Running head: TEACHABLE AGENTS AND THE PROTÉGÉ EFFECT

Teachable Agents and the Protégé Effect:

Increasing the Effort Towards Learning

Catherine Chase, Doris B. Chin, Marily Oppezzo, & Daniel L. Schwartz

Stanford University

In press. Journal of Science Education and Technology.

Corresponding Author:

Catherine Chase
Stanford University School of Education
485 Lasuen Mall
Stanford, CA 94305
650.723.2109
cchase@stanford.edu

ABSTRACT

Betty's Brain is a computer-based learning environment that capitalizes on the social aspects of learning. In Betty's Brain, students instruct a character called a Teachable Agent (TA) which can reason based on how it is taught. Two studies demonstrate the protégé effect: students make greater effort to learn for their TAs than they do for themselves. The first study involved 8th-grade students learning biology. Although all students worked with the same Betty's Brain software, students in the TA condition believed they were teaching their own TAs, while in another condition, they believed they were learning for themselves. TA students spent more time on learning activities (e.g. reading) and they also learned more. These beneficial effects were most pronounced for lower achieving children. The second study used a verbal protocol with 5th-grade students to determine the possible causes of the protégé effect. As before, students learned either for their TAs or for themselves. Like study 1, students in the TA condition spent more time on learning activities. These children treated their TAs socially by attributing mental states and responsibility to them. They were also more likely to acknowledge errors by displaying negative affect and making attributions for the causes of failures. Perhaps having a TA invokes a sense of responsibility that motivates learning, provides an environment in which knowledge can be improved through revision, and protects students' egos from the psychological ramifications of failure.

Teachable Agents and the Protégé Effect:
Increasing the Effort Towards Learning

The interactive potential of the computer naturally draws comparisons to social behavior. For example, the Turing (1950) test proposed that if a human interacts with a computer, and the human believes the computer is a person, then the computer has achieved human intelligence. A number of computer programs were engineered to challenge the validity of the Turing test. ELIZA, for instance, successfully impersonated the dialog of a Rogerian therapist, but the computer used such simple rules that it would be absurd to consider it truly intelligent (Weizenbaum, 1976). Whether or not the Turing test is adequate for deciding the intelligence of a computer, it is useful to note that the Turing test is really about the social behavior of the computer. There could have been other tests of human intelligence; for example, could the computer learn language? But, instead the test assessed whether people would treat the computer as a social entity. Here, we use the natural social attractions of the computer to improve students' science learning.

Computers readily draw forth people's social schemas. Even when they explicitly know they are interacting with a computer, people will behave in socially appropriate ways (Reeves & Nass, 1998). People's tendency to attribute social intelligence to computers has fueled the creation of graphical worlds that comingle human and computer intelligence. Examples include Second Life, the Sims, and World of Warcraft – where people interact with graphical characters that may represent a live person or a computer character. These human-computer hybrids not only boost natural social inclinations, they can also produce novel social configurations that sustain unusual psychological states. For instance, game players can program graphical

characters to act (and interact) in virtual social worlds even when the players are no longer at their computer.

The novel social configuration presented here involves software agents that blend student and computer intelligence. We have created a computer-based learning environment that features a Teachable Agent (TA) -- a graphical computer character that students teach. The TA uses artificial intelligence to learn and reason about what it has been taught. Teachable Agents are a hybrid; they reflect their owners' knowledge, yet have minds of their own. This social arrangement has benefits for learning. For example, students are likely to adopt their TAs' reasoning methods (Schwartz et al., in press). Here, we focus on the motivational consequences.

We begin with a brief review of agents and avatars, which are the two main classes of virtual characters used in educational applications. We then introduce Teachable Agents, which combine properties of agents and avatars. This sets the stage for two studies that demonstrate what we term the protégé effect: students make greater effort to learn for their TAs than they do for themselves. The first study produces this effect, even when the only difference between conditions is whether students *believe* they are teaching TAs or not. The second study shows the social nature of the interaction with the TA and how it contributes to the protégé effect. We conclude with some initial thoughts on the role of TAs in creating a distinctly social set of motivations to learn, which are supported by an ego-protective buffer, an incrementalist approach to learning, and a sense of responsibility.

Learning and Motivation with Agents, Avatars, and Hybrids

Interactive computer characters traditionally come in one of two forms: avatar and agent (Bailenson & Blascovich, 2004). An avatar is a character that represents and is controlled by a human. For example, in a video game, the characters manipulated by the players are avatars. In

contrast, an agent is a character controlled by the computer. When people play a hockey video game by themselves, they each control their own avatars, while the computer controls the other players (agents) on the team. One of the interesting things about these computer games is that the users can jump from character to character, so they control whichever player happens to have the hockey puck. This is a nice example of a novel social configuration that computers support.

Agents and avatars each have advantages for education. A number of useful learning situations can be created by agents (for a nice collection of instances, see Baylor, 2007). For example, agents can provide role models for how to think or act. Ryokai, Vaucelle, and Cassell (2003) used an embodied conversational agent named Sam to engage children in collaborative story-telling. Children who interacted with Sam adopted his conversational behaviors and used more advanced narrative skills than children who conversed with peers. Another type of agent is a pedagogical agent, which provides advice to learners. For instance, Shimoda, White, and Frederiksen, (2002) used a panoply of agents to deliver meta-cognitive tips during scientific inquiry. Clarebout, Elen, Johnson, and Shaw (2002) have created a typology of pedagogically-relevant agent behaviors such as showing, explaining, and questioning.

Agents can also be used to improve motivation. Lester et al. (1997) experimented with five varieties of Herman the Bug, a pedagogical agent who worked with middle school students as they designed a plant. In a condition where the agent gave no advice but exhibited social behaviors of encouragement, students gave the agent high ratings on entertainment value and chose to have Herman help them with homework. Lester et al. dubbed this the *persona effect*, claiming that the socialness of the agent helped to engage students with the software. Similarly, Baylor and Kim (2005) found that pedagogical agents equipped with encouraging dialogue were perceived as more motivating and showed a moderate trend for enhancing student self-efficacy.

Like agents, avatars (which humans control) may also have benefits for learning. For example, people may learn to take on the attributes of their avatars. Yee and Bailenson (2007) termed this the *proteus effect*. In one study, participants were assigned to use either a tall or short avatar. They then played a negotiation game with another person in virtual reality. The people who played as the tall avatar were tougher negotiators and were more likely to come out ahead. Presumably, they took on the stereotype that height confers power and authority. This tendency for adoption has educational potential, when the attributes to be adopted are useful dispositions for learning.

Avatars can also motivate students to take risks. If the avatar makes a mistake, the user does not necessarily suffer the consequences. When getting checked into the boards in a virtual hockey game, the players not only do not get hurt, but they can also "laugh it off." Just as computer simulations of nuclear fusion are physically safer than the real thing (Perkins et al., 2006), avatars can make it psychologically safer to try new things, without experiencing the real consequences of failure.

A hybrid agent/avatar blends the properties of an agent and an avatar. It is a character that includes a bit of the computer and bit of the human user. A key element of a hybrid agent/avatar is its ability to behave without explicit human control while still reflecting prior interactions with a human user. A growing number of hybrids vary the mix of human dependence and independence. Some applications have the user try to "program" a character so it lives and acts exactly the way the user intends (Gerhard, Moore, & Hobbs, 2004; Imbert & de Antonio, 2000). For example, in The Sims, a popular commercial game, computer characters behave based on the attributes supplied by their users plus some amount of their own apparent "free will."

