Towards Academically Productive Talk Supported by Conversational Agents

Gregory Dyke, David Adamson, Iris Howley, and Carolyn P. Rosé

Carnegie Mellon University, Pittsburgh, PA, USA
{gdyke, dadamson, ihowley, cprose}@cs.cmu.edu

Abstract. In this paper, we investigate the use of conversational agents to scaffold on-line collaborative learning discussions through an approach called *academically productive talk*. In contrast to past work, which has involved using agents to elevate the conceptual depth of collaborative discussion by leading students in groups through directed lines of reasoning, this approach lets students follow their own lines of reasoning and promotes productive practices such as explaining, stating agreement and disagreement, and reading and revoicing the statements of other students. We contrast two types of academically productive talk support for a discussion about 9th grade biology and show that one type in particular has a positive effect on the overall conversation, while the other is worse than no support. This positive effect carries over onto participation in a full-class discussion the following day. We use a sociolinguistic style analysis to investigate how the two types of support influence the discussion and draw conclusions for redesign. In particular, our findings have implications for how dynamic micro-scripting agents such as those scaffolding academically productive talk can be used in consort with more static macro- and micro- scripting.

Keywords: conversational agents, discussion scaffolding, collaboration scripting

1 Introduction

In recent years there has been a series of successful results in the area of conversational agents to support learning in chat environments [2][4][6-11]. Such agents have provided social support, affording the agents a more credible social standing in the group and helping to diffuse tension and create a productive learning environment. Furthermore, they have provided conceptual support, designed to elicit more depth by leading students through directed lines of reasoning, referred to as *knowledge construction dialogues* (KCDs).

While KCDs have been shown to lead to increased learning gains, particularly in situations where the conversational agents also provide social support [8], the necessity of designing them statically, with a pre-defined line of reasoning in mind both makes them hard to adapt to new subject material and

adfa, p. 1, 2011.

does not fully exploit the benefits of collaborative learners following their own spontaneous lines of reasoning.

We have therefore drawn on extensive work related to support of classroom discourse [12-14] and collaborative learning [3,15] to investigate the use by conversational agents of facilitation moves that promote *academically productive talk* (APT). The aim of APT facilitation moves is to increase the amount of *transactivity* [3], by dynamically reacting to student discussions, encouraging them to build on each other's reasoning. Furthermore, as APT refers both to learners social positioning to each other and their conceptual positioning to knowledge, this provides us with a theoretical framework to better integrate the social and conceptual support aspects of conversational agents.

In this paper, we analyse our first study involving an agent performing APT moves in the context of a 9th grade biology classroom. We contrast two forms of support (one in which the agent performs the facilitation and a second in which the agent prompts another student to perform these moves) and a null condition with no support. We show that the presence of APT moves is correlated with improved student reasoning but also discover that while the first form of APT support shows promise, the second produces much less reasoning than would be expected. In order to better understand how the agents shape the conversation, both productively and unproductively, we employ a linguistic style process analysis to inform the next iteration of development of academically productive talk agents.

2 Academically Productive Talk

The notion of Academically Productive Talk stems from frameworks that emphasize the importance of social interaction in the development of mental processes, and has developed in parallel to similar ideas from the computer-supported collaborative learning community. Michaels, O'Connor and Resnick [12] describe some of the core dialogic practices of Accountable Talk along three broad dimensions:

- Students should be accountable to the learning community, listening to the contributions of others and building on them to form their own.
- Students should be accountable to accepted standards of reasoning, emphasizing logical connections and drawing reasonable conclusions
- Students should be accountable to knowledge, making arguments which are based explicitly on facts, written texts or other public information.

In order to introduce such practices in the classroom where they do not exist, it is necessary both to introduce students to unfamiliar dialogic interaction forms and to provide teachers with the means to scaffold these interaction forms. Drawing on over 15 years of observation and study, Michaels, O'Connor and Resnick [12] propose a number of core "moves" that teachers can draw upon in order to encourage the development of academically productive classroom discussion, among which are:

1. Revoicing: “So let me see if I’ve got your thinking right. You’re saying XXX?” (with time for students to accept or reject the teacher’s formulation);
2. Asking students to restate someone else’s reasoning: “Can you repeat what he just said in your own words?”;
3. Asking students to apply their own reasoning to someone else’s reasoning: “Do you agree or disagree and why?”;
4. Prompting students for participation: “Would someone like to add on?”;
5. Asking students to explicate their reasoning: “Why do you think that?” or “How did you arrive at that answer?” or “Say more about that”.
   These moves have in common that they encourage reasoning statements (where the reasoning is made explicit) and they encourage transactivity [3], in which a reasoning *operates on* previous reasoning statement.

