conversation.

Table 3 shows an excerpt of an actual conversation between a team of three students and our tutor agent Avis. Names are changed for preserving anonymity. Turn 142 concludes a lesson about the relationship of stress and ease of use of a wrench. As that plan step is completed, the social controller which was blocked during the lesson gets control and it decides to complement (strategy 1d) one of the students in turn 144 for participating in the discussion. Following that turn, the plan controller is able to proceed to the next step as we see from turn 150 onwards. Also note that one of the students exhibits enthusiasm about their team's designs in turns 143 and 146. The tutor exhibits cheerfulness (strategy 2b) in turn 145 and 147 to reciprocate the student's enthusiasm.

Table 3. Excerpt of an conversation between three students and Avis (Tutor)

	Speaker	Contribution
139	Tutor	One last thing on this topic, Does more (or less) stress in a wrench make it easier to use?
140	S95	no change?
141	Tutor	You are correct about that.. Stress doesn't determine ease of use.
142	Tutor	It's the moment achieved by the wrench that determines the ease of use.
143	S89	yay!
144	Tutor	It's good to have your contributions in the discussion Jackie :-)
145	Tutor	Go team :-)
146	S89	Go team yay
147	Tutor	I am happy to be working with our team
148	S89	Me too
150	Tutor	Unfortunately maximum stress (12800) in our design1 is …

As mentioned earlier, the amount of social behavior the tutor can perform is regulated using a social ratio parameter which is set at 20% for our social tutors. This is based on the percentage of positive socio-emotional turns reported in Bales' empirical experiments with small groups. Note that the social ratio parameter is only the upper limit of the percentage of social turn.

4 Evaluation

The rest of the paper describes the procedures and results from a controlled experiment we conducted to validate our hypothesis mentioned in section 2.1.

4.1 Design

We designed a study with three conditions. In the experimental condition (*Social*), students interacted with a tutor that was equipped with the eleven social interaction strategies, unlike the control condition (*Task*) which is our lower baseline condition. In a third gold standard condition, a human tutor was allowed to perform social interaction while the students interacted with a tutor similar to the *Task* condition. The human tutors were instructed to not give any task-related information/instructions. They were asked to trigger appropriate social prompts (from the same list the automated tutor uses) when they thought it was appropriate. Human tutors were allowed to make modifications to the prompts before triggering them. They were also allowed to type in new prompts.

In all three conditions, students would receive the same task-related information (instructions / lessons / feedback) through the automated tutor. Based on the examples in Table 2, we notice that in the task condition tutor has features that most common tutors do e.g. asking questions, giving feedback, etc. The time allotted for the interaction is the same for each group. The only manipulation in this design is the amount of social interaction which varies from minimal (*Task*) to computationalizable (*Social*) to ideal (*Human*). According to our hypothesis, the *Social* tutors and the *Human* conditions will outperform the *Task* condition.

4.2 Procedure and Outcome Measures

We conducted a between subjects experiment during a college freshmen computer-aided engineering lab project. 98 mechanical engineering students participated in the experiment, which was held over six sessions spread evenly between two days. The two days of the experiment were separated by two weeks. Students were grouped into teams of three to four individuals. Each group communicated using a private chatroom[8]. No two members of the same group sat next to each other during the lab. The groups were evenly distributed between the three conditions in each session.

Table 4. Items about Tutor and Learning Task rated by students on a 7-point Likert Scale

Q1	I liked the tutor very much.
Q2	The tutor was very cordial and friendly during the discussion
Q3	The tutor was providing very good ideas for the discussion
Q4	The tutor kept the discussion at a very comfortable level socially
Q5	The tutor was part of my team
Q6	The tutor received the ideas and suggestions I contributed to the discussion positively
Q7	I am happy with the discussion we had during the design challenge
Q8	My group was successful at meeting the goals of the design challenge
Q9	The design challenge was exciting and I did my best to come up with good designs

Each session started with a follow-along tutorial of computer-aided analysis where the students analyzed a wrench they had designed in a previous lab. A pre-test with 11 questions (7 multiple choice questions and 4 brief explanation questions) was administered after the analysis tutorial. The experimental manipulation happened during the Collaborative Design Competition after the pre-test. Students were asked to work as a team to design a better wrench taking three aspects into consideration: ease of use, material cost and safety. Students were instructed to make three new designs and calculate success measures for each of the three aspects under consideration. They were also told that a tutor will help them with the first and the second designs so that they are well-prepared to do the final design. No additional details about the tutor were given. Besides receiving lab credit for participating in the design competition, students were told that every member of the team that learns the most will receive a $10 gift-card as prize.

