Intern. J. Comput.-Support. Collab. Learn
DOI 10.1007/s11412-016-9246-2

Conversational agents for academically productive talk: a comparison of directed and undirected agent interventions

Stergios Tegos[1] · Stavros Demetriadis[1] · Pantelis M. Papadopoulos[2] · Armin Weinberger[3]

Received: 31 July 2016 / Accepted: 3 November 2016

Abstract Conversational agents that draw on the framework of academically productive talk (APT) have been lately shown to be effective in helping learners sustain productive forms of peer dialogue in diverse learning settings. Yet, literature suggests that more research is required on how learners respond to and benefit from such flexible agents in order to fine-tune the design of automated APT intervention modes and, thus, enhance agent pedagogical efficacy. Building on this line of research, this work explores the impact of a configurable APT agent that prompts peers to build on prior knowledge and logically connect their contributions to important domain concepts discussed in class. A total of 96 computer science students engaged in a dialogue-based activity in the context of a Human-Computer Interaction (HCI) university course. During the activity, students worked online in dyads to accomplish a learning task. The

Q3 References marked as (XX, in press) and (XX, 2016) in the text for blind review:
Tegos, S., & Demetriadis, S. (in press). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. Educational Technology & Society.
Tegos, S. (2016). *Web-Based Conversational Agents for Collaborative Learning Support* (Doctoral dissertation). Aristotle University of Thessaloniki.

✉ Stergios Tegos
stegos@csd.auth.gr

Stavros Demetriadis
sdemetri@csd.auth.gr

Pantelis M. Papadopoulos
pmpapad@tdm.au.dk

Armin Weinberger
a.weinberger@mx.uni-saarland.de

[1] School of Informatics, Aristotle University of Thessaloniki, Ethn. Antistaseos 16 str, 55133 Kalamaria, Greece

Q2 [2] Centre for Teaching Development and Digital Media, Aarhus University, Aarhus, Denmark

[3] Department of Educational Technology, Saarland University, Saarbrücken, Germany

study compares three conditions: students who collaborated without any agent interference (control), students who received undirected agent interventions that addressed both peers in the dyad (U treatment), and students who received directed agent interventions addressing a particular learner instead of the dyad (D treatment). The results suggest that although both agent intervention methods can improve students' learning outcomes and dyad in-task performance, the directed one is more effective than the undirected one in enhancing individual domain knowledge acquisition and explicit reasoning. Furthermore, findings show that the positive effect of the agent on dyad performance is mediated by the frequency of students' contributions displaying explicit reasoning, while most students perceive agent involvement favorably.

Keywords Conversational agent · Academically productive talk · Computer-supported collaborative learning · Peer dialogue

Introduction

Language is argued to be the most powerful mediating tool for cognitive development, while dialogue is the foundational act of language (Resnick, Michaels, & O'Connor, 2010). Drawing on the strong associations of peer dialogue with learning outcomes in a variety of contexts, research in the area of computer-supported collaborative learning (CSCL) has repeatedly emphasized the importance of fruitful dialogical interactions among learners (Stahl et al. 2014). The depth and quality of peer interactions, such as conflict resolution, mutual regulation or explicit argumentation, have been found to play a catalytic role in the extent to which students comprehend the topic in question and learn from collaborative activities (e.g., Asterhan and Schwarz 2016).

However, although peer interactions constitute a significant learning mechanism, their presence is not always assured since students' dialogue is often unproductive (Dillenbourg and Tchounikine 2007). Even in structured dialogue-based activities, placing students together and asking them to discuss a topic with each other does not ensure their engagement in effective collaborative behavior (Vogel et al. 2016). Therefore, aside from such methods as manipulating the design of collaborative tasks, researchers typically explore how to increase the likelihood of constructive interactions by monitoring small-group dialogue and delivering supportive interventions when appropriate (Webb 2009). Still, some questions readily emerged, as to what an effective dialogue should be like, or as to how design-based research could contribute to the development of CSCL environments by providing scaffolding during group discussions (Ludvigsen and Mørch 2010). Could agent technologies utilize discourse facilitation strategies used in classroom to help students sustain a productive peer-to-peer dialogue in diverse learning situations? (Goodman et al. 2005; Howley et al. 2013) How should an agent intervene during a peer dialogue considering that not all one-learner-setting assumptions (e.g., the near-even student participation one) apply to a multi-user setting? (Harrer et al. 2006; Kumar and Rosé 2011).

This study explores the impact of conversational agent interventions on peer dialogue, specifically focusing on and analyzing the differences arising from varied intervention modality. Agent interventions have been modeled after a discourse facilitation strategy that is commonly implemented by teachers in class. The study provides evidence that the level of peers' explicit reasoning and subsequent learning outcomes are affected by the way that the agent addresses peers in a dyad during online discussions.

Academically productive talk

A classroom discourse framework, namely academically productive talk (also known as APT or Accountable Talk), has emerged through teachers' exploration of effective classroom discussion practices on how to promote academic learning and reasoned student participation (Michaels and O'Connor 2013; Michaels et al. 2008; Resnick et al. 2010; Sohmer et al. 2009). This framework focuses on the key role of social interaction in learning. According to APT (Resnick et al. 2010), students' discussions should be accountable to:

The learning community: students should listen to and build upon their partners' ideas, learning from each other as the discussion unfolds.

Accurate knowledge: students should support the validity of their contributions using explicit evidence and making references to a pool of knowledge accessible to the group (e.g., a textbook or presentation).

Rigorous thinking: students should focus on logically connecting their claims in a reasonable manner, evaluating the soundness of their arguments and drawing valid inferences.

Following extensive research base on classroom discourse, APT encourages instructors to utilize a set of strategic interventions (talk moves). The latter have been conceptualized as useful tools for triggering and modeling valuable forms of students' discourse (Sohmer et al. 2009) and for responding to challenges teachers face in facilitating discussions (Michaels and O'Connor 2013). The effective implementation of APT interventions, such as the ones depicted in Table 1, can help maintain a rigorous, coherent, engaging and equitable discussion (Michaels et al. 2010). There is also converging evidence that such APT facilitation strategies can deepen students' understanding of complex material and lead to academic achievements in diverse classroom situations and educational contexts (Michaels et al. 2008).

An important aspect of APT is that it prioritizes students' reasoning over correctness and does not expect the teacher to maintain complete control over students' discussions (Michaels et al. 2010). This distinguishes it from other widely used classroom discourse formats, such as the IRE/F (initiation-response-evaluation/feedback), where the teacher initiates discussion by asking a question, awaits a response from the student, and closes down discussion after evaluating the student's response and providing suitable feedback (Michaels and O'Connor 2013). APT aims to relinquish instructor's authority on the topic under discussion and orchestrate a more student-centered discussion, where students are motivated and challenged

Table 1 A selection of APT interventions

Intervention	Example	Accountability
1. Add-on	"Would you like to add something to what … said about?"	Learning community
2. Agree-disagree	"Do you agree or disagree with what your partner said about …? Why?"	Learning community
3. Re-voice	"So, are you saying that…? Is that correct?"	Learning community
4. Press for accuracy	"Could you identify that in a reference book?"	Accurate knowledge
5. Build on prior knowledge	"How does this connect with what we know about …?"	Accurate knowledge
6. Press for reasoning	"Why do you think that?"	Rigorous thinking
7. Expand reasoning	"That's interesting! take your time elaborating on that…"	Rigorous thinking

to think profoundly and make use of scientific reasoning skills to solve problems. In an academically productive peer discussion, students are expected to engage intellectually. Students actively participate and contribute to the conversation of their group, communicate their reasoning, pay attention to their partners' contributions and construct logical arguments utilizing accurate evidence (Michaels et al. 2010).

APT emphasis on students' explicit reasoning coincides with the view of many researchers exploring features conducive to a productive peer dialogue. Although pertinent studies have been conducted from both a cognitive and a socio-cultural perspective, it has been shown that the formalized identification of an effective dialogue can be a complex challenging task (Weinberger and Fischer 2006). The theories that have emerged vary in conceptualization and terminology (e.g., productive agency, social modes of co-construction and transactivity); yet, they share the view that knowledge construction during peer dialogue occurs through a series of steps where learners' mental models are explicitly shared, mutually examined and possibly integrated (Stahl and Rosé 2011).