Another example is the Tamagotchi -- a digital pet housed in a small, egg-shaped computer. Children are responsible for feeding, cleaning, and nurturing their Tamagotchis. The pets respond and grow based on the children's care. From the children's point of view, Tamagotchis exhibit behaviors that are both independent and dependent. Children (especially girls) find the responsibility and nurturing highly motivating (Pesce, 2000). The research presented here shows that a sense of responsibility towards a hybrid can lead to educationally relevant outcomes as well.

A Teachable Agent (TA) is a "sentient" hybrid agent/avatar that has been specifically designed for educational outcomes. The TA engages learners in a teacher-pupil metaphor and takes on the role of protégé. The student teaches the TA, so the TA is dependent on the student. At the same time, the TA contains artificial intelligence that allows it to behave independently. For instance, the TA can reason, answer questions, and complete various assessments based on how it was taught. Moreover, a TA possesses the educational benefits of both agents and avatars. Like an agent, a TA provides an independent social presence that motivates students to interact with it, plus it offers new models of thinking and reasoning. Like an avatar, the TA has properties that students can adopt, without the intellectual risks that come with doing something on one's own.

[FIGURE 1 GOES HERE]

A Teachable Agent Called Betty's Brain

There are several types of Teachable Agent software (see Schwartz, Blair, Biswas, Leelawong, & Davis, 2007); here we focus on Betty's Brain. Betty was designed to model chain-like mechanisms of cause and effect relationships. For example, when the brain's temperature set point rises, several multi-step pathways cause the body's temperature to increase

and develop a fever (see Figure 1). Betty is especially relevant to science domains where long chains of qualitative causes are a useful way to explain phenomena. Biology content like food webs and ecosystems, bodily systems, and global warming are well-modeled by Betty's architecture.

Before teaching in Betty's Brain, each student names and designs the appearance of her own TA (Betty's Brain is the name of the software; students create characters for themselves). A student then teaches her TA by creating a concept map of nodes connected by qualitative causal links; for example, 'heat release' decreases 'body temperature'. The map fancifully symbolizes the interior of the TA's brain. Once taught, a TA can answer questions. For instance, Betty includes a simple query feature. Using basic artificial intelligence techniques, the TA animates its reasoning process by successively highlighting each node and link in a causal chain (see Biswas, Leelawong, Schwartz, Vye, & TAG-V, 2005). In Figure 1, the TA uses the map it was taught to answer the query, "If blood flow to skin increases, what happens to body temperature?" A student can trace her TA's reasoning, and then remediate its knowledge (and her own) if necessary. A TA always reasons logically, but depending on the nodes and links it was taught, it will reach a right or wrong answer.

Betty's Brain is not meant to be the only means of instruction, but rather to provide a way for students to organize and reason about content they have learned in the classroom (Schwartz, Blair, Biswas, Leelawong, & Davis, 2007). Betty is intended to complement many styles of instruction, not replace them. One of her complementary strengths is feedback. Betty comes with a number of software options that provide feedback in various forms, some of which can spark classroom discussion. The option shown in Figure 2a enables a teacher to project multiple TAs' maps using a classroom projector. The teacher can ask the same question of all the TAs

simultaneously, then zoom in to focus the discussion on one or two maps. Figure 2b shows the All Possible Questions (APQ) matrix – a tool that asks the TA every possible question. It then compares the answers of the TA with those of a hidden, pre-programmed expert map to produce a grid that indicates which questions the TA got right and wrong.

[FIGURE 2 GOES HERE]

Several of Betty's attributes were designed to encourage students to treat their TAs as social beings. For instance, a TA can draw inferences from questions, take quizzes, play games, and even comment on its own knowledge (depending on the configuration of the software). Betty's Brain also comes with narratives and graphical elements to help support the mindset of teaching. Finally, each student can customize her TA's appearance and give it a name, which makes her TA more personal than a sterile, generic computerized icon. In reality, students are simply programming their TAs in a high-level graphical language, and children know the computer is not really alive. Nevertheless, as we demonstrate in Study 2, students suspend disbelief enough to treat the computer as possessing knowledge and feelings (e.g., Reeves & Nass, 1998; Turkle, 1995).

One of a TA's most social elements is its ability to externalize its thought processes. When a TA animates its reasoning on the screen, it literally makes its "thinking" visible. A study with 6th-graders indicated that students do learn from the TA's overt model of causal reasoning (Schwartz et al., in press). In one condition, students worked with their TAs to organize what they had learned from various readings, films, and hands-on activities. In another condition, students learned the same content, but worked with a commercial concept mapping program called Inspiration. Students took periodic paper and pencil tests across three weeks of a curriculum about global warming. Over time, the TA students increasingly outperformed the

Inspiration students, and TA students demonstrated the greatest advantage on questions that required longer chains of causal inference. These results indicate that students adopted the reasoning process modeled by the TAs in Betty's Brain.

Other studies have also found learning benefits when students work with Betty's Brain. A two-month study had 5th-graders learn river ecology (Wagster, Tan, Wu, Biswas, & Schwartz, 2007). In the Teach condition, each student taught Betty (in this study all students taught the same graphical character called Betty rather than creating their own TAs). In the Being-Taught condition, Betty's image was replaced with a "mentor agent" named Mr. Davis. In the Being-Taught condition, students also created maps. When a student asked a question of her map, the mentor agent traced through the map (in exactly the same way that Betty did for students in the Teach condition). Thus, the primary difference between conditions was quite subtle – the mindset of teaching versus being taught. Students in the Teach condition produced more accurate concept maps. The benefits also transferred to a unit on land ecology, when the students were no longer in their respective treatments. Students who had been in the Teach condition again made better concept maps.

Overview of the Motivation Studies

Given evidence of cognitive gains, the current research was designed to get a closer look at the motivational properties of Teachable Agents. The first study demonstrates the protégé effect: students are willing to work harder to learn for their TAs than for themselves, and this is especially true for low-achieving students. The second study finds that students treat their TAs as social, thinking beings. Students closely monitor and take responsibility for their TAs' failures, which motivates them to revise their own understanding so they can teach better. The studies were short in duration, only one to three hours, so there was minimal expectation of

finding learning differences. Instead, the research focused specifically on affective elements that may have contributed to the learning benefits found in earlier research.

In the current studies, one of Betty's features was particularly important – the Triple-A-Challenge Gameshow. The Gameshow is an online environment where multiple TAs, each taught by a different student, can interact and compete with one another (Figure 3). Students can log on from home to teach their TAs (by accessing the Betty software), chat with other students, and eventually have their TAs play in a game. During game play, the host poses questions of the form, "If X increases/decreases, what happens to Y?" After each question, the student wagers from 0 to 500 points, and the TA answers based on what it has been taught. Then, the host reveals the correct answer and awards points. Students normally play the Gameshow in rounds, with each round consisting of about six questions, and subsequent rounds including more difficult questions (i.e. requiring longer chains of reasoning).

[FIGURE 3 GOES HERE]

The Gameshow was developed to make homework more interactive, social, and fun. In one study, Schwartz et al. (in press) found high levels of homework compliance when students used the Gameshow with TAs, and the Gameshow prepared students to learn related content in class over the next few days. In the current studies, the Gameshow was not used for homework, but instead used in the classroom in Study 1, and for individual sessions in Study 2. In both studies, the manipulation was whether the character in the software represented a TA, or whether the character was an Avatar that represented the student. In the TA condition, the TAs answered the host's questions while students wagered on their protégés. In the Avatar condition, the students answered the host's questions and wagered on themselves.