3 An Agent to Facilitate Academically Productive Talk

In this study, 50 students in four 9th grade biology periods were involved in an activity about diffusion and osmosis over two 42-minute periods on consecutive days. On the first day, they went through a 20 minute discussion in groups of three, in which a conversational agent presented them with three similar experimental setups, asking them to make predictions, watch a video, record their observations and provide explanations. This agent also provided APT scaffolding according the condition to which the groups were assigned. Furthermore the students were assigned roles related to APT scaffolding, with each student being responsible for performing one type of scaffold when appropriate. On the second day, the students participated in full class discussions, led by their teacher, at the end of which they took a post-test. Our research goal was to evaluate two forms of APT support. Our educational goal was to prepare the students as well as possible for the second day’s discussion so that they might each benefit from it as much as possible.

3.1 Agent Support for Academically Productive Talk

The APT conversational agent was setup to accomplish two roles, neither of which provided any conceptual support. The first was to guide and instruct students through each phase of the activity. The second was to provide various levels of scaffolding using three of the “moves” proposed for the scaffolding of APT: prompting students to restate each other’s reasoning, asking students whether they are in agreement with each other or not, and asking students to further explicate their reasoning.

The levels of support formed the three experimental conditions of our study:

- **Unsupported**: provide no APT support (only guiding through phases of activity)
- **Direct**: directly prompt students using APT moves ("John, could you say what Ann said in your own words")
- **Indirect**: prompt students to fulfill their assigned role ("Susan, could you ask John to say in his own words what Ann said").

In a pilot study using human "wizards of Oz" to provide APT support, students reacted unfavorably to the tutors – we hypothesized that in such a social situation a computer agent might not have the authority and credibility to make APT move requests of the human participants. The Indirect condition was designed to mitigate this situation by prompting learners to fulfill a role which had already been assigned to them *in lieu* of the agent.

Student1	I think it's going to get heavier.
Tutor	Student2, do you agree with what Student1 just said?
Student2	Wait I'm confused, please explain this again.
Student1	The egg will get bigger… heavier
Tutor	Student3, do you agree with what Student1 just said?
Student3	I can't understand.
Student3	oh, ok, I get it.

In the example above, when the agent detects that a student has made a prediction, it tries to get the other students to challenge the prediction. In this case, the response is that both of the other students admit that they are confused. This is actually a productive response since voicing confusion can be a precursor to a useful clarification dialogue. If students don't voice their confusion, they are less likely to achieve clarity within the conversation. In the Indirect condition, the Tutor would have said: *Student3, check with Student2 if they agree with Student1*.

4 Analysis

In our analysis, presented below, we initially examine the students' conversations and the effect of the ATP support conditions, by coding utterances for accountable talk moves, reasoning, and transactivity. Reasoning movesWe then examine the effect on participation in the following day's full class discussion and the learning outcome subsequent to that discussion. This shows that the Direct condition outperforms the None and the Indirect.

We then perform a more detailed process analysis of linguistic style, to investigate why the Indirect condition performs so poorly. We investigate specific areas in the conversations where Indirect seems different from the other two conditions and isolate some of the issues which will be a focal point for APT agent redesign.

4.1 Reasoning in Conversations

We first coded for APT moves (which follow a set template), reasoning (0.72κ interrater reliability), and transactivity (0.70κ).

Table 1. APT Moves, Reasoning, Transactivity per student, across all conditions

Condition	Student APT Moves	APT Moves (including tutor)	Reasoning	Transactiv-ity
Unsupported	.56 (2.7%)	1.6 (1.8%)	1.6 (11%)	.55 (2.7%)
Indirect	1.2 (4.9%)	3.8 (3.6%)	.53 (3.8%)	.13 (1.1%)
Direct	.67 (6.4%)	4.25 (7%)	2 (17%)	.92 (5.1%)