Under this prism, some consistencies were identified while investigating vital conversational characteristics and behaviors fostering meaningful learning (Sionti et al. 2012). One of these was reported to be the explicit articulation of students' reasoning (Stahl and Rosé 2011). Indeed, a common issue is that sometimes learners do not make their perspectives explicit to the group so that a common ground can be negotiated and a consensus be reached (Weinberger et al. 2007). According to Brandom (1998, page xviii), making something explicit can be described as the process of putting a claim into "a form in which it can be given as a reason, and reasons demanded for it". This is especially important in written dialogue where the externalization of students' reasoning can be essential to both the development of explicit references, thus enhancing dialogue coherence (Oehl and Pfister 2010), and the facilitation of peer interactions and grounding processes that affect the outcome of students' collaboration (Papadopoulos et al. 2013). The explicitness of students' reasoning can also be regarded as a prerequisite for dialogue transactivity, itself considered to be a valuable indicator of the learning taking place during peers' discourse (Sionti et al. 2012). Transactivity can be described as the degree to which learners use their partners as resources, referring to and building on each other's reasoning as the dialogue unfolds (Noroozi et al. 2013). This form of dialogue is found to positively impact learning outcomes and argumentative knowledge construction in collaborative scenarios (Chi 2009).

Learners rarely engage in transactive, academically productive talk spontaneously (e.g., Noroozi et al. 2013). Among the threats to APT is diffusion of responsibility of learners stepping back from a task with peer learners present. Learners may engage in a collaborative task to different degrees, but still benefit from the teamwork equally (Slavin 1992). Moreover, heuristics of how to engage in APT may be more or less readily available to the learners (Fischer et al. 2013).

One approach to addressing these problems is to guide and prompt learners to execute specific, productive discourse moves with set scripts that could either be trained or implemented in CSCL environments (Fischer et al. 2007). Scripts can help individual group members to engage in specific discourse moves, but may also alter mutual expectations regarding the roles and responsibilities within a group (Weinberger 2011). However effective, with instructional scripts typically being inflexible to situational changes or to needs of individual group members, scripts may become redundant and learners' perception of their usability may falter quickly.

Promoting academically productive discussions with conversational agents

Over the years, advances in computational linguistics and the rapidly expanding role of artificial intelligence in education have aroused a growing interest in developing conversational agents as tools to providing adaptive, flexible support in collaborative learning activities (e.g., Adamson et al. 2014; Kumar and Rosé 2011). In educational settings, conversational agents are commonly regarded as pedagogical agents that typically communicate with the learners in natural language in an attempt to act a pedagogical role, such as a tutor, coach or learning companion (Gulz et al. 2011).

Unlike research focusing on agents engaging in a one-to-one tutorial dialogue with the learner (e.g., Rus et al. 2013), researchers have also explored the design and usage of conversational agents aiming to scaffold productive group discussions (e.g., Adamson et al. 2014; Dyke et al. 2013; Stahl 2015; Tegos et al. 2015). Inspired by the work on APT, these types of agent are usually designed to act as peer dialogue facilitators during collaborative activities, promoting students' engagement in fruitful conversational interactions through a series of APT interventions (Stahl 2015). Such agents typically have a limited range of how they can navigate natural discourse and often display simple prompts that aim at eliciting student thinking instead of providing content-specific explanations or instructional assistance. Drawing on a considerable body of work suggesting that APT facilitation strategies can be beneficial for learning across a wide range of subject areas (e.g., Michaels et al. 2008), a major advantage of this flexible form of dialogue support is that, to a certain extent, it can be domain-independent and scalable.

Adamson et al. (2013) investigated the impact of an Agree-Disagree agent intervention mode, which prompted students to comment on their partners' statements (e.g. "What do you think about John's idea? Do you agree or disagree?") (Table 1, item 2). The study was conducted in the context of a chemistry university course and involved undergraduate students working in small groups to accomplish a collaborative task. Findings revealed that the agent had a marginal positive effect on students' learning and intensified knowledge exchange during group discussions. Following a similar rationale, a study explored the impact of an agent intervention mode that delivered both Agree-Disagree and Add-On interventions (Table 1, items 1 and 2) during an online dialogue-based activity, which took place in the context of a computer science university course (Tegos et al. 2015). The results were in line with Adamson et al.'s (2014), indicating that agent interventions encouraging peers to think together can amplify students' explicit reasoning processes and improve learning performance at both the individual and group level. Another study employing a similar intervention strategy showed that unsolicited APT interventions, automatically triggered and displayed by the agent, can be more efficient in increasing the level of explicit reasoning as compared to solicited APT interventions, triggered automatically but only displayed upon students' request (Tegos et al. 2014).

In a study involving 9th grade biology classes, Adamson and Rosé (2013) compared an Agree-Disagree intervention mode with a Revoicing one (Table 1, item 3), which aimed to help students externalize, expand and clarify their own thinking (e.g. "So what I hear you saying is 'X'. Is that right?"). The results revealed that the Revoicing strategy was more beneficial than the Agree-Disagree one for this age group. Following a similar rationale, Dyke et al.'s (2013) study in the same domain contrasted the performance of a Revoicing mode to an APT Feedback intervention mode, providing encouragement for students engaging in APT-based behaviors (e.g. "Thanks for offering an explanation, John"). Although Feedback interventions did not affect students' learning, study findings indicated a positive learning effect of the Revoicing intervention mode, which led to a more intensive reasoning exchange between peers. Two months later, another study was conducted involving the same participants in a similar context (Adamson et al. 2014). This time, no significant learning effect was detected for Revoicing. It was assumed that the difference in results was owed to the fact that the material of the latter study was easier for the students since at that time students got familiar with the subject. Interestingly, a last study in the context of an engineering university course reported a negative learning effect for the Revoicing intervention mode (Adamson et al. 2014).

Though encouraging, the findings emerging from the studies in this area suggest that the efficacy of APT agents may significantly vary depending on factors such as the type of intervention employed (Table 1), the difficulty of the instructional domain or students' background knowledge. Even though an Agree-Disagree agent intervention mode can be appropriate for advanced learners who are somewhat experienced in the subject and have solid argumentation skills, a Revoicing mode, which focuses on eliciting self-oriented conversational moves, appears to be beneficial only for novices or young learners not always capable of articulating their own ideas.

In this perspective, more fine-grained experimentation is needed to understand the potential benefits of APT agents and determine the context in which each intervention mode can perform most effectively (Adamson et al. 2014). Additionally, apart from the need to investigate usability and student acceptance issues, such as how the learners perceive and respond to the agent interventions, intriguing questions arise concerning the optimal design and configuration of such agents. Further research could be conducive to developing more efficient and agile APT agents, especially considering that most human instructors tend to be highly adaptive and responsive to multiple class parameters when selecting a specific APT intervention strategy and the timing or the target of their intervention (Hmelo-Silver 2013). For instance, given that an important aspect in CSCL systems design is how interventions are presented and address learning partners (Magnisalis et al. 2011), could the efficacy of an APT agent be drastically affected by whether its interventions target a single student or the whole group?

Research objectives

In view of the above research questions and line of research, this work investigates the utilization of a Building-on-Prior-Knowledge intervention mode (Table 1, item 5), which is operated by a configurable conversational agent in the context of a collaborative activity in higher education. Expanding on prior research on how to promote accountability to the learning community via dynamic APT agent interventions (e.g., Adamson and Rosé 2013; Tegos et al. 2015), this study explores the impact of an APT agent intervention mode that aims to promote accountability to accurate knowledge by encouraging students to link their current contributions to important domain concepts or principles discussed in class (e.g., "Does the KLM model have anything to do with the hotkeys selection you are talking about? Please, elaborate."). In this manner, students are asked to support their claims by making reference to previous knowledge that they have access to (Michaels et al. 2010). Overall, the goal of this study is twofold: (a) to confirm a previous study finding indicating the effectiveness of an agent intervention mode that urges peers to build on their prior knowledge (Tegos and Demetriadis in press) and (b) to explore whether a directed intervention method (D: the agent addresses one particular student) can be more beneficial than an undirected intervention

Q1

method (U: the agent addresses both partners in the dyad) in terms of enhancing learning and explicit reasoning. We expect the results of this study to inform instructors and researchers what pedagogical benefits may arise and how to best utilize such rapidly deployable agent facilitation technologies operating on the basis of APT interventions.

Method

Participants and domain

A total of 96 undergraduate computer science students participated in the study (15 female; 81 male; age: 19–26, M = 20.58, SD = 1.41). All participants were enrolled in the second-year course "Human-Computer Interaction" (HCI), in which students become acquainted with methodologies of prototyping and evaluating human-centered interfaces and user experience (Preece et al. 2015). Additionally, students learn about the principles of cognition and perception required for effective interaction design. Hence, the learning goals encompass theoretical knowledge and its application to solving concrete design tasks. The study language was Greek and students' participation was a compulsory course assignment. Students were informed that their conversations would be recorded during the activity and consented for their data to be anonymously used for research.