Our predictions were simple. Students in both conditions would be engaged by the novelty of the technologies, especially in the context of school. However, the TA would yield a specific type of engagement. Students would be more motivated to learn for their protégés than for themselves. Specifically, they would spend more time reading and revising their knowledge. Furthermore, this motivation would be partially driven by the "make believe" that their TAs have thoughts and feelings and by the sense of responsibility students would develop towards their digital pupils.

Study 1: The Protégé Effect

One of the interesting benefits of new technologies is that they permit "clean tests" that are hard to match in the physical world. For example, most research that claims to have demonstrated a benefit of social interaction on learning has been confounded by the many differences between a social and non-social interaction (e.g., Kuhl, Tsao, & Liu, 2003; Moreno, Mayer, Spires, & Lester, 2001). For example, demonstrating that an individual learns more by working in a group than working alone may be attributed to the increase of information exchange and not to the fact that the individual was in a social exchange. Chi, Roy, and Hausmann (2008), recognizing this distinction, proposed that learning from social interaction may be due to the same processes involved in self-explanation (e.g. elaborating on a topic by explaining to oneself).

New technologies provide fresh possibilities for untangling these matters (Blascovich, Loomis, Beall, Swinth, Hoyt, & Bailenson, 2002). For example, Okita, Bailenson, and Schwartz (2007) had adults interact with a graphical character in immersive virtual reality. The participants and the character discussed the biological mechanisms that sustain a fever. The interactions were tacitly scripted so that each participant said and heard the same things at the

same times. The experimental manipulation was simply whether the participants were told that the character was a computer agent or that the character represented a person in another room (in reality, it was always a computer program). When participants thought the character was the avatar of another person, they learned more about fever mechanisms and were able to apply their learning to new situations. They also showed higher levels of arousal as measured by skin conductance, and this arousal was correlated with how well they had learned. Even though all the information and behaviors were held constant, the mere belief of a social interaction led to better learning. More recent research (Chen, Shohamy, Ross, Reeves, & Wagner, 2009) suggests that believing an experience is social activates the brain's reward circuitry, which helps to cement the learning of new associations (e.g., Davachi, Mitchell, & Wagner, 2003).

The current study also adopts a "mere belief" manipulation. In the Okita et al. study, social was operationalized as interacting with another person versus interacting with a computer. In the current studies, social is operationalized as other versus self, or to be more precise, protégé versus self. On the first day of the study, the sole difference between conditions was whether students thought they were teaching their TAs or making concept maps for their own learning. Ideally, the results of this clean comparison will illuminate some of the mechanisms that underlie the benefits of learning-by-teaching more generally (e.g., Renkl, 1995), and not just those found in this particular technology environment.

The study was designed to examine whether students would produce greater effort to learn for their TAs than for themselves. In addition to the direct comparison of treatments, a second question was whether the TA treatment would have positive effects for lower achieving students. In prior implementations of the Teachable Agent software, teachers reported that their lower achieving students seemed to benefit especially from the Teachable Agents. It is

conceivable that TAs may protect the students from being wrong themselves (it was their TAs and not them who got it wrong). Moreover, the TA provides a new way to learn. Students who have not had much success with traditional approaches may find this a welcome change. In either case, it is important to gather direct evidence regarding the teachers' observations.

[FIGURE 4 GOES HERE]

In this study, 8th grade students used Betty's Brain over two 50-minute class periods. During this time, they learned how to use the software, read about fever mechanisms, created and tested their concepts maps, chatted with each other online, and played the Gameshow. Figure 4 is a screen shot of the expert map that was used by the software to judge the TA's or student's knowledge (depending on condition). Students did not see this map. It is included here to show the complex interrelationships represented in the content. To learn about the mechanisms of a fever, students could access a one-page reading document through the Gameshow environment (see appendix for fever passage).

[FIGURE 5 GOES HERE]

Figure 5 shows the time course of the study. The key points of difference between conditions are underlined. In both conditions, students used Betty's Brain to create concept maps. In the TA condition, the characters represented the students' pupils, and students were told they were making and testing concept maps to help their protégés learn. In the Avatar condition, the characters represented the students themselves, and they were told to use the concept mapping activities to help themselves learn. In either case, the software was intelligent and could answer questions based on the maps the students had created. For example, students in either condition could submit their maps to a quiz feature that scored the maps on a set of questions. The difference on Day 1 was only in the cover story, and students in the TA condition

did not know their TAs would be playing in a Gameshow. On Day 2, the manipulation was less subtle. All students played the Gameshow. Students in the Avatar condition answered questions for themselves, while students in the TA condition watched their TAs answer the Gameshow questions.

Methods

Participants. Sixty-two 8th-graders, drawn evenly from four different classes, participated in the study. The children attended a diverse San Francisco Bay Area middle school, including 35% Asian, 25% Hispanic, 22% Filipino, 11% White, and 4% African-American students. Thirty-seven percent of the students qualified for free or reduced lunch programs. All students had the same 8th grade science teacher. Halves of each of the classes were assigned intact to treatment, so that half of two classes completed the Avatar condition and half of two classes completed the TA condition (the other class halves completed an entirely different study). Stratified random sampling of the children from each class ensured that pre-existing achievement scores were the same across the two conditions (M_{Avatar} = 78.5, SD = 6.5, M_{TA} = 78.2, SD = 8.5). Achievement was based on the cumulative score the children had earned over the prior eight months in science class. Nevertheless, issues of intact assignment need to be kept in mind when attempting to generalize the results.

Design and Procedures. There were two conditions: TA and Avatar. In the TA condition, the graphical characters represented the students' protégés; students used the mapping software to teach about fever mechanisms; the students answered the questions *themselves* in the Gameshow on Day 1; and on Day 2 the students' TAs answered the questions. In the Avatar condition, the graphical characters represented the students; students used the mapping software to learn about fever mechanisms themselves; and the students themselves answered the Gameshow questions

on both days. Since two to four students (or TAs) played against each other at once, there were up to nine different games going at the same time within a class.

On Day 1, all students logged on to the Triple-A Gameshow system. They learned to customize their TAs, chat, access reading resources, create causal maps, ask questions of the maps, and use the quiz feature. Students received the relevant fever nodes, and their task was to link them up using the reading passage as a guide. The manipulation was given in the instructions and framing of the concept mapping software: students were either making concept maps for themselves or to teach their TAs. The last ten minutes were devoted to showing students the Gameshow, how to join a game, wager, and answer Gameshow questions. At this time, all students played the game in self-answering mode.

On Day 2, all students logged on to play a preliminary game. Students in the Avatar condition continued to answer the Gameshow questions themselves. However, unlike the day before, students in the TA condition now had the questions answered for them by their TAs. After this preliminary round, all students received a brief tutorial on "best practices" for making a map, followed by eight minutes of map revision time (during which they could also chat, read, and so forth). Each student then played the Gameshow against one other opponent. Afterwards, the class was given free time to prepare for and/or continue play in the Gameshow. On Day 3, all students completed a paper and pencil posttest on the mechanisms of fever.

Measures and coding. The study included three sources of data. One was the computer-generated logging data that indicated how students used their time with the software. A second source of data was the quality of the concept maps. At the end of each day after the students were gone, each map was evaluated using automated scoring as described in the Results section. The final data source was the posttest, which had three levels of questions: factual, integration,

and application (see Appendix). Factual questions asked about facts that were stated explicitly in the passage. Integration questions required integration of information across the passage. Application questions required applying the fever mechanisms to situations not discussed in the passage. Each question was scored on a 0 to 2 point scale for incorrect, partially correct, and fully correct answers. Two independent coders scored a minimum of 30% of the data for each question. Reliability ranged from 95% to 100% for all questions. A single coder then scored the remaining data.