It should first be admitted that, overall, these results are lower than we had expected, with little reasoning and transactivity, mainly because of the difficulty the students had in carrying out the activity. The biggest difference between conditions shows up in terms of explicit displays of reasoning. Here there is a marginal effect on total number of reasoning moves per session $F(2,42) = 2.46$, $p < .1$, whereby students in the Direct condition produce a significantly greater number of reasoning moves than students in the Indirect condition, with the Unsupported condition not being significantly different from either (this same effect is significant when considering reasoning moves as a percentage $F(2,42) = 4.47$, $p < .05$). We did not see any statistical relationship between the number or percentage of Academically Productive Talk moves from the tutor and either student reasoning displays or transactive moves, however, we did see a significant but weak correlation between total percentage of Academically Productive Talk moves in a chat transcript from any source and the percentage of student contributions that were explicit displays of reasoning $R2 = .11$, $p < .05$. Given this result, and the non-significant trend of the Indirect condition having more APT moves (both from the students and from all participants), it is surprising that the Direct condition outperformed the Indirect condition in producing reasoning.

4.2 Effect on Full-Class Discussion Participation

We examined the effect on class participation by counting contributions to the teacher-led discussion. Because the data were far from normally distributed, we first did a log transformation on the counts of contributions. We then performed an ANOVA analysis to determine whether there was a significant effect of condition. Since there was also a big difference in participation (and ability) across class periods, we retained class session as an additional factor in the ANOVA analysis. Both class session ($F(3,21) = 7.0$, $p < .005$) and condition ($F(2,26) = 4.2$, $p < .05$) were statistically significant[1]. A post-hoc analysis using t-tests demonstrated that students in both the Direct and Indirect conditions contributed to the whole group discussion significantly more frequently than students who had been in the Unsupported condition. In both cases the effect size was about .75 standard deviations.

Table 2. Classroom discussion participation by Period and Condition

	Unsupported	Indirect	Direct
Period 1	4.2 (3.7)	8.0 (5.9)	3.7 (2.1)
Period 3	N/A	19 (8.5)	60 (49.5)
Period 6	1 (0)	3.2 (2.1)	5.8 (5.3)
Period 9	1 (0)	20 (0)	7 (0)

4.3 Learning Gains

The major factor influencing post-test results was the class period. The performance of all but the first period was so poor that no results of any significance were observable. To increase statistical power, we examined the effect of condition only on the first period (grouping Direct and Indirect conditions into the Supported condition) and only on questions related to providing generic explanations (as opposed to fact recall and observation understanding). Students in the Supported conditions scored significantly higher than those in the Unsupported $F(1,46) = 4.3$, $p < .05$, with an effect size of 1.1sd.

Table 3. Post-test score on Explain for Period 1, by condition (mark is out of 4 points)

	Supported	Unsupported
Explain	2 (.7)	1.1 (.9)

4.4 Process Analysis of Linguistic Style

From the above analyses it is surprising that the Indirect condition produced such poor reasoning compared to the Direct. We therefore examined the conversations in greater detail. In addition to Transactivity, which shows how students reason and operate on each others' reasoning, we coded the discus-

[1] Because of the difficulty in indentifying participating students in our audio recordings of the class discussion, this data is incomplete and the analysis may not accurately reflect the effect of participation on discussion. On the other hand, there is no reason to assume that our ability to identify students was biased by condition.

sions for Heteroglossia (0.77κ inter-rater reliability), which shows how participants frame their assertions. The Heteroglossia framework is operationalized from Martin and White's theory of engagement (Martin & White, 2005), and here we describe it as identifying word choice that allows or restricts other possibilities and opinions. This creates a rather simple divide in possible coding terms for contributions (among statements that are ontask assertions):

- Heteroglossic-Expand (HE) phrases tend to make allowances for alternative views and opinions (such as "She claimed that glucose will move through the semi-permeable membrane.")
- Heteroglossic-Contract (HC) phrases attempt to thwart other positions (such as "The experiment demonstrated that glucose will move through the semi-permeable membrane.")
- Monoglossic (M) phrases make no mention of other views and viewpoints (such as "Glucose will move through the semi-permeable membrane.")

Overall, we find a positive and strong correlation between the average percentage of HE contributions in a discussion and the percentage of a student's contributions that are explicit reasoning displays, R2 = .5, p < .0001. We also see a significantly smaller percentage of student contributions that are Heteroglossic Expand F(2,41) = 6.79, p < .005 in the Indirect condition.