Conversational agent system

The MentorChat prototype conversational agent system was used for the purpose of this study (Tegos 2016). MentorChat is a configurable chat-based environment, which enables students to participate in online synchronous collaborative activities. A MentorChat activity may include multiple phases, each asking students to collaborate in small groups to jointly answer an open-ended domain topic (Fig. 1a). The system components include the learner, the teacher and the conversational agent modules.

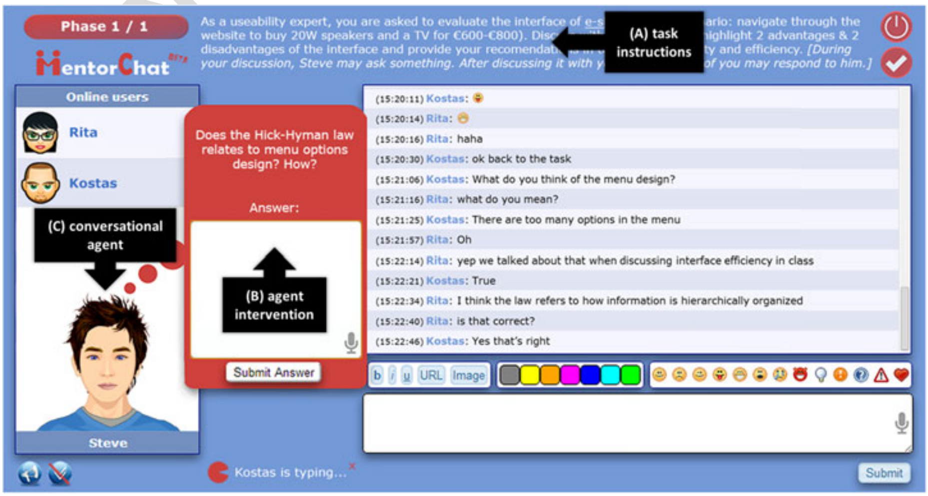

Fig. 1 MentorChat learning environment

The learner module provides an instant messaging interface (Fig. 1), allowing learners to communicate with each other through text or voice, using the speech recognition function to compose their messages. Students' discussions are monitored by a conversational agent. The agent decides to intervene displaying APT-oriented prompts to realize the experimental conditions building on a specific procedure described below. The agent interventions are displayed outside (on the left of) the main chat window (Fig. 1b). This mechanism serves as an *attention grabbing* strategy and enables peers to have constant access to the agent message so that they can respond to it whenever they choose. The agent possesses an animated 2D human-like representation (Fig. 1c). A text-to-speech (TTS) engine is also employed so that the agent can read its messages aloud.

MentorChat was developed to provide teachers with opportunities to apply concrete dialogue-based activities in their daily teaching. Using the administration panels a teacher can set up an online activity consisting of a series of phases (collaborative tasks), monitor students' discussions in real time, and configure the domain model of the conversational agent for each activity phase. The configuration of the agent domain model is accomplished through an integrated concept mapping tool (Fig. 2). In order to create a concept map, the teacher enters a set of simple statements (Fig. 2b), comprising three basic parts: a subject (concept A), an object (concept B), and a verb or verbal phrase (relationship of concepts). The system then renders and visualizes these elements in a concept map (Fig. 2a), which serves as the knowledge representation of the agent for the particular activity phase. Each agent concept map is then stored in a system library, which aims to facilitate the domain modeling process by enabling the reusability of the agent concept maps.

While a detailed analysis of the system components can be found in XX (2016), it should be noted that the conversational agent operates on the basis of a pipeline architecture, which includes three core models: the peer interaction, the domain, and the intervention models. In a nutshell, the peer interaction model is responsible for analyzing students' utterances and keeping track of the group chat history. Utilizing the agent concept map (Fig. 2b), a

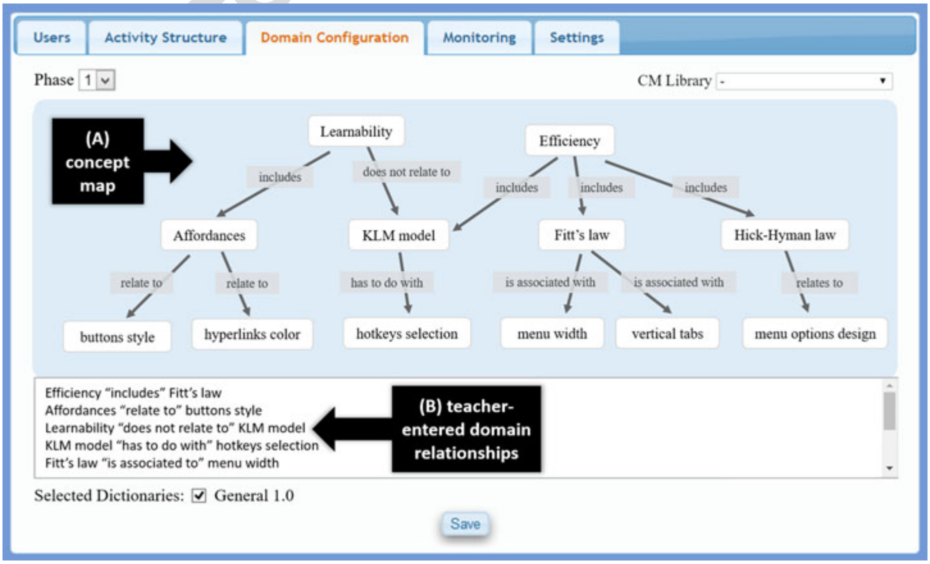

Fig. 2 MentorChat domain configuration panel

Q1

WordNet lexicon and a set of pattern matching and string similarity algorithms, this model creates a concept map for every student based on the concepts discussed by each peer. These maps are dynamically enriched with new concepts introduced by the peers as their discussion advances.

Next, the agent domain model compares the learners' concept maps with the agent concept map (Fig. 2a) in order to decide whether an agent intervention would be appropriate. For example (Table 2), in the version of the system used in this study, once the agent detects that students are discussing one of the concepts included in the agent concept map (e.g. "menu options design"), the agent may propose an intervention asking students to logically connect the concept being discussed with an associated higher-level concept of the map (e.g., the "Hick-Hyman law"). This may only occur if the particular higher-level concept has not been previously discussed.

In case an intervention is suggested by the agent domain model, the agent intervention model handles the synthesis of the intervention text on the basis of the teacher-defined statements (Fig. 2a) and a pool of pre-stored APT-based phrases including system variables. This model also manages the display time of each intervention by investigating a series of micro-parameters, such as time passed since the last agent intervention or the frequency of chat posts. Eventually, the examination of these variables enables the system to decide whether the agent intervention should be displayed or suppressed in order to avoid a potentially excessive interference from consecutive agent interventions appearing in a short time frame.

Procedure

The course instructor set up an activity in MentorChat by entering all participants' information as well as the task description. The instructor also created the agent concept map by entering a set of statements as the ones displayed in Fig. 2b. The activity requested students to (a)

Table 2 A dialogue excerpt showcasing an agent intervention addressing both peers

	User	Message
1.	Kostas:	What do you think of the menu design?
2.	Rita:	what do you mean?
3.	Kostas:	There are too many options in the menu
4.	Agent:	Does the Hick-Hyman law relates to menu options design? How?
5.	Rita:	Oh
6.	Rita:	yep we talked about that when discussing interface efficiency in class
7.	Kostas:	True
8.	Rita:	I think the law refers to how information is hierarchically organized
9.	Rita:	is that correct?
10.	Kostas:	Yes that's right
11.	Kostas:	*[Submitted Answer]*For increased efficiency, the N options of the menu should be presented in thematically organized categories so that a user searching for an item does not require a long time to click on it. Hick-Hyman suggests that this time depends on log2N
12.	Rita:	Great!
13.	Rita:	5 min to go

collaboratively assess the interface of an online shop in terms of efficiency and learnability, and (b) submit a joint answer to a learning question. The latter asked students to highlight (at least) two advantages and disadvantages of the interface and propose potential improvements, based on the usability principles discussed in the course.

The study involved three main phases: a pre-task, a collaborative and a post-task phase (Fig. 3). In the first pre-task phase students were automatically directed to an online pre-test after logging into MentorChat. The test was administered individually within a 20-min time frame.

In the second phase, after completing the pre-test, the students were randomly matched with other students waiting to engage in the collaborative activity (text-based chatting). Eventually, 48 dyads were formed and randomly allocated by the system to one control (16 dyads) and two treatment conditions (16 dyads in each). All dyads participated in the chat phase that lasted 40 min. Students were distributed between two university labs so that each dyad member would communicate with their partner using a computer in a different room.