Results

When students worked with the software, they could complete a number of different activities that ranged from chatting to reading to game playing. A fairly prototypical sequence of activities for the first day comes from John Doe in the TA condition. John spent the first eight minutes customizing his agent and chatting with other students on-line. He then read the science passage for three minutes. He spent the next nine minutes alternating between connecting the nodes in the agent's map and referring to the reading passage. After having made headway with his agent's map, John spent a minute formulating a question from the drop down menus, and then observing his agent's answer. He gave his agent one of the pre-made quizzes and edited the map based on the feedback for two more minutes. For the following nine minutes, he alternated between reading the passage, formulating and asking his agent questions, and editing the map based on the reading and the feedback. In the next four minutes he chatted on-line while looking for other students to play with in the gameshow. He then played the gameshow and chatted for the remaining time.

Other students followed similar patterns of moving between different activities. Some of the activities were directly relevant to learning such as reading the passage, creating the map,

formulating questions, and seeking feedback and revising. Other activities were less directly relevant to learning, for example, chatting, customizing the look of the character, and playing the game. The differences between the two conditions appeared in the relative distributions of activities that were directed towards learning and those that were not. The following sections describe the differences in activity distributions, and the evidence that students in the TA condition learned more.

[FIGURE 6 GOES HERE]

Effort Towards Learning. Students in the TA condition showed greater effort towards learning. Figure 6 shows how students spent their time in the software. The key difference is the greater time the TA condition spent on learning activities (working on the map or reading the passage). A repeated measures analysis crossed the factors of Day and Condition using proportion of time spent on learning activities as the dependent measure. There was a main effect for Day, with students making greater effort to learn on Day 1, $F(1, 59) = 431.7$, $MSE = .008$, $p < .001$. More importantly, there was a main effect of Condition, with TA students spending a greater proportion of their time learning, $F(1, 59) = 21.9$, $MSE = .015$, $p < .001$. There were no interactions. So, despite the attractions of chatting and playing, the TA students chose to learn for their TAs.

[TABLE 1 GOES HERE]

Table 1 shows the average number of times that students engaged in different learning activities (excluding reading, which is treated below). *Map Edits* refers to adding, deleting, or changing a link in the concept map. *Quizzes* refers to how many times students submitted their maps to get scored against a set of questions. *Asks* refers to how often students asked their maps to answer questions they posed. *Explains* refers to how often students asked their maps to trace

out the details of an answer in more detail. These variables were entered in a multivariate analysis with Condition as a between-subjects variable and Day as a within-subjects variable. Both Day, $F(4, 56) = 26.6$, $p < .0001$ and Condition, $F(4, 56) = 2.7$, $p < .05$ showed significant main effects. Looking at specific activities, the number of map edits and quizzes were significantly greater for the TA condition, p's $< .01$. Thus, students in the TA condition spent more time working on the concept maps and checking those maps with a quiz. It is also worth noting that students in the Avatar condition took advantage of the intelligence of the system by using the quiz, ask, and explain features. Though both conditions appreciated the same interactive affordances, the TA students used them more.

[FIGURE 7 GOES HERE]

The TA students' extra effort towards learning was not confined to working on the map, which might be expected on Day 2 because performance in the Gameshow was contingent on the map in the TA condition. The TA students also spent nearly twice as much time studying the fever passage. Figure 7 shows the time spent reading the passage. A repeated measures analysis used Day as a within-subjects factor and Condition as a between-subjects factor with Reading Time as the dependent measure. Students in the TA condition read longer, $F(1, 59) = 10.9$, $MSE=17.5$, $p < .005$. Students in both conditions read more on Day 1, $F(1, 59) = 213.1$, $MSE = 9.8$, $p < .001$. There was also an interaction, $F(1, 59) = 9.2$, $p < .005$, which indicates that the TA students showed the greatest reading difference on Day 1, even before they knew there was a performance venue for their TAs (i.e. the Gameshow). The mere belief of teaching a TA led to greater effort towards learning than did studying for oneself.

Effects on Learning. Given the extra effort towards learning, the next question is whether it led to better learning, as measured by the posttest. Based on prior research (Schwartz et al., in press), we did not expect differences on the basic fact questions. Rather, differences, if any, would show up on the harder integration and applications questions that required reasoning through causal chains. A second question was whether there would be a condition by prior achievement interaction. To get the most precise data possible, we removed five students who did not complete the full implementation. One student was not present on all three days of the study. Four students did not complete any questions on the posttest (fortuitously, they were distributed equally across condition and achievement level).

[FIGURE 8 GOES HERE]

A repeated measures analysis crossed Question Type with Condition, and used prior Achievement as a covariate crossed with the other two factors. There was a Condition by Question Type interaction with the largest TA advantage on the harder problems, $F(2, 102) = 3.8$, $MSE = 0.5$, $p < .05$. There was also an Achievement by Condition by Question Type interaction, $F(2, 102) = 4.2$, $p < .05$. Figure 8 shows the average scores on each of the Question Types by Condition. It indicates the effect of Achievement by breaking it into a high and low variable (using the median of all the students as the break point), instead of a continuous variable as used in the statistical analyses. One way to interpret the complex interaction is to compare the low-achieving TA students with the high-achieving Avatar students. As the questions become more complex, going left to right, the low-achieving TA students catch up with the high-achieving Avatar students.

[FIGURE 9 GOES HERE]

In-Game Correlates of Achievement Effects on Learning. Given the positive effects of the TA condition for the low-achieving students, we examined the log files to see if there was an identifiable activity that contributed to the effect. A multivariate analysis used Condition, Day, and Achievement (high-low on a median split) as crossed factors with the frequencies of the various learning activities as the dependent measures. The only variable to exhibit a significant Condition by Achievement interaction was the time spent editing the maps; $F(1, 56) = 5.3$, $MSE = 52.8$, $p < .05$. Figure 9 shows that the low-achieving TA students took advantage of the map editing feature much more than the low-achieving Avatar students. They were working harder to get their maps just right.

[FIGURE 10 GOES HERE]

One potential concern is that the low-achieving students in the TA condition may have just been rapidly adding and deleting links in a trial and error fashion rather than in a thoughtful way. An analysis of the students' concept maps indicates this was not the case. The maps were scored automatically against the expert map. Figure 2b, which shows the All-Possible-Questions (APQ) matrix, helps clarify how the scoring was completed. The APQ matrix indicates the agent's accuracy on all possible questions of the form, "If X increases what happens to Y?" where X and Y are nodes from the expert map. From this matrix, we derived an APQ index, which is the percentage of correct answers for those questions that relate two nodes with a traceable path in between. The APQ index naturally weights more central nodes in the concept map, because they are involved in more questions.

Compared to other measures of system use, the APQ index was the best correlate of posttest performance; APQ_{Day1} by posttest $r = .46$, and APQ_{Day2} by posttest $r = .37$, p's $< .01$. The most telling data compare the APQs for the low-achieving students from the two conditions,

as shown in right-hand panel of Figure 10. Compared to the maps of the low-achieving students in the Avatar condition, the maps of the low-achieving students in the TA condition were twice as good on Day 1 (M_{TA} = 18. 3, M_{Avatar} = 9.5), and three times as good on Day 2 (M_{TA} = 28.5, M_{Avatar} = 9.9). This indicates that the low-achieving students in the TA condition were not just changing their maps arbitrarily. Rather, they were putting in the effort to make their maps better, and they were succeeding.