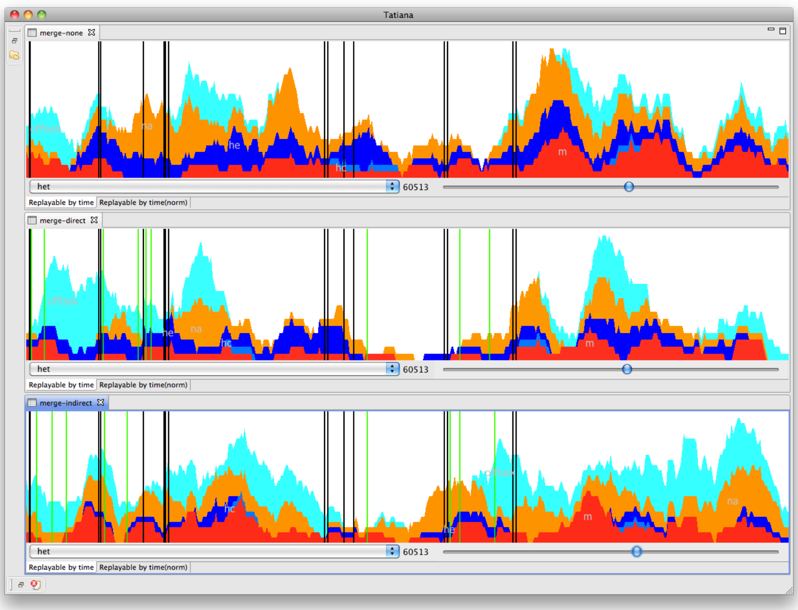

Fig. 1. Heteroglossia (M=red, HE=blue, HC=light blue, Non Assertion=orange, Offtask=cyan) distributed over time (horizontal axis) in Unsupported (top), Direct (center) and Indirect (Bottom). Dark bars indicate tutor turns present in all conditions

and groups, green bars are the APT moves specific to individual conditions and groups. The dip in the middle is where the students watch the video.

To better understand what was happening in the indirect condition, we used Tatiana [5] to construct a visualization showing the running average distribution of heteroglossia codes over time within each of the three conditions (cumulating across the groups of each condition, cf. Fig. 1). We can see that during the prediction phase, before going to see the video but after several ATP moves by the agent, there is a marked lack of HE turns and a marked presence of M turns in the Indirect condition compared to the two others. The HE turns remain low throughout. By investigating these phases of the conversation more closely, we saw that HE statements tended to be predictions and explanations, whereas the M statements tended to be statements of incomprension.

Closer examination revealed that this was often triggered by the agent's macro-scripting of the activity (instructions of what to do) interfering with its micro-scripting of the APT (e.g. Fig. 2). Furthermore, the agent frequently lost credibility as performing moves in inappropriate places was doubly harmful when the students followed its instructions (e.g. asking for a revoice when there was nothing useful to revoice).

Tutor	You should now move on to discussing Condition C
S041	The world is going to end in 2012
Tutor	S027, now would be a good time to ask S034 to build on what S041 is saying.
S034	I'm so confused!
S027	034, would you like to build onto what 041 is saying? And me too!
Tutor	When you agree, write down your predictions on your worksheet.

Fig. 2. Inappropriate timing of request for revoicing in the Indirect condition

5 Discussion and implications for redesign

The analysis of heteroglossia helped us identify the major way in which the Indirect discussions were different from those in the other two conditions. It also showed that lack of heteroglossic turns and presence of monoglossic turns could be an indicator of areas where the tutor was negatively impacting the discussion. Aside from the interference between static macro- and dynamic micro- scripting, we also saw that students were often frustrated when the agent was unable to provide them with any conceptual insight.

In redesign for re-conducting a similar study this year, the agent scripting issues were solved as discussed in [1] by implementing a better coordination algorithm so as to avoid collisions from different agent types. The activity was simplified so as to be better tailored to the students abilities. Last, and most importantly, the revoicing and addon APT moves were adopted because of their ability to provide simple conceptual nudges. The revoicing agent compares student utterances to a set of sentences that it knows to be true about the domain. When it finds a sentence which is semantically similar, it asks the students whether it is an appropriate revoice (e.g. Fig. 3). The addon agent knows about a list of concepts which ought to be discussed and will prompt struggling groups to discuss these concepts. Both these agents allow for rapid extension to new domains without the need to develop elaborate lines of reasoning.