Lastly, in the post-task phase, students had 25 min to complete the post-test individually, plus an additional 10-min period to fill in the opinion questionnaire. One week after the activity, students also participated in a semi-structured focus group session.

Research design

A pre-test post-test experimental design was used to investigate the effects of two Building-on-Prior-Knowledge (BPK) agent intervention methods. More specifically, the study employed a between-subjects research design and compared three conditions:

students collaborating in dyads to accomplish a learning task without any agent intervention (control condition);

students who received undirected BPK interventions while collaborating in dyads to accomplish the same task (U treatment condition);

students who received directed BPK interventions while collaborating in dyads to accomplish the same task (D treatment condition).

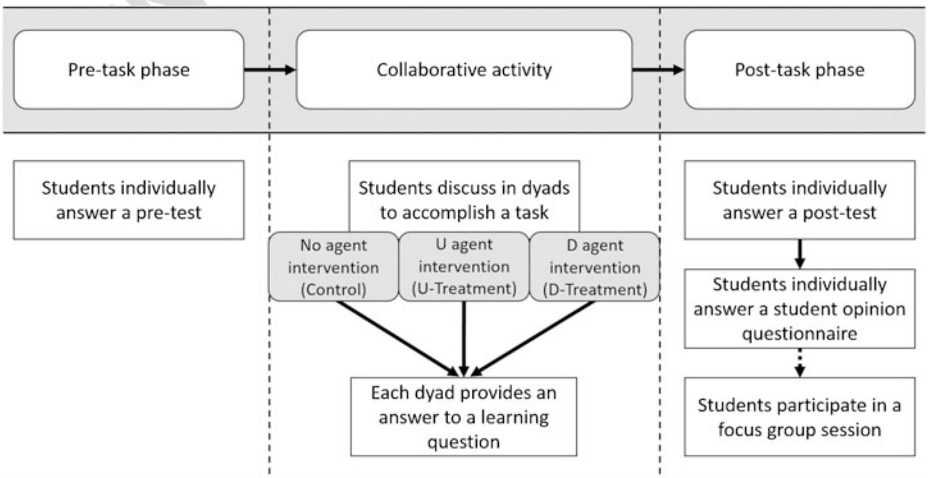

Fig. 3 Study workflow

The independent variable was the agent support, which varied in the different research conditions as discussed in the next section. The main dependent variables were the student learning, the dyad in-task performance, and the degree of explicit reasoning exhibited during students' discussions.

Study conditions

The students in the control condition collaborated without any interference from the conversational agent, which remained deactivated during the collaborative activity. However, as in all conditions, static system prompts were displayed in the chat window in order to support learners' awareness (e.g., "John has logged out") or provide simple instruction on interface features (e.g., "Submit an answer by clicking…").

In contrast to the control condition, the conversational agent operating in the treatment conditions displayed unsolicited dynamic interventions. Considering the Building-on-Prior-Knowledge APT facilitation strategy employed by the agent in this study (Table 1, item 5), the main objective of the agent interventions was to encourage students to support their claims leveraging knowledge acquired at a previous time. Particularly, the agent was tailored to ask students to link their current contribution revolving around a key domain concept to a relevant domain principle discussed during the course (Table 2, row 4).

Regarding the first treatment condition, the agent delivered undirected (U) interventions, which were simultaneously presented to both peers in the dyad (Fig. 4a). The dialogue excerpt presented in Table 2 illustrates such an agent intervention. As stated in the activity guidelines, the students of the U treatment condition were expected to respond to the agent in a coordinated way (one of them) using the agent answer box. When the student submitted a response, the answer box closed and the response remained available in the main chat panel.

In the second treatment condition, the agent was tailored to deliver directed (D) interventions. Although these interventions were displayed to both partners, as in the other treatment condition, in this condition only the student specified by the agent could submit a response using the agent answer box (Fig. 4b). Similarly to the U treatment condition, any response submitted remained accessible to both peers. The D intervention method addressed only the partner of the student who had triggered the agent intervention by introducing a key domain

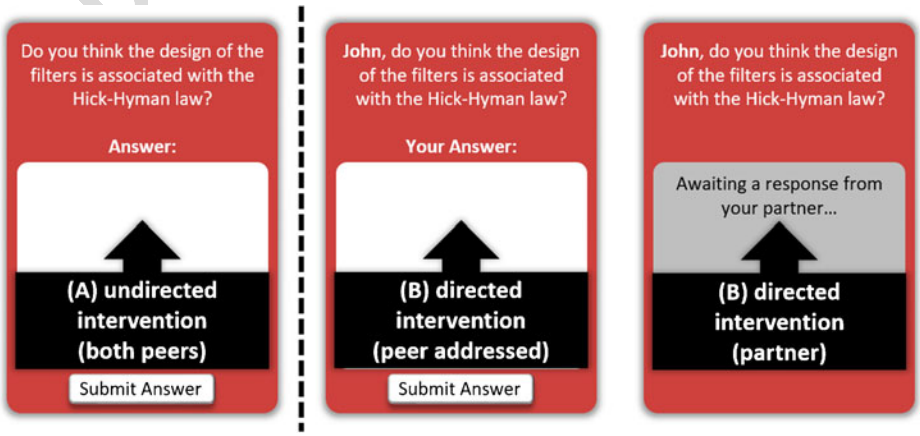

Fig. 4 The graphical interface for the U and D intervention methods

concept. As illustrated in Table 3, the assumption of the agent in the particular dialogue turn was that Jason might have a lesser understanding than Philip about the concepts brought up by Philip. Therefore, the agent decided to direct its question to Jason encouraging him to respond (Table 3, row 4).

Data collection and analysis

A .05 level of significance was set for all the statistical analyses conducted. Parametric tests were used only when the respective test assumptions, such as the data normality or homogeneity of variances, were not violated.

Individual learning

In order to measure students' domain knowledge before and after the experimental activity, students' pre-test and post-test answers were evaluated.

The pre-test consisted of two sections (10 points each). The first one included 10 multiple-choice questions and targeted at the lowest level of Bloom's taxonomy (Huitt 2011), focusing on recognition and memory retrieval. The second section included 4 open-ended questions and aimed at the second level of Bloom's taxonomy, requiring students to comprehend and interpret domain information based on their prior learning. Students' answer sheets were mixed and scored independently by two raters who had extensive experience in the HCI domain. Holistic rubric scales were used for the assessment of the open-ended questions. The intra-class correlation coefficient indicated a high inter-rater reliability (ICC = .99). The overall

t3.1 **Table 3** A ☺ intervention

		User	Message
t3.2		User	Message
t3.3	1.	Philip:	That doesn't seem right. The menu closes instantly if you move your mouse pointer out of the popup.
t3.4	2.	Jason:	That's correct.
t3.5	3.	Philip:	OK
t3.6	4.	Agent:	Jason, do you believe the Accot-Zhai law somehow relates to mouse movement in menu?
t3.7	5.	Jason:	Hmm, please give me a minute to respond.
t3.8	6.	Philip:	ok no problem, ask me if you need anything
t3.9	7.	Jason:	Do you remember the mathematical expression?
t3.10	8.	Philip:	T = a*b*(D/S) ☺
t3.11	9.	Jason:	*[Submitted Answer]*The Accot-Zhai (or steering) law, which predicts the time required to steer a pointing device through a 2D tunnel (T = a*b*(D/S)), relates to the top cascading menu as it does not provide users with the necessary time to navigate through the hierarchical menu options without closing.
t3.12	10.	Philip:	That's true because, although the menu appears to be OK in size, it could be improved by expanding the 'active' pointer region or placing menu items closer to each other
t3.13	11.	Jason:	That's a nice suggestion actually, we could also add a delay so that the sub-menu does not close immediately while moving the pointer between menu items
t3.14	12.	Philip:	Let's move on

pre-test construct (20-point scale), resulting from summing the scores of the two questionnaire sections, had a satisfactory internal consistency ($\alpha = .72$).

The post-test included six open-ended questions (20-point scale) and targeted at the second level of Bloom's taxonomy. Students' answers were scored by the same raters as in the second pre-test section. Their intra-class correlation coefficient was reported to be high (ICC = .96).

Both tests assessed students' knowledge on the same sub-domain ("Human-Computer Interaction: Designing for efficiency"), and were validated by the course instructor, an expert in the domain. It should be noted that the post-test purposely included only open questions since the inclusion of multiple-choice questions could constitute a source of bias in favor of the treatment students, who would have recently seen the concepts displayed by the agent, and thus could display improved performance by simply 'recalling' rather than displaying their 'understanding'.