Discussion

Study 1 was designed to determine whether students would make greater effort towards learning for their TAs than they would for themselves. On the first day, the TA students were told they were instructing their Teachable Agents, whereas the Avatar students were told they were making concept maps to help themselves learn. They used identical software, and the only difference was their belief state. The differences in the effort towards learning on the first day testify to the power of protégés to influence learning behaviors. Students had attractive alternatives to reading and map editing, namely, the opportunity to chat online and play a game with other students. Furthermore, on Day 1, performance in the Gameshow was not contingent on the maps for either condition. Nevertheless, students in the TA condition spent more time editing their maps and quizzing them, and they spent nearly twice as long reading the fever passage as students in the Avatar condition. Instead, Avatar students spent proportionately more time using the chat feature and playing the Gameshow.

On the second day, students in the TA condition saw their TAs play in the Gameshow, whereas students in the Avatar condition played the game themselves. Again, the TA students spent more time working on their maps, as would be expected, because their TAs had to have accurate maps to do well in the Gameshow. Interestingly, this was especially true for the low-

achieving students in the TA condition, who spent much more time improving their maps than the low-achieving students in the Avatar condition. These differences led to relative gains in learning as measured by the posttest. Students in the TA condition did better on the harder questions, and this was especially noticeable for the low-achieving students. On the hard application questions, they performed as well as the high-achieving Avatar students.

It is useful to note that the motivational differences between the conditions should not be attributed to students having "more fun." Students in both conditions enjoyed chatting and playing the Gameshow, and it is hard to imagine that reading would be more "fun" in this context. On a set of moment-to-moment measures of engagement, not reported here for the sake of simplicity, there were few reliable differences between the conditions. Rather, students in the TA condition were motivated to put greater effort towards learning. This seems like a useful motivational target for designers of educational games, where students often just want to play.

An important question is how to sustain these motivational benefits for months and not just days. One can imagine that the fiction of teaching an agent might lose its luster, and students could stop working so hard to learn on its behalf. One way to address this question is to imagine what would happen if protégés were put into games that included several motivational elements such as rich narratives, clear goals, and incremental challenges. We hypothesize that these motivators would spill over to help sustain the teaching metaphor. For example, students would be energized to learn so they could help their protégés advance to the next level in the game, perhaps even more so than if they were playing only for themselves.

Study 2: Psychological Concomitants of the Protégé Effect

Study 1 demonstrated the protégé effect: students put forth greater effort to learn for their TAs than for themselves. However, the behavioral and learning data collected in Study 1 do not

shed light on the underlying mechanisms of this effect. To uncover possible causes, participants in Study 2 were asked to think aloud, externalizing their thoughts and emotions, while they worked with either a TA or an avatar. These data begin to uncover the psychological machinery behind the TA students' increased motivation to learn.

Study 2 had a similar design as the first; half of the students were in a TA condition and half were in an Avatar condition. 5th-grade children received the same fever passage and an identical set of nodes to connect within the concept map. Students were videotaped as they worked for approximately one hour. The children were encouraged to think aloud, and their protocols were transcribed and coded. Analysis of the data focused on three primary questions. The first question was whether there would be a replication of the protégé effect, where students make greater effort to learn for their TAs than for themselves. The other two questions focused on the psychological mechanisms behind the protégé effect.

The first psychological question was whether the students would treat their TAs as independent, sentient beings. For example, would they talk about their TAs' thoughts? Would they distribute responsibility for performance in the Gameshow across themselves and their TAs? If so, this would indicate that students treated the TA as a protégé, because its behavior was partially due to themselves but partially independent. This could create a sense of responsibility that would lead students to try harder for their TAs than for themselves. For example, previous research with Betty's Brain documented anecdotal evidence of students developing a feeling of responsibility towards their TAs (Biswas, Leelawong, Schwartz, Vye, & TAG-V, 2005).

The second psychological question was how students would respond to failure with the TA as a mediator. This was especially relevant to the positive effects found for the low-achieving

students in the preceding study. In a performance situation, students with self-perceived low ability often avoid difficult learning tasks or give up quickly, because they are afraid of failure (Elliot & Dweck, 1988). More generally, sustained experiences of personal failure may lead students to opt out, losing interest altogether for certain learning activities. The TA, however, creates a situation in which responsibility for failure is distributed across teacher and pupil. Instead of blaming their own knowledge and abilities, students may fault their TA or their poor teaching. This may allow them to both acknowledge failures and address them by working harder to learn.

Methods

Participants. Twenty-four 5th-grade students from a high-performing private school participated (10 male and 14 female). Students were predominantly Caucasian and Asian American and came from a common high-achievement profile, as determined by the school. Because this population of students is younger yet higher achieving than in the prior study, attempts to generalize findings across studies must be done with caution.

Design & Procedure. The TA and Avatar conditions were similar to those of Day 2 in Study 1. Students in the TA group taught their TAs and watched them answer Gameshow questions. Students in the Avatar condition learned on their own and answered Gameshow questions themselves. Unlike in the prior study, Betty's reasoning was turned off for the Avatar students. They were simply using graphical tools to make concept maps, while TA students were able to ask the TAs questions and view their reasoning. Also unlike the prior study, the children in both conditions played the Gameshow alone; other children were not logged on at the same time. Dependent measures included verbalizations made during the Gameshow, time spent on learning-relevant behaviors, and scores on an oral posttest of fever mechanisms.

Before beginning the protocol study, students received software training in a 45-minute classroom session. During the session, students personalized characters that would represent themselves or their TAs, depending on condition. Students then watched the experimenter give a demonstration of the software. The experimenter interacted with the whole class to build a practice map projected at the front of the room.

One to three weeks later, each participant completed an individual, 60-minute session. Three researchers, each trained in the research protocol, ran the sessions, randomly switching between conditions. All sessions were videotaped for later analysis. Each session had four phases: Prepare, Play, Revise, and Posttest.

In the Prepare phase, students first read the fever passage aloud. Each student then used the software to construct a concept map of fever mechanisms. The TA condition was told "Teach your agent the best you can by making this concept map," while the Avatar condition was told "Learn the best you can by making this concept map." Participants could spend as much time as they wanted building their concept maps or looking back at the passage, and this time was recorded.

During the Play phase, students first practiced doing a think-aloud while playing Sudoku. Students then played the Gameshow while thinking aloud. There were a total of six Gameshow questions, which varied in difficulty. Every participant saw the same six questions, and the system provided feedback on answer accuracy. If students were silent for ten seconds, they were prompted with "what are you thinking now?" If students were not verbalizing at all, they were prompted with the following questions: (1) What is the answer and why? (2) Why wager that amount? (3) Will the answer be right or wrong? (4) Why is the answer wrong?

In the Revise phase, students were told they would soon play a more difficult round, and if they chose, they could prepare by reviewing the feedback from the Gameshow, re-reading the passage, and/or working on the concept map. Students received as much time as they wanted to prepare for the second round, except for one student, who spent so much time in the Prepare phase that there was no time for revision (although she wanted to revise).

In the Posttest phase, students were told, "We have run out of time to play Round Two. I'd like to ask you a few questions before sending you back to class." Students answered nine questions orally (see Appendix). Similar to Study 1, posttest questions were scored on a scale of 0 to 2.

Results

The students in the TA condition replicated the findings of Study 1 in that they put forth more effort towards learning. The new findings come from the protocol analyses. Students in the TA condition treated their agents as sentient and partly responsible for getting an answer right or wrong. The TA students were also much more likely to acknowledge when an answer was wrong by exhibiting negative affect and making attributions. The following analyses, which also include samples of student dialog, detail these findings and suggest several ways that teachable agents lead students to put greater effort towards learning.