After the students spent 35 minutes on the design competition, a post-test was administered. Following the test, student filled out a perception survey. The survey comprised of eighteen items to be rated on a seven point Likert-scale ranging from Strongly Disagree (1) to Strongly Agree (7). Six of the items were based on Burke's survey [18] rephrased to elicit ratings about the tutor's behavior. Three questions were designed to elicit ratings of task satisfaction, satisfaction with group discussion and perceived task legitimacy. These questions are shown in Table 4. The other questions were about group climate and perceptions of other group members.

4.3 Results and Analysis

Learning Outcomes. Using an ANOVA, we saw that there was no significant differences ($p = 0.680$) between pre-test scores for the three conditions (*Task*, *Social*, *Human*). To evaluate the effect of the tutor's social capability on the post-test achievement, we used an ANCOVA model with day of the experiment and the condition as independent variables. Pre-test score was used as a covariate. We found a significant main effect of the condition variable $F(2, 93) = 10.56$, $p < 0.001$. A pairwise Tukey test post-hoc analysis revealed that both the *Human* and *Social* conditions were significantly better than *Task* condition. This is consistent with our hypothesis. The *Social* and *Human* conditions were not significantly different on this measure. The relative effect sizes with respect to the *Task* condition was 0.93 standard deviations (σ) for the *Human* condition and 0.71σ for the *Social* condition. There was no main effect of day of experiment on this outcome.

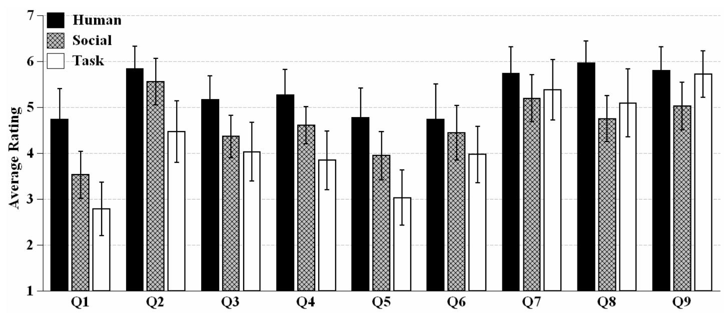

Fig. 2. Average ratings for the Tutor (Q1-Q6) and the Learning Task (Q7-Q9) 95% confidence intervals are also shown

Perception Ratings. Figure 2 shows the average ratings by the students for the survey items. Using condition and day of the experiment as independent variables in an ANOVA, we modeled the ratings for the items. There was a significant main effect of condition ($p < 0.05$) on the first five items. There was no significant difference on the item about tutor agreeing with the students (Q6). Also, there was no main effect of day of experiment on these outcomes. Pairwise Tukey test post-hoc analysis showed the only tutors in the *Human* condition were significantly ($p < 0.05$) better than *Task* condition for the first five questions (Q1-Q5). The tutor in *Social* condition was rated significantly ($p < 0.05$) better only for Q2 (being friendly) and marginally better ($p < 0.08$) for Q5 (being part of the team). The social tutors were not significantly better than our lower baseline (*Task*) on the other four items (Q1, Q3, Q4, Q6).

On the task satisfaction item (Q8) there was significant main effect of both condition $F(2,92) = 4.91$, $p < 0.01$ as well as day of experiment $F(1, 92) = 11.57$, $p <$

0.001. The *Social* condition was the worst of the three conditions on this measure. However, only the difference between *Human* and *Social* conditions was significant. There were no main effects on Q7 and Q9.

4.4 Analysis and Discussion

Table 5. Average number of social behavior turns displayed by tutor

Behavior	Strategy	Social	Human
Doing Introductions	1a	2.67	3.80
Being Friendly	1b-1e	5.61	8.10
Doing Conclusions	1f	0.97	1.80
Trying to Release Tension	2a-2c	5.81	1.77
Agreeing	3a-3b	1.78	4.90
Pushing			0.57
Being Antagonist			1.23

In order to compare our implementation of the social tutors and our gold standard (*Human*), we counted the instances of actual display of social behaviors by those tutors. The turns were classified as one of seven behaviors listed in Table 5 based on the social prompt closest to the turn. Table 5 also shows the average turn counts for the seven types of social behavior for the two types of tutors. All the differences between the tutors shown in Table 5 are significant. We note that except the number of turns related to tension release strategies (2a, 2b, 2c), the human tutors performed significantly more social turns. Also, we note that the human tutors performed additional social behaviors that were not part of the social strategies implemented in our social tutors on some occasions. Both the Pushing and Being Antagonist behavior classify as negative socio-emotional interaction categories in Bales' IPA scheme [10].