To compare students' prior knowledge in the different conditions, a one-way analysis of variance (ANOVA) was conducted on pre-test scores. To determine the effect of the two agent intervention modes on students' learning, a one-way analysis of covariance (ANCOVA) was performed using the pre-test score as the covariate and the post-test score as the dependent variable. Additionally, since individual knowledge acquisition occurred during a collaborative session and agent interventions varied among the dyads, we introduced the dyad as a nested factor in our analysis of individual learning outcomes and performed a two-level nested ANOVA. This hierarchical analysis was chosen since there was one measurement variable (post-test score) and two nested nominal variables (conditions and dyads in conditions).

Dyad performance in task

In order to measure dyad in-task performance, all dyads' answers submitted in response to the main learning question of the activity were evaluated. The same raters who participated in the data analysis phase of the pre- and post-test questionnaires followed predefined instructions and used a 20-point rubric scale in order to score each dyad's answer submitted at the end of the collaborative activity. The scale demonstrated a satisfactory intra-class correlation coefficient (ICC = .94). A Kruskal-Wallis H test was run to determine if there were differences in the scores of the answers provided in the three conditions.

Explicit reasoning in discussion

A discourse analysis was performed to measure the level of explicit reasoning exhibited during peer discussions. Two of the authors proceeded to code students' contributions in two phases. In the initial phase, the authors independently coded a subset of students' discussions. Following a Cohen's kappa analysis, which indicated that there was satisfactory agreement between the two coders' judgements ($\kappa = .87$), any discrepancies found were addressed until consensus was reached. In the second phase, the authors collaboratively performed a line-by-line analysis of all students' contributions.

The coding process was based on an extended version of the IBIS discussion model, which is regarded as an effective model for analyzing conversational interactions occurring in online small-group collaborative activities (Liu and Tsai 2008). On top of the main categories of the IBIS model comprising issue, position and argument, the study scheme incorporated two additional (finer-grained) categories, named explicit position and explicit argument, both focusing on the detection of 'explicit reasoning displays'. The formulation of what an explicit reasoning display involved was primarily derived from the work of Sionti et al. (2012). The identification of contributions containing explicit reasoning did not require students' reasoning to be correct and mainly focused on students' attempts to think in a logical way, beyond what was given in the task instructions, leveraging previously acquired theoretical constructs and concepts. In this manner, a student's contribution could be identified either as an argument or an explicit argument based on whether it simply supported/objected to a previously articulated position (e.g., "true, this seems to be the case in this screenshot") or also displayed some form of explicit reasoning on domain concepts (e.g., "this is correct because the option has not nearly enough width in order to be easily selected – Fitts' law model"). A similar distinction was also made between positions and explicit positions. Table 4 depicts the scheme categories used in the discourse analysis along with some examples.

The frequencies of the above categories were calculated for each dyad based on the dyad contributions. A one-way ANOVA was conducted to determine whether there are any differences in the explicit position and explicit argument frequencies between the research conditions. Our aim was to explore whether the agent interventions had a significant impact on the display of students' reasoning.

In an attempt to investigate whether the agent interventions affected the distribution of explicit contributions within the dyads, we calculated a percentage for the learning partners in each dyad based on how many explicit contributions (explicit positions and explicit arguments) each peer had contributed. The term 'less explicit' was used conventionally for the learning partner with the lower percentage of explicit contributions in their dyad. A Kruskal-Wallis H-test was conducted to determine if there were any significant differences in the percentages of the 'less explicit' peers in all conditions.

Table 4 Discourse analysis scheme (XX, in press)

Category	Description
Off task	Contributions that do not relate to the task and often play a purely social function (e.g., "Hello", "Bye")
Repetition	Reiterations of prior contributions often repeated after some time for a better understanding.
Team management	Management-oriented utterances used for task coordination (e.g., "We do not have enough time let's submit our response")
Common understanding	Short utterances used to establish common understanding on the subject (e.g., "OK")
Issue	What needs to be done or resolved to proceed with the overall task (e.g., "What other laws are relevant?")
Position	Opinions usually related to the resolution of the issue raised (e.g., "Fitts' law applies here")
Argument	Opinions supporting or objecting to a position (e.g., "You are absolutely right")
Explicit position	Positions that explicitly outline reasoning on domain concepts (e.g., "According to Hick-Hyman, the reaction time increases logarithmically as the number of options increases")
Explicit argument	Much as explicit positions, arguments displaying explicit reasoning on domain concepts (e.g., "I disagree, Hick's law cannot be used for randomly ordered lists requiring linear time")

Q1

Moreover, a statistical mediation analysis was conducted following the procedure proposed by Hayes (2013). Our study investigated whether the frequency of explicit contributions in dyad discussions can serve as a mediator (M), carrying the influence of the agent intervention methods (X) on the dyad performance (Y). The test was performed using the PROCESS SPSS macro, which employed a bootstrap-based method with bias-corrected confidence estimates (Hayes 2013). The 95 % confidence interval of the indirect effects was obtained with 5000 bootstrap resamples.

Explicit response ratio

To probe into the agent effect on the generation of explicit contributions, we proceeded to mark as 'agent-induced' every explicit contribution stimulated by the agent. A contribution was marked only if it was closely related to an agent intervention, either as a direct response to the agent or as a follow-up comment.

Following the above process, an explicit response ratio (ERR) was calculated for each dyad in the treatment conditions. This ratio was computed by dividing the agent-induced explicit contributions of the dyad with the number of agent interventions appearing in the chat. Thus, the ERR value of a dyad indicated the average number of explicit contributions stimulated by each agent intervention. An independent-samples t-test was conducted to compare the ERRs between the two treatment conditions.

Student opinion

The student opinion questionnaire was used to measure students' perceptions of the collaborative activity and the agent role. Students expressed their opinion about a series of statements using a 5-step Likert scale (1: disagree; 5: agree). The instrument consisted of two parts. The first part recorded students' subjective views on their overall learning experience and the system usability. The second part, available only for the treatment conditions, elicited students' opinions about the conversational agent.

The treatment students also participated in a semi-structured focus group session aiming to collect complementary data about the perceived benefits or drawbacks of the agent intervention methods. Students' responses were transcribed verbatim and analyzed with the constant comparative method (Boeije 2002).

Results

Individual learning

The means and standard deviations of students' pre- and post-test scores are presented in Table 5. The one-way ANOVA comparing students' pre-test scores revealed that the three conditions were comparable regarding students' prior knowledge, $F(2, 93) = .100$, $p = .905$, $\omega^2 = .002$.

The ANCOVA examining the agent impact on students' learning revealed a statistically significant, large difference in students' post-test scores between the conditions, $F(2, 92) = 13.630$, $p = .000$, partial $\eta^2 = .229$. A post hoc analysis, performed with a Bonferroni adjustment, showed that the D treatment condition outperformed significantly the U treatment

Table 5 The students' scores in the pre- and post-test questionnaires

		Pre-test		Post-test	
	n	M	SD	M	SD
Control	32	10.94	5.41	10.13	4.48
U Treatment	32	10.65	3.84	11.95	3.99
D Treatment	32	10.45	4.06	14.01	3.68

(M_{diff} = 2.173, p = .023) and the control condition (M_{diff} = 4.163, p = .000). The control condition had the lowest post-test scores, which was significantly lower than the U treatment condition (M_{diff} = 1.990, p = .043).

Given that students worked in dyads within the research conditions, a nested ANOVA also reported a significant variation in means between the conditions and confirmed that the conditions had a significant contribution to the overall variability in the post-test scores, $F_{condition}(2, 45) = 8.267$, p = .001. As opposed to the condition factor, the effect of dyads nested within research groups was not statistically significant.

Dyad performance in task

After evaluating the answers provided by the dyads to the activity learning question, a Kruskal-Wallis H test indicated a statistically significant difference between the three conditions, $\chi^2(2) = 10.964$, p = .004. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. The post-hoc analysis revealed significant differences in the scores between the control (mean rank = 15.44) and U treatment conditions (mean rank = 27.25) (p = .045) as well as the control and D treatment conditions (mean rank = 30.81) (p = .005), albeit not between the two treatment conditions.

Explicit reasoning in discussion

A total number of 3909 students' contributions were identified in the discussions of all dyads (n = 48, M = 81.44, SD = 15.66). Table 6 presents the overall results of the discourse analysis conducted.