Efforts Towards Learning. The TA students demonstrated greater effort towards learning as measured by their combined reading and map editing times. A repeated measures analysis crossed the factors of Occasion (Prepare or Revise) by Condition with combined reading and map editing times as the dependent measure. There was a main effect for Occasion, with students spending more time in preparation than revision, $F(1, 21) = 17.2$, $p < .001$. There was also a main effect for Condition, with TA students spending more time overall, $F(1, 21) = 25.1$, $p <$

.001. The interaction of Occasion by Condition was not significant, but descriptively, Figure 11 shows the advantage for the TA group was greatest during the Revise period. Only 64% of Avatar students chose to revise at all, compared with 100% of TA students. Even if the analysis only includes the Avatar students who did choose to revise, the TA students persisted three times longer during the Revise phase, $t(16) = 4.88$, $p < .001$, ($M_{TA} = 8.6$ min, SD = 3.2, $M_{Avatar} = 2.5$ min, SD = 1.8). As in Study 1, students in the TA condition were more likely to choose to refine their understanding, and they spent more time doing so.

[FIGURE 11 GOES HERE]

These differences in learning behaviors, however, did not translate into differences in learning outcomes. The posttest scores did not significantly differ by condition (per question on a 0-2 scale, $M_{Avatar} = 0.95$, SD = 0.39; $M_{TA} = 0.85$, SD = 0.39). Given the short duration of the treatment and the relative complexity of the materials (which had been designed for 8th graders), this finding was not surprising.

Coding of Protocol Data. The verbal record provides some insight into the protégé effect and why TA students were motivated to make greater effort towards learning. Verbal protocols taken during Gameshow play were transcribed and coded at the statement level. A statement was defined as any phrase or series of phrases that expressed a single sentiment or thought. Statements were first classified into three major categories: mental attributions, responsibility attributions, and affective statements.

Mental attributions were defined as statements that assigned credit for thoughts or mental actions to an entity. These statements were further coded as attributing credit to the self ("I don't

understand"), the TA ("He knows it!"), or some combination of both ("I know she knows this one").

Responsibility attributions assigned credit for successes or failures on questions in the Gameshow (i.e. getting a question right or wrong). Like mental attributions, responsibility attributions were classified as crediting the self ("Yeah! I did it!"), the TA ("Thanks a lot, Queenworld."), or both ("That's one of the things I didn't teach her"). They were further subdivided into whether the attribution assigned credit for a failure ("I didn't know that one") or a success ("We did it!").

Affective statements were expressions of students' emotions. They were coded as positive or negative. Positive statements expressed enjoyment, excitement, hope, or relief (e.g. "Cool!", "This is fun," or "Now I'm kind of relieved"). Negative statements expressed anger, annoyance, pity, or sadness (e.g. "Poor Diokiki," "I'm not a good teacher" or "Oh shoot!"). Affective statements were also categorized by whether they occurred in response to success or failure in the Gameshow.

Using a subset of the transcripts (30%), two researchers applied the codes (one blind to the hypotheses). Inter-rater reliability ranged from 77% to 100%, with an average agreement rating of 90% across coding categories. A primary researcher coded the remaining transcripts. The results were tallied into three scores so that each student had a mean number of mental attributions, responsibility attributions, and affective statements per Gameshow question.

Attributions Towards the TAs. These data demonstrate that students saw the TA's performance as a reflection of their own knowledge but also viewed the TA as a separate entity that had thoughts of its own. One-fifth of the TA students' mental attributions were made exclusively

towards the TA, suggesting that they gave the TA credit for having its own knowledge ("He totally knows this one") and reasoning skills ("He could probably figure it out"). One-fourth of the TA students' mental attributions were made towards a combination of student and TA (e.g. "I, err… he didn't know it"), as if students were confused about who was doing the thinking – themselves or their digital pupils. Finally, 55% of TA students' mental attributions were self-directed compared with 100% in the Avatar group (see the left side of Table 2). Students in the Avatar condition did not perceive their Avatars as sentient, and therefore made all attributions to themselves.

[TABLE 2 GOES HERE]

In addition to mental attributions, students also attributed responsibility to the TA for Gameshow outcomes (both successes and failures). The right side of Table 2 shows that the TA students apportioned responsibility equally across themselves ("I got it right"), the TA ("He got it wrong"), and some combination of both ("We did it!"). To some extent, students treated the TA as a separate entity with social status, while the combined attributions of self and TA indicate they considered the TA a protégé (part self, part other). Again, the Avatar students made only self-attributions.

Response to Failure. Students from the two conditions demonstrated strikingly different affective and attributional profiles in response to an incorrect answer in the Gameshow. Table 3 shows that on average, TA students displayed more negative emotion in response to failure. Sixty-percent of the TA students made at least one statement of negative affect after failure compared to only 7% of Avatar students. Table 3 also shows that TA students were not simply

more emotive or less positive. What differentiated these two groups' affective profiles was their negative emotional response to failure.

[TABLE 3 GOES HERE]

In addition to the difference in emotions expressed after failure, students in the TA condition were more likely to assign responsibility for a failure. Table 4 shows that students in the TA group made far more responsibility attributions per failed question than Avatar students. In addition, every TA student made at least one attributional statement in response to failure, compared with only 64% of Avatar students. TA students tended to distribute the blame for failure evenly amongst themselves ("I didn't know that one"), their TAs ("He got it wrong"), or both ("We're gonna lose this one" or "I guess I didn't teach him that"). Avatar students, on the other hand, had no one to blame but themselves. In comparison to TA students, Avatar students made hardly any attributions after failure. However, in response to success, Avatar and TA students made similar numbers of attributions.

[TABLE 4 GOES HERE]

Discussion

As in the first study with 8th-grade students, this study found that 5th-grade students who worked with TAs spent more time on learning activities. During the initial preparation phase, they spent more time reading and constructing their maps. After playing the Gameshow, more TA students chose to revise, and they spent more time revising. This was expected, since the only way for a TA student to improve in the Gameshow was to edit the map. Study 1 provided evidence that simply believing one was teaching a TA led to greater effort towards learning, even without the incentive of the Gameshow. The main purpose of Study 2 was to gather students' thoughts to examine possible mechanisms behind the increased learning effort.

Verbal protocols revealed that students acted as though the TA were a sentient, semi-independent being who engaged in mental activity and deserved partial credit for outcomes in the Gameshow. TA students indicated this by distributing and co-mingling mental and responsibility attributions between themselves and their TAs. One student even named his TA "Echo," illustrating the symbolic role of the TA as protégé. Students viewed the TA as a social being that was partly them and partly another.

TA students acknowledged failure more often than the Avatar students by making more attributions for failure and expressing more negative affect. While TA students sometimes articulated frustration with their TAs ("Ughhh! Why does he keep saying large increase!?!"), most expressed sympathy ("Poor Diokiki… I'm sorry, Diokiki"). Often, these sympathetic statements were followed by statements of intention to take action to help their TAs perform better, as in the case of one student who said, "Whoa, I really need to teach him more." From this dialog, one gets the sense that students felt responsible for their TAs' performance in the Gameshow, because the TAs were enacting their teachings. At the same time, the students did not have to accept all the blame. TA students apportioned responsibility for failure across themselves, their TAs, and some combination of both (often in reference to poor teaching). The General Discussion considers how these factors may contribute to the increased effort towards learning.

General Discussion

Two studies demonstrated the existence of a protégé effect: students are more willing to make the effort towards learning on behalf of a computerized protégé than for themselves. The first study, which used a classroom-level intervention, revealed that students who taught TAs spent more time on learning behaviors and ultimately learned more than students who learned for

themselves. The protégé effect was particularly beneficial for low-achieving students who, through increased effort, developed an understanding of the complex biology content that was on par with the high-achieving students who did not use TAs.