This year's study has just been conducted and preliminary analysis shows that the amount of reasoning is greater by at least an order of magnitude than the amount shown in this condition. The agents also were considered more credible and useful interlocutors than in the study presented in this paper.

S25	it mixed with the water and it turned green because of how much glucose is in the water
Tutor	Would that be the same as saying "a glucose test strip changes from yellow to green when glucose is there?"
S25	Yes

Fig. 3. In this excerpt from the new revoicing agent, the tutor's utterance both reframes the student utterance more precisely, and leaves the responsibility for producing and evaluating reasoning with the student.

6 Conclusion

In this paper, we presented a new kind of conversational agent for learning, based on the theoretical framework of Academically Productive Talk. Such agents are designed to behave as beneficial generic participants in collaborative learning discussion situations. We described a study in which two different implementations of such APT support are compared and contrasted with an Unsupported condition. While our activity proved to be slightly too difficult, the Supported conditions are shown to provide better learning outcomes and increased participation in subsequent classroom discussion. The Direct condition is shown to outperform the Indirect condition in increasing the amount of student reasoning. A process analysis of linguistic style is used to investigate this difference more closely, revealing several issues with the agents as implemented. In a promising redesign, we implemented new kinds of APT moves such as revoicing and adding on and a better coordination mechanism for loosely coupled agents. We believe APT agents open the doors to creating agents which can be reused in a variety of contexts with minimal adaptation effort. Furthermore, they provide new opportunities for controlled research into the effects and pertinence in context of various APT and other discussion scaffolding moves.

7 Acknowledgments

This work was funded by NSF SBE-0836012.

8 References

1. Adamson, D., Rosé, C. P. Coordinating Multi-dimensional Support in Collaborative Conversational Agents, in Proceedings of Intelligent Tutoring Systems (2012)
2. Ai, H., Kumar, R., Nguyen, D., Nagasunder, A., Rosé, C. P. Exploring the Effectiveness of Social Capabilities and Goal Alignment in Computer Supported Collaborative Learning, in Proceedings of Intelligent Tutoring Systems (2010)
3. Berkowitz, M., & Gibbs, J. Measuring the developmental features of moral discussion. Merrill-Palmer Quarterly, 29, 399-410 (1983)
4. Chaudhuri, S., Kumar, R., Howley, I., Rosé, C. P. Engaging Collaborative Learners with Helping Agents, submitted to Artificial Intelligence in Education (2009)
5. Dyke, G., Lund, K.,& Girardot, J.-J. Tatiana: an environment to support the CSCL analysis process. *CSCL 2009*, Rhodes, Greece,58—67 (2009)
6. Howley, I., Chaudhuri, S., Kumar, R., Rosé, C. P. Motivation and Collaboration On-Line, submitted to Artificial Intelligence in Education (2009)
7. Howley, I., Mayfield, E., Rosé, C. P. Missing Something? Authority in Collaborative Learning, in Proceedings of Computer Supported Collaborative Learning (2011)
8. Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. Tutorial Dialogue as Adaptive Collaborative Learning Support, Proceedings of Artificial Intelligence in Education (2007)
9. Kumar, R., Ai, H., Beuth, J., Rosé, C. P. Socially-capable Conversational Tutors can be Effective in Collaborative Learning Situations, in Proceedings of Intelligent Tutoring Systems. (2010)
10. Kumar, R. & Rosé, C. P. Architecture for building Conversational Agents that support Collaborative Learning, IEEE Transactions on Learning Technologies special issue on Intelligent and Innovative Support Systems for Computer Supported Collaborative Learning (in press)
11. Martin, J. & White, P. The Language of Evaluation: Appraisal in English, Palgrave (2005)
12. Michaels, S., O'Connor, C., & Resnick, L.B. Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. Studies in Philosophy and Education (2007)
13. Resnick, L. B., Bill, V., & Lesgold, S. Developing thinking abilities in arithmetic class. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), Neo-Piagetian theories of cognitive development: Implications and applications for education (pp. 210-230). London: Routledge (1992)
14. Resnick, L., O'Connor, C., and Michaels, S. Classroom Discourse, Mathematical Rigor, and Student Reasoning: An Accountable Talk Literature Review (2007)
15. Weinberger A., Fischer F. A framework to analyze argumentative knowledge construction in computer supported collaborative learning. Computers & Education; vol 46, pp.71– 95 (2006)