The one-way ANOVA performed on dyad frequency values showed that the frequency of explicit positions varied significantly between the conditions, F(2, 45) = 10.800, p = .000, ω^2 = .290. In particular, the frequency value increased from the control (M = 9.47, SD = 4.16), to U treatment (M = 13.56, SD = 3.02) to D treatment (M = 15.88, SD = 4.53) conditions, in that order. Tukey post hoc analysis yielded two significant differences. More specifically, the mean increase from control to U treatment was statistically significant (M_{diff} = 4.09, p = .015), as well as the increase from control to D treatment (M_{diff} = 6.42, p = .000).

Likewise, the frequency of explicit arguments also varied significantly between the three conditions, F(2, 45) = 7.320, p = .002, ω^2 = .208. The explicit argument frequency increased from the control (M = 3.95, SD = 3.29), to U treatment (M = 7.22, SD = 3.96) to D treatment (M = 8.29, SD = 2.64) conditions, in the same order. Tukey post hoc analysis demonstrated

Table 6 Overall discourse analysis results

		Control (n = 16 dyads)		U Treatment (n = 16 dyads)		D Treatment (n = 16 dyads)	
		Total	Freq. (%)	Total	Freq. (%)	Total	Freq. (%)
1.	Off Task	104	9.50	96	6.41	88	6.69
2.	Repetition	16	1.46	22	1.47	10	0.76
3.	Team Management	200	18.26	288	19.23	338	25.68
4.	Common Understanding	126	11.51	201	13.42	167	12.69
5.	Issue	97	8.86	172	11.48	113	8.59
6.	Position	221	20.18	192	12.82	139	10.56
7.	Argument	186	16.99	219	14.62	152	11.55
8.	Explicit Position	101	9.22	199	13.28	201	15.27
9.	Explicit Argument	44	4.02	109	7.28	108	8.21

that only the increase from control to U treatment ($M_{diff} = 3.27$, $p = .022$) and the increase from control to D treatment ($M_{diff} = 4.34$, $p = .002$) were statistically significant.

Figure 5 presents the distribution of explicit contributions within all dyads in the three conditions. A Kruskal-Wallis H-test indicated that the percentages of the explicit contributions calculated for the 'less explicit' peers varied significantly between the conditions, $\chi^2(3) = 6.305$, $p = .043$. In particular, the average percentage of the 'less explicit' peer was found to increase from the control (29.60 %), to the U treatment (35.90 %), to the D treatment (42.18 %) conditions, in that order. Pairwise comparisons showed a statistically significant difference between the control (mean rank = 18.94) and D treatment (mean rank = 31.19) ($p = .039$), but not in any other condition combination.

Furthermore, multiple regression analyses were performed, investigating whether the frequency of explicit contributions mediated the effect of the agent intervention method on dyad performance. Results revealed that the agent intervention method was a significant predictor of explicit reasoning ($B = 3.680$, $t(94) = 3.070$, $p = .004$) as well as dyad performance ($B = 1.906$, $t(94) = 2.640$, $p = .011$), while explicit reasoning was a significant predictor of

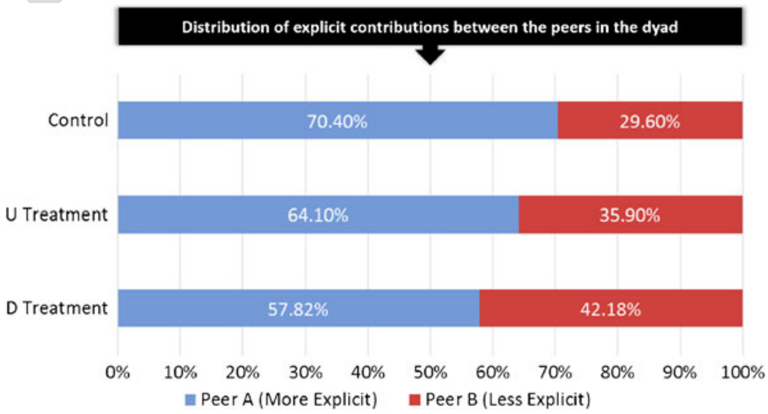

Fig. 5 Explicit contributions balance

dyad performance (B = .317, t(94) = 4.157, p = .000). These results supported the mediational role of explicit reasoning (b = 1.192, 95 % CI [.385, 2.345]) and were consistent with full mediation as the agent intervention method was no longer a significant predictor of students' learning performance after controlling for the mediator (b = .739, t(94) = 1.084, p = .284). Regression coefficients and standard errors are illustrated in Fig. 6.

Explicit response ratio

Table 7 presents major descriptive statistics about the agent interventions displayed in the treatment conditions, as well as the explicit positions and explicit arguments induced by the two agent intervention methods. The independent samples t-test conducted on explicit response ratio (ERR) mean values (Table 7, item 4) showed a statistically significant difference in favor of the D agent intervention method, t(30) = 2.079, p = .046, d = .759.

Student opinion

The examination of the data emerging from the student opinion questionnaires and the focus group session led to the key findings presented in Table 8.

Discussion

In agreement with the findings of our previous study (Tegos and Demetriadis in press), the first set of results demonstrated that the APT agent interventions improved students' learning outcomes significantly. Although students' knowledge levels were comparable prior to the experimental activity, the post-test results revealed that the students who interacted with the conversational agent in the two treatment conditions came out of the collaborative activity with a domain knowledge advantage over the students of the control condition (Table 5). This is corroborated by the results of the student opinion questionnaire, which showed that the students of the control condition perceived the collaborative activity as less helpful for enhancing their domain knowledge than the treatment students (Table 8, item 2). Furthermore, an interesting finding was that the D treatment condition performed significantly better than the U treatment condition in terms of knowledge comprehension. Indeed, the

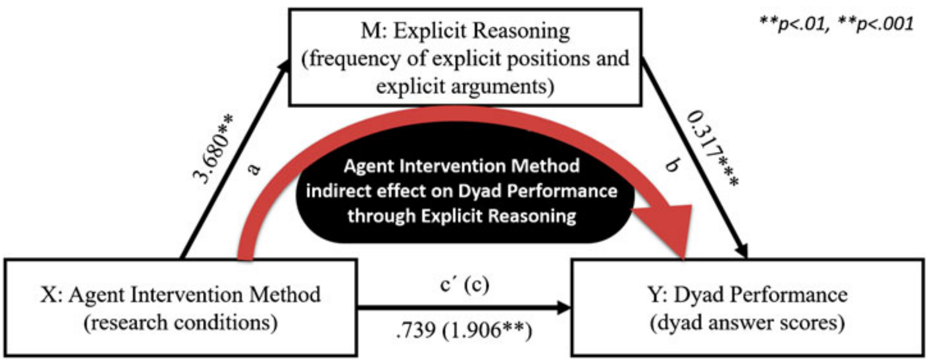

Fig. 6 Mediation diagram

Q1

Table 7 Agent intervention effect on the stimulation of explicit contributions

		U Treatment		D Treatment	
		(*n* = 16)		(*n* = 16)	
		M	SD	M	SD
1.	Agent interventions	2.94	0.77	2.75	0.68
2.	Agent-induced explicit positions	4.00	2.42	4.56	2.16
3.	Agent-induced explicit arguments	2.50	2.19	3.44	1.93
4.	Explicit response ratio (ERR)	2.16	0.88	2.91	0.88

students in the D condition were able to better illustrate their understanding in the post-test as compared to the students of the U condition.

Apart from the agent learning effect measured at individual level, the agent also had a positive impact on dyad performance in the task. More specifically, the dyads in the treatment conditions were found to provide more accurate and comprehensive answers to the learning question of the activity. The answers submitted in the treatment conditions received higher ratings and appeared to be more conceptually solid and complete than the ones of the control condition. No significant differences were reported between the U and D treatment conditions, indicating that the alteration of the agent intervention method in the treatment conditions did not significantly affect dyad performance.

A possible explanation for the above effect may be that the agent urged peers to link their chat contributions more strongly and accurately to the main theoretical principles of the course while co-constructing their dyad answers. Thus, the treatment teams were able to utilize some

Table 8 Student opinion findings

Student opinion questionnaire

1. No major issues were reported concerning the system usability (*n* = 96, M = 4.13, SD = .85) or performance (*n* = 96, M = 4.42, SD = .68). There were no significant differences between the conditions.
2. The D and U treatment students expressed greater agreement (*n* = 64, M = 4.01, SD = .70) than their control counterparts (*n* = 32, M = 3.56, SD = .98) with the statement: "the collaborative activity improved my domain knowledge", U = 1267.5, z = 2.058, *p* = .040, *r* = .210.
3. The U treatment students expressed greater agreement (*n* = 32, M = 3.81, SD = 0.81) than the D students (*n* = 32, M = 3.27, SD = 1.05) with the statement: "the agent questions did not disrupt my discussion with my partner", U = 664.5, z = 2.084, *p* = .037, *r* = .260.
4. The treatment students (*n* = 64) had a fairly positive reaction to the following statements: "the agent questions displayed during the discussion were simple and understandable" (M = 4.30, SD = .69), "the agent questions helped me recall or retrieve useful domain information for the evaluation of the e-shop interface" (M = 3.94, SD = .69), "the timing and the content of the agent questions were consistent with the on-going discussion" (M = 3.98, SD = .68).