The second study, which gathered individual verbal protocols, also found that students spent more time engaging in learning activities for their TAs than for themselves. The verbal data provided possible reasons for the students' greater effort towards learning. For these students, the TA existed in a middle ground between avatar and agent. Like an agent, the TA was treated as an independent, social being that was attributed with cognitive states and responsibility for the quality of its answers. And like an avatar, students viewed the TA as a reflection of themselves. The students did not simply treat the TA as computer software they had programmed. In fact, TA students were particularly attentive and emotionally responsive when their protégés failed, and they often expressed regret that they had not taught their TAs well enough.

By occupying the unique social position of part self, part other, the TA incited motivation to work harder to learn. This type of motivation is unusual in computer environments, because it removes students from the very thing that is motivating them; students leave their TAs to read. The protégé effect can be contrasted with common motivational features added to computer environments, like gaining points towards some quantitative goal and engaging in fantasy contexts, which keep the student at the computer terminal longer. In these latter cases, learning is a side effect of sustained engagement. With the TA, the students were motivated to learn *per se* – so much so, that they chose learning activities over attractive and novel alternatives like chatting and playing games.

Three factors may contribute to the protégé effect: an ego-protective buffer, the adoption of an incrementalist theory of TA intelligence, and a sense of responsibility. In broad strokes, the students' egos are spared enough that they can acknowledge failure; they know there is a clear way to ameliorate the failure by teaching better; and, they are inspired to do so because they feel they owe it to their TAs.

A protégé offers an *ego-protective buffer* (EPB). The EPB shields students from forming negative beliefs about themselves, because the blame for failure can reside elsewhere. For instance, when a TA is failing, it can absorb part of the blame. Moreover, the TA's failure can be attributed to poor teaching, which also deflects the blame away from an "internal" property of the student. Failure attributions that identify poor teaching as the source of the error also occur in human-human teaching. Ross, Bierbrauer, and Polly (1974) and Ames (1975) found that professional and non-professional teachers instructing human students attributed failures to their own teaching. Without the EPB provided by the TA, learners have only themselves to blame and may be more likely to fault their own intellects.

The EPB helps students acknowledge the need for revision, but to take action, students must also believe that revision will be fruitful. Dweck's theory of incremental versus entity beliefs about intelligence is relevant here (Dweck, 2000). Individuals who have an entity theory believe their intellectual ability is fixed and unchangeable. Incrementalists, on the other hand, believe that intelligence is malleable and fluid. To them intelligence is more like knowledge than an innate ability. According to Dweck, students with an incremental theory put greater effort towards learning because they believe their efforts can change their intellectual abilities.

Through the protégé effect, children appear to become incremental theorists about their TAs' abilities. With the TA, it is obvious how to make incremental progress – teach better by

getting the links and nodes right. TA students are more willing to put in effort because they believe it can improve their TAs. For students who learn for themselves, there is no transparent mechanism that links a specific learning behavior to improved performance (especially for 5th graders, who may not have the metacognitive wherewithal to strategically improve their understanding). In other words, TA students know how to enhance their TAs' knowledge while Avatar students may not believe it is possible to change their own intelligence (or may not know how to). This difference may have been especially significant for the low-achieving students in the first study. On Day 2, the low-achieving TA students made many edits to the concept maps, whereas the low-achieving Avatar students made almost none. Similarly, Dweck has found that both low and high-achieving incrementalists persist through challenging tasks by adopting high-quality learning behaviors, while low-achieving entity theorists tend to adopt self-sabotaging characteristics that signify a state of learned helplessness (Dweck, 2000).

The third factor in the protégé effect is a sense of responsibility, which can help explain why the TA students spent more time on learning activities before they received any success or failure feedback. The verbal data in Study 2 suggests that students felt responsible for their TAs' learning. Just as parents nurture and care for their children and coaches spend time with their players, students do the same for their TAs. Recall the students who said, "Whoa, I really need to teach him more," and "Poor Diokiki… I'm sorry, Diokiki." This sense of responsibility may have propelled students to persist and revise, which could explain the TA students' greater reading and revision times.

These three factors comprise a distinctly social story, even though the children were interacting with a computer program. Social motivations provoked by the TA were strong enough that students wanted to learn, even more than they wanted to chat with other students

online. This demonstrates the potential power of sociable technologies for learning. The EPB, incremental theorist, and responsibility explanations require further research to establish their validity, but a key aspect of these accounts is that students treat their TA as a protégé – a separate but dependent "other" with social and sentient attributes.

To further isolate the significance of the social, one possible study design could replace the Avatar condition with a condition where students are told to write a computer program. This would help distinguish the role of general production (programming) versus social production (teaching). Given our hypothesis that the protégé effect is due to social motivations, we would expect students in the programming condition to be less inclined to acknowledge errors, more inclined to think the errors reflect their intelligence, and less inclined to feel responsible to their computer programs. Ultimately, these students would make less effort to learn.

Conclusion

Over the next few years, we anticipate that avatars and intelligent agents will be increasingly blended. In a virtual environment, for example, a player's character may provide feedback by disobeying when the player makes too many bad decisions (Arena, Schwartz, & Bailenson, 2009). Or in a simulation of classroom interactions, a user may create students with various traits and observe how they would behave as a group. TAs and other hybrid technologies such as these present innovative educational opportunities while raising new questions about learning. For instance, what kinds of social relationships besides tutor-tutee might be beneficial for learning? Just how "social" must the interaction between human and computer be to motivate learning? What are the boundaries of the term "social"? If future research addresses these questions it may uncover new psychological phenomena that occur in

the social interactions between human and computer. In turn, this research can help create a new generation of effective educational technologies filled with social intelligence.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grants EHR-0634044, SLC-0354453, and by the Department of Education under grant IES R305H060089. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agencies.

REFERENCES

Ames, R. (1975). Teacher's attributions of responsibility: Some unexpected non-defensive effects. Journal of Educational Psychology, 67(5), 668-676.

Arena, D.A., Schwartz, D.L., & Bailenson, J.N. (2009, June). Effects of social belief on spatial learning in virtual reality. Paper presented to the Games, Learning and Society Conference 5.0, Madison, WI.

Bailenson, J. N. & Blascovich, J. (2004). Avatars. In W.S. Bainbridge (Ed.), Berkshire encyclopedia of human-computer interaction (Vol 1, pp. 64-68). Great Barrington, MA: Berkshire Publishing Group.

Baylor, A. L. (2007) Pedagogical agents as a social interface. Educational Technology, 47(1), 11-14.

Baylor, A. L. & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. International Journal of Artificial Intelligence in Education, 15(2), 95-115.

Biswas, G., Leelawong, K., Schwartz, D. L., Vye, N., & TAG-V (2005). Learning by teaching: A new agent paradigm for educational software. Applied Artificial Intelligence, 19, 363-392.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J.N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. Psychological Inquiry, 13, 103-124.

Chen, J., Shohamy, D., Ross, V., Reeves, B., & Wagner, A. D. (2009). The impact of social belief on the neurophysiology of learning and memory. Abstracts from the Annual Meeting of the Society for Neuroscience. San Francisco, CA.

Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. Cognitive Science, 32, 301-348.

Clarebout, G., Elen, J., Johnson, W.L., & Shaw, E. (2002). Animated pedagogical agents: An opportunity to be grasped? Journal of Educational Multimedia and Hypermedia 11(3), 267-286.

Davachi L., Mitchell J.P., & Wagner A.D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. Proceedings of the National Academy of Science, 100(4), 2157-62.

Dweck, C. S. (2000). Self-Theories: Their role in motivation, personality, and development. Philadelphia: Taylor & Francis Group.

Elliot, E., & Dweck, C. (1988). Goals: An approach to motivation and achievement. Journal of Personality and Social Psychology, 54(1), 5-12.

Gerhard, M., Moore, D., & Hobbs, D. (2004). Embodiment and copresence in collaborative interfaces. International Journal of Human-Computer Studies, 61(4), 453-480.

Imbert, R., & de Antonio, A. (2000). The Bunny Dilemma: Stepping Between Agents and Avatars. Proceedings of the 17th Twente Workshop on Language and Technology.

Kuhl, P., K., Tsao, F-M., & Liu, H-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. Proceedings of the National Academy of Sciences, 100, 9096-9101.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 359-366.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? Cognition and Instruction, 19, 177-213.

Okita, S.Y., Bailenson, J., Schwartz, D. L. (2007). The mere belief of social interaction improves learning. In D. S. McNamara & J. G. Trafton (Eds.), The Proceedings of the 29th Meeting of the Cognitive Science Society (pp. 1355-1360). August, Nashville, USA.

Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S, Wieman, C., & LeMaster, R. (2006, January). PhET: Interactive simulations for teaching and learning physics. The Physics Teacher, 44 (1), 18-23.

Pesce, M. (2000). The playful world: Interactive toys and the future of imagination. New York: Ballantine Books.

Reeves, B., & Nass, C. (1998). The media equation: How people treat computers, television, and new media like real people and places. New York: Cambridge University Press.

Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. Learning and Instruction, 5, 21-36.

Ross, L., Bierbrauer, G., & Polly, S. (1974). Attribution of educational outcomes by professional and nonprofessional instructors. Journal of Personality and Social Psychology, 29(5), 609-618.

Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. Journal of Computer Assisted Learning, 19, 195-208.

Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (2007). Animations of thought: Interactivity in the teachable agents paradigm. In R. Lowe & W. Schnotz (Eds).

Learning with Animation: Research and Implications for Design (pp. 114-40). UK: Cambridge University Press.

Schwartz, D.L., Chase, C., Chin, C., Oppezzo, M., Kwong, H., Okita, S., Biswas, G., Roscoe, R.D., Jeong, H., & Wagster, J.D., (in press). Interactive Metacognition: Monitoring and Regulating a Teachable Agent. To appear in xxxx (Eds.), Handbook of Metacognition in Education, xxx: xx.

Shimoda, T. A., White, B. Y, & Frederiksen, J. R. (2002). Student goal orientation in learning inquiry skills with modifiable software advisors. Science Education, 86, 244-63.

Turing, A.M. (1950). Computing Machinery and Intelligence. Mind, 59(236), 433-460.

Turkle, S. (1995). Life on the screen: Identity in the age of the internet. New York: Simon & Schuster.

Wagster, J., Tan, J., Wu, Y., Biswas, G., & Schwartz, D. L. (2007). Do learning by teaching environments with metacognitive support help students develop better learning behaviors? The Proceedings of the 29th Meeting of the Cognitive Science Society (pp. 695-700). August, Nashville, USA.

Weizenbaum, J. (1976). Computer power and human reason. San Francisco, CA: W. H. Freeman.

Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior. Human Communication Research 33, 271-290.

TABLES

Table 1. Frequency of Different Learning Activities (and Standard Errors of Means)

	Map Edits	Quizzes	Asks	Explains
Day 1				
TA	16.7 (1.4)	2.9 (0.5)	1.6 (0.4)	0.5 (0.2)
Avatar	13.8 (1.4)	2.1 (0.5)	1.0 (0.4)	0.2 (0.2)
Day 2				
TA	8.6 (1.2)	4.6 (0.8)	0.6 (0.3)	0.0 (0.1)
Avatar	2.4 (1.2)	1.3 (0.8)	0.6 (0.3)	0.2 (0.1)
Day Average				
TA	12.7 (1.0)**	3.8 (0.5)**	1.1 (0.3)	0.3 (0.1)
Avatar	8.1 (1.0)	1.7 (0.6)	0.8 (0.3)	0.2 (0.1)

** $p < .01$

Table 2. Mean Number of Attributional Statements per Question (with Standard Errors of Means)

	Mental Attributions			Responsibility Attributions		
	Self	TA	Both	Self	TA	Both
TA	1.15 (.19)	0.43 (.17)	0.52 (.19)	0.43 (.11)	0.48 (.19)	0.45 (.13)
Avatar	1.92 (.30)	n/a	n/a	0.56 (.10)	n/a	n/a

Table 3. Mean Number of Affective Statements per Success or Failure (with SE)

	Positive After Success	Total Positive	Negative After Failure	Total Negative
TA	0.64 (.30)	0.67 (.32)	0.62 (.20)**	0.60 (.20)
Avatar	0.48 (.12)	0.58 (.14)	0.02 (.02)	0.51 (.22)

** Z = 2.9, $p < .01$, Mann-Whitney

Table 4. Mean Number of Responsibility Attributions per Success or Failure (with SE)

	Attributions to Success				Attributions to Failure			
	Self	TA	Both	Total	Self	TA	Both	Total
TA	.17(.12)	.27(.12)	0.0 (.0)	.44 (.16)	.54(.13)	.47(.21)	.66(.19)	1.67(.28)**
Avatar	.53(.10)	n/a	n/a	.53(.10)	.65(.22)	n/a	n/a	.65 (.22)

** Z = 2.7, $p < .01$, Mann-Whitney

FIGURE CAPTIONS

Figure 1. The Teachable Agent Betty's Brain. Using the Betty software, each student teaches her own TA (in this case, named "Dee") by constructing a concept map as its "brain." Through basic artificial intelligence techniques, the TA can answer questions based on the relationships depicted in its map. Students can query the TA using a pull down menu. The highlighted links and nodes in the figure show how the TA answers the question, "If 'blood flow to skin' increases, what happens to 'body temperature'?"

Figure 2. Software Options for Various Types of Feedback. Panel A shows a front-of-the-class (FOC) display, where teachers project and query multiple "brains" (maps) simultaneously. The highlights around each concept map indicate correct and incorrect answers. Panel B shows the All-Possible-Questions (APQ) matrix. The matrix indicates a TA's accuracy when asked the complete population of possible questions in a hidden expert map. All concepts are displayed on both axes. Each cell displays feedback to the question, "If Y increases, what happens to X?" For both applications, green indicates a correct answer, red indicates incorrect, and yellow indicates correct but by the wrong causal path. A version of the Betty's Brain environment and teacher tools can be found at <aaalab.stanford.edu>.

Figure 3. Triple-A-Challenge Gameshow. (A) Students log on from home or school. (B) They customize the look of their individual TAs and give them names. (C) They teach their TAs. (D) Students can chat, see their progress, and find other students who want to play a game. (E) Students can play in a game show, where a host asks questions, and they wager on whether their TAs will answer correctly.

Figure 4. Expert Map of the Fever Passage. Students received the same nodes as in the expert map, but the links were removed and the nodes were not neatly organized. The expert map was used to check the accuracy of answers and to generate questions for the quizzes and Gameshow.

Figure 5. Overview of Study 1. The underlined elements indicate experimental differences between treatments.

Figure 6. How Students Used Their Time When Logged On.

Figure 7. Reading Times by Condition and Day.

Figure 8. Posttest Scores Separated by Question Type, Condition, and Achievement Level.

Figure 9. Average Number of Map Edits Separated by Achievement Level, Day, and Condition.

Figure 10. APQ Index Scores Separated by Achievement Level, Day, and Condition.

Figure 11. Time Spent on Learning Activities During Prepare and Revise Phases.

FIGURES