5 Conclusion

First and foremost, the study presented in this paper shows that conversational tutors used in collaborative learning scenarios can be improved significantly by making them socially-capable while keeping the task (tutoring) related behavior the same. Specifically, we have shown that a tutor with human-level social capability can achieve a 0.93σ of learning effect compared to a tutor without any social interaction capability. We also see that our upper baseline (*Human*) tutors are perceived significantly better on five out of six items on a survey.

Furthermore, we have described an approach to bridge research in small group communication and conversational tutors using the flexibility provided by the Basilica architecture for developing such interactive agents. The first implementation of a tutor with social interaction capabilities using this approach showed a significant learning effect of 0.71σ compared to the lower baseline. However, on the perception metrics, this implementation of the tutor did not perform significantly better than the *Task* baseline unlike the *Human* gold standard.

Overall, the results presented here show a promise in further pursuing this line of investigation. Several improvements need to be made to our current set of social interaction strategies and their implementation to match human performance both on the performance as well as perception measures to ensure the observed effects can be consistently manifested in deployable conversational tutors. Our next step in that direction is guided by the observation that tutors in the *Human* condition performed many more social interaction turns than our implementation of the social tutors. This suggests that insufficient amount of social behavior performed by our social tutors could be a reason for their inferior perception compared to the human tutors.

References

1. Arnott, E., Hastings, P., Allbritton, D.: Research Methods Tutor:Evaluation of a dialog based tutoring system in the classroom, Behavior Research Methods, vol. 40, pp. 694-698 (2008)
2. Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A.: Tutorial Dialogue as Adaptive Collaborative Learning Support. In: Proc. of AI in Education (2007)
3. Graesser, A. C., Chipman, P., Haynes, B. C., Olney, A.: AutoTutor: An Intelligent Tutoring System with Mixed-initiative Dialog. IEEE Trans. in Education, vol. 48, pp. 612-618 (2005)
4. Liu, Y., Chee, Y. S.: Designing Interaction Models in a Multiparty 3D learning environment. In: Proc. of Intl. Conf. on Computers in Education (2004)
5. Kumar, R., Gweon, G., Joshi, M., Cui, Y., Rosé, C. P.: Supporting students working together on Math with Social Dialogue. Speech and Language Technology in Education (2007)
6. Chaudhuri, S., Kumar, R., Joshi, M., Terrell, E., Higgs, F., Aleven, V., Rosé, C. P.: It's Not Easy Being Green: Supporting Collaborative Green Design Learning. In: Proc. of Intelligent Tutoring Systems (2008)
7. Chaudhuri, S., Kumar, R., Howley, I., Rosé, C. P.: Engaging Collaborative Learners with Helping Agents. In: Proc of AI in Education (2009)
8. Bhatt, K., Evens, M., Argamon, S.: Hedged responses and expressions of affect in human/human and human/computer tutorial interactions. In: Proc. of the Cognitive Science Society (2004)
9. McGrath, J. E.: Groups: Interaction and Performance. Prentice-Hall, NJ (1984)
10. Bales, R.F.: Interaction process analysis: A method for the study of small groups, Addison-Wesley, Cambridge, MA (1950)
11. Bion, W. R.: Experiences in groups: And other papers. Basic Books, New York, NY (1961)
12. Graesser, A. C., Person, N., Harter, D., TRG: Teaching tactics and dialog in AutoTutor. Intl. journal of AI in Education. vol. 12(3), pp. 257-279 (2001)
13. Wang, N., Johnson, L. W.: The Politeness Effect in an intelligent foreign language tutoring system. In: Proc. of Intelligent Tutoring Systems (2008)
14. Rosé, C. P., Moore, J. D., VanLehn, K., Allbritton, D.: A Comparative Evaluation of Socratic versus Didactic Tutoring, In: Proc of Cognitive Sciences Society (2001)
15. D'Mello, S. K., Craig, S. D., Gholson, B., Frankin, S., Picard, R., Graesser, A. C.: Integrating Affect Sensors in an Intelligent Tutoring System. In: Proc. of Workshop on Affective Interactions: The Computer in the Affective Loop at IUI (2005)
16. ConcertChat, http://www.ipsi.fraunhofer.de/concert/index_en.shtml?projects/chat
17. Kumar, R., Rosé, C. P.: Basilica: An architecture for building conversational agents. In: Proc. of NAACL-HLT, Boulder, CO (2009)
18. Burke, P. T.: The development of Task and Social-Emotional Role Differentiation. Sociometry, vol. 30 (4), pp. 379-392 (1967)