Focus group

5. The majority of the treatment students (*n* = 64, F = 76.56 %) stated that the agent interventions helped them resolve the learning task.
6. A group of students from the D treatment condition (*n* = 32, F = 21.88 %) disliked the fact that sometimes the agent did not allow them to submit a response.
7. Some treatment students (*n* = 64, F = 14.06 %) stated that they would like to have the option to temporarily hide an agent intervention.

of the topics discussed throughout the course more effectively in order to bolster their arguments and better support the claims presented in their conceptually richer answers. Overall, the conversational agent seemed to play a critical role in supporting accountability by asking students to consider themselves responsible for the accuracy and validity of their claims, and “be committed to getting the facts right” (Wolf et al. 2005, p. 6). Even though many students assume that there is no need to explicitly discuss what is common knowledge in the community, encouraging students to make their knowledge sources explicit is considered vital in academic settings for increasing collective reasoning levels and improving collaborative learning outcomes (Michaels et al. 2010; Papadopoulos et al. 2013).

The discourse analysis revealed that the agent interventions had a significant effect on the levels of explicit reasoning exhibited during the collaborative activity. In particular, the frequencies of explicit positions and explicit arguments were measured to be substantially higher for the treatment conditions than the control condition (Table 6, items 8 and 9). Considering the number of explicit contributions identified as agent-induced (Table 7, items 2 and 3), we argue that the increased generation of students’ explicit contributions is largely owed to the activation of the agent interventions, which promoted students’ sound reasoning by pressing them for clear statements backed up by concrete evidence. This is consistent with Dyke et al.’s (2013) findings, suggesting that an agent prompting students to follow academically productive practices can amplify students’ expression of scientific reasoning.

The mediation analysis conducted in the study revealed that the display of explicit reasoning played a significant mediating role, carrying the influence of the agent intervention method on dyad performance. As illustrated in Fig. 6, the agent interventions significantly affected explicit reasoning (a path), explicit reasoning had a significant unique effect on dyad performance (b path), agent interventions significantly affected dyad performance in the absence of explicit reasoning (c’ path), and the effect of the agent on dyad performance shrunk upon the addition of explicit contributions frequency to the model (c path). Thus, our proposed model suggests that the impact of an APT agent on dyad performance varies based on how well the agent can trigger conversational interactions whereby learners explicitly display their reasoning on conceptual knowledge.

The explicit response ratio (ERR) metric revealed that the D agent intervention method was more efficient in stimulating subsequent explicit contributions from the students than the U method (Table 7). On the basis of our observations and evidence obtained throughout the discourse analysis phase, when the agent addressed a specific student in the D condition it seemed that the student felt personally responsible for giving a comprehensive response to the agent. In fact, the peers addressed sometimes asked for the assistance of their partners, who often commented on the agent intervention and provided additional information. Directing prompts to individual learners by an agent seems to be a feasible way to reduce diffusion of responsibility and facilitate equal participation in reasoning processes without setting up specific incentive structures (cf. Slavin 1992). The way the agent was deployed in this experimental condition fully aligns, however, with principles of individual accountability and interdependence. Addressing one specific student was not a covert process. Hence, both learners could understand how the agent implemented shared dialogue rules. At times, the above behavior seemed to result in a transactive form of dialogue, where students built on each other’s reasoning in order to provide a more comprehensive response to the agent. A future discourse analysis focusing on the identification of transactive contributions could provide valuable insights in this matter. Still, it appears that the directed agent approach acknowledges

Q1

the situational characteristic of transactivity. While any non-adaptive prompting for transactive dialogue may turn into an additional routine task for learners, the agent flexibly calling on the respective ‘less explicit’ student to respond helps learners to simultaneously connect to peer input as well as to the theoretical principles to be learned.

Some peers in the U condition appeared to have little coordination and occasionally did not communicate et al.l with each other before responding to the agent question. In most of these cases, the student who triggered the agent intervention by discussing an important task-related concept took the initiative to respond to the agent question without discussing the matter with their partner. As expected, this behavior resulted in some relatively unbalanced discussions, where the most active student explicated their thoughts far more frequently than their partner. This is supported by examining the distribution of explicit contributions between the learning partners (Fig. 5). As revealed by our analysis, the discussions in the D treatment condition were far more balanced in terms of explicit reasoning than those in the U treatment and even more so in the control condition. According to our viewpoint, the directed interventions of the agent promoted more equitable student participation by occasionally taking control of turns at talk. We consider this implicit turn-taking strategy to be associated with the better individual learning outcomes of the D condition since the D agent interventions encouraged the ‘less explicit’ partners, who might have remained relatively inactive in the U condition, to actively participate and explicitly display their reasoning.

Even though most students had an overall positive perception of the agent (Table 8, items 4 and 5), the students in the D condition perceived agent interventions as more disruptive than those of the U condition (Table 8, item 3). Although further research is required to understand the implications of this perceived increase in the interruption effect of the D intervention method, this finding may relate to the fact that the D interventions introduced more situational constraints than the U interventions by imposing students to follow a specific student-agent interaction protocol. With individual students being put on the spot, students’ perceptions of freedom and, thus, their opinion of the agent may have been negatively affected, given that turn-taking strategies are known to have a significant impact on perceived agent personality, attitude and handling of interruptions (Cafaro et al. 2016). In future research, learners’ perception of agents and prompts need to be investigated further through a more qualitative analytic approach. While we have found that learners made sense of and followed agent instructions in the lab scenario, there is a need to develop an insight into what criteria and circumstances play into how learners interpret agent instructions. Nevertheless, considering that collaborative knowledge construction in unstructured chat sessions relies on the successful coordination of peers’ conversational turns (Oehl and Pfister 2010), we argue that the D agent interventions structured student-agent interactions in a robust manner that facilitated group awareness and increased dialogue coherence.

Despite the promising findings of this study, its limitations should be taken into account as well. First, it should be noted that only after further research can the findings relating to the increased efficacy of the directed intervention mode be generalized across different group sizes and task characteristics, since the agent impact may substantially vary over these parameters. For instance, although directed interventions may be more appropriate for relatively simple tasks, in a complex problem-solving activity where participants tend to work on different parts of the task, an undirected intervention may be more efficient than a directed one by allowing the more involved student - the one currently working on the part pertaining the intervention - to address the agent question. Furthermore, another fact that should be considered while interpreting these findings is that all participants were aware of their discussions being monitored. This has probably altered the conversational behavior of treatment students, who may have responded to agent interventions more systematically than they would have in a more informal learning setting, as for example in the context of a massive open online course (MOOC). Lastly, it is all too clear that the conversational agent used in this study could only display simple prompts without possessing the intelligence required to engage in full-fledged discussions with the learners. Still, this is in line with our broad research objective of developing easily configurable and deployable agents, which can operate in diverse educational contexts with substantial learning benefits.

In closing, we would like to 'zoom out' and comment on the potential fruitfulness of the research line of this study. It is clear that further studies need to explore the design space of APT agents, probe into interesting dimensions of agent-induced peer interactions and provide evidence on how agent effectiveness may vary on the basis of specific design decisions. In broader terms, however, we see as important that teacher-verified strategies (beyond APT) could be modeled and integrated in e-learning environments providing the basis for the development of domain-independent pedagogically 'skillful' agents.

Conclusion

Despite the above limitations, this study provides adequate evidence on the potential benefits of unsolicited APT agent interventions that attempt to promote accountability to accurate knowledge by encouraging students to build on their prior knowledge in order to support their claims and arguments. It is suggested that such agent interventions may enhance students' learning, increase the level of explicit reasoning exhibited during students' discussions and improve the in-task performance of dyads working online in higher education settings. Interestingly, the increase in explicit reasoning levels seems to mediate the positive effect of the agent interventions on dyad performance. Furthermore, the agent impact on individual learning appears to be amplified when the agent employs a directed intervention method targeting a particular peer, rather than an undirected intervention method, addressing both peers in a dyad simultaneously. In a similar manner, the efficacy of the agent in triggering explicit reasoning processes and engaging students in constructive interactions seems to be higher for the directed intervention method as compared to the undirected method.

Despite these promising study findings, more research is required in order to investigate how a series of enigmatic factors, such as the task nature and complexity, the maturity of students, and the nature of the discipline being learned, may or may not drastically affect agent efficacy. Future studies could be conducive to the exploration and formalization of such factors in an attempt to amplify the pedagogical effectiveness of conversational agents operating in a collaborative learning context. These studies could also enlighten the research community on the potential benefits and shortcomings of employing specific intervention techniques, such as the delivery of privately directed interventions, i.e. displayed only to a group member instead of the public group chat. In this perspective, we perceive our work to have established an argument in favor of further systematic research on APT agents from a quantitative as well as a qualitative methodological standpoint.

Acknowledgments We are appreciative of Fotini Bourotzoglou's contribution to this work.

Q1

References

Adamson, D., & Rosé, C. P. (2013). Academically Productive Talk: One Size Does Not Fit All. In *Artificial Intelligence in Education (AIED) 2013 Workshops Proceedings* (p. 51–60).

Adamson, D., Ashe, C., Jang, H., Yaron, D., & Rosé, C. P, (2013). Intensification of group knowledge exchange with academically productive talk agents. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings* (vol. 1, pp. 10–17).

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education, 24*(1), 92–124.

Asterhan, C. S., & Schwarz, B. B. (2016). Argumentation for learning: well-trodden paths and unexplored territories. *Educational Psychologist, 51*(2), 164–187.

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality and Quantity, 36*(4), 391–409.

Brandom, R. (1998). *Making it explicit: reasoning, representing, and discursive commitment*. Harvard University Press.

Cafaro, A., Glas, N., & Pelachaud, C. (2016). The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions. In J. Thangarajah, K. Tuyls, C. Jonker, & S. Marsella (Eds.), *15th International Conference on Autonomous Agents and Multiagent Systems* (pp. 911–920).

Chi, M. T. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105.

Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of Computer Assisted Learning, 23*(1), 1–13.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics, 6*, 241–252.

Dyke, G., Adamson, D., Howley, I., & Rosé, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *Learning Technologies, IEEE Transactions on, 6*(3), 240–247.

Fischer, F., Kollar, I., Mandl, H., & Haake, J. (Eds.) (2007). *Scripting computer-supported communication of knowledge. Cognitive, computational and educational perspectives*. New York: Springer.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*(1), 56–66.

Goodman, B. A., Linton, F. N., Gaimari, R. D., Hitzeman, J. M., Ross, H. J., & Zarrella, G. (2005). Using dialogue features to predict trouble during collaborative learning. *User Modeling and User-Adapted Interaction, 15*(1–2), 85–134.

Gulz, A., Haake, M., Silvervarg, A., Sjödén, B., & Veletsianos, G. (2011). Building a social conversational pedagogical agent: design challenges and methodological approaches. In D. Perez-Marin & I. Pascual-Nieto (Eds.), *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices* (pp. 128–155). IGI Global.

Harrer, A., McLaren, B. M., Walker, E., Bollen, L., & Sewall, J. (2006). Creating cognitive tutors for collaborative learning: steps toward realization. *User Modeling and User-Adapted Interaction, 16*(3–4), 175–209.

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. New York, NY: Guilford Press.

Hmelo-Silver, C. E. (2013). Multivocality as a tool for design-based research. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions* (pp. 561–573). New York: Springer.

Howley, I., Kumar, R., Mayfield, E., Dyke, G., & Rosé, C. P. (2013). Gaining insights from sociolinguistic style analysis for redesign of conversational agent based support for collaborative learning. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions (pp. 477–494)* .Springer US

Huitt, W. (2011). Bloom et al.'s taxonomy of the cognitive domain. In *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University.

Kumar, R., & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies, 4*(1), 21–34.

Liu, C. C., & Tsai, C. C. (2008). An analysis of peer interaction patterns as discoursed by on-line small group problem-solving activity. *Computers & Education, 50*(3), 627–639.

Ludvigsen, S., & Mørch, A. (2010). Computer-supported collaborative learning: basic concepts, multiple perspectives, and emerging trends. In B. McGaw, P. Peterson, & E. Baker (Eds.), *The international encyclopedia of education 3rd edition, volume 5* (pp. 290–296). Oxford, UK: Elsevier.

Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: a review of the field. *IEEE Transactions on Learning Technologies, 4*(1), 5–20.

Michaels, S., & O'Connor, C. (2013). Conceptualizing talk moves as tools: professional development approaches for academically productive discussion. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through talk and dialogue*. Washington DC: American Educational Research Association.

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*(4), 283–297.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick L. B. (2010). *Accountable Talk Sourcebook: For Classroom That Works*. University of Pittsburgh Institute for Learning. Retrieved on May 1, 2016, from http://ifl.pitt.edu/index.php/download/index/ats.

Noroozi, O., Teasley, S., Biemans, H. A., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning, 8*(2), 189–223.

Oehl, M., & Pfister, H. R. (2010). E-collaborative knowledge construction in chat environments. *E-Collaborative Knowledge Construction: Learning from Computer-Supported and Virtual Environments*, 54–72.

Papadopoulos, P. M., Demetriadis, S., & Weinberger, A. (2013). 'Make it explicit!': improving collaboration through increase of script coercion. *Journal of Computer Assisted Learning, 29*(4), 383–398.

Preece, J., Sharp, H., & Rogers, Y. (2015). *Interaction design: beyond human-computer interaction*. John Wiley & Sons.

Resnick, L. B., Michaels, S., & O'Connor, C. (2010). How (well structured) talk builds the mind. In R. Sternberg & D. Preiss (Eds.), *From genes to context: new discoveries about learning from educational research and their applications* (pp. 163–194). New York: Springer.

Rus, V., D'Mello, S., Hu, X., & Graesser, A. (2013). Recent advances in conversational intelligent tutoring systems. *AI Magazine, 34*(3), 42–54.

Sionti, M., Ai, H., Rosé, C. P., & Resnick, L. (2012). A framework for analyzing development of argumentation through classroom discussions. In N. Pinkwart & B. McLaren (Eds.), *Educational Technologies for Teaching Argumentation Skills (pp. 28–55)*. Bentham Science Publishers.

Slavin, R. E. (1992). When and why does cooperative learning increase achievement? Theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups. The theoretical anatomy of group learning* (pp. 145–173). Cambridge: Cambridge University Press.

Sohmer, R., Michaels, S., O'Connor, M. C., & Resnick, L. (2009). Guided construction of knowledge in the classroom. In B. Schwarz, T. Dreyfus, & R. Hershkowitz (Eds.), *Transformation of knowledge through classroom interaction* (pp. 105–129). New York: Taylor and Francis.

Stahl, G. (2015). Computer-supported academically productive discourse. *Socializing intelligence through academic talk and dialogue*, 213–224. Retrieved on May 1, 2016 from http://gerrystahl.net/pub/lrdc2015.pdf.

Stahl, G., & Rosé, C. P. (2011). Group cognition in online teams. In E. Salas & S. M. Fiore (Eds.), *Theories of team cognition: cross-disciplinary perspectives*. New York, NY: Taylor & Francis.

Stahl, G., Cress, U., Ludvigsen, S., & Law, N. (2014). Dialogic foundations of CSCL. *International Journal of Computer-Supported Collaborative Learning, 9*(2), 117.

Tegos, S. (2016). *Web-based conversational agents for collaborative learning support*. Aristotle University of Thessaloniki: Unpublished Doctoral dissertation.

Tegos, S., & Demetriadis, S. (in press). Conversational agents improve peer learning through building on prior knowledge. *Educational Technology & Society.*

Tegos, S., Demetriadis, S., & Karakostas, A. (2014). Conversational Agent to Promote Students' Productive Talk: The Effect of Solicited vs. Unsolicited Agent Intervention. In *IEEE 14th International Conference on Advanced Learning Technologies* (pp. 72–76).

Tegos, S., Demetriadis, S., & Karakostas, A. (2015). Promoting academically productive talk with conversational agent interventions in collaborative learning settings. *Computers & Education, 87*, 309–325.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2016). Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: a Meta-Analysis. *Educational Psychology Review*, 1–35.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology, 79*(1), 1–28.

Weinberger, A. (2011). Principles of transactive computer-supported collaboration scripts. *Nordic Journal of Digital Literacy, 6*(3), 189–202.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: concepts and assessment. *Learning and Instruction, 17*(4), 416–426.

Wolf, M. K., Crosson, A. C., & Resnick, L. B. (2005). *Accountable talk in reading comprehension instruction.* Regents of the University of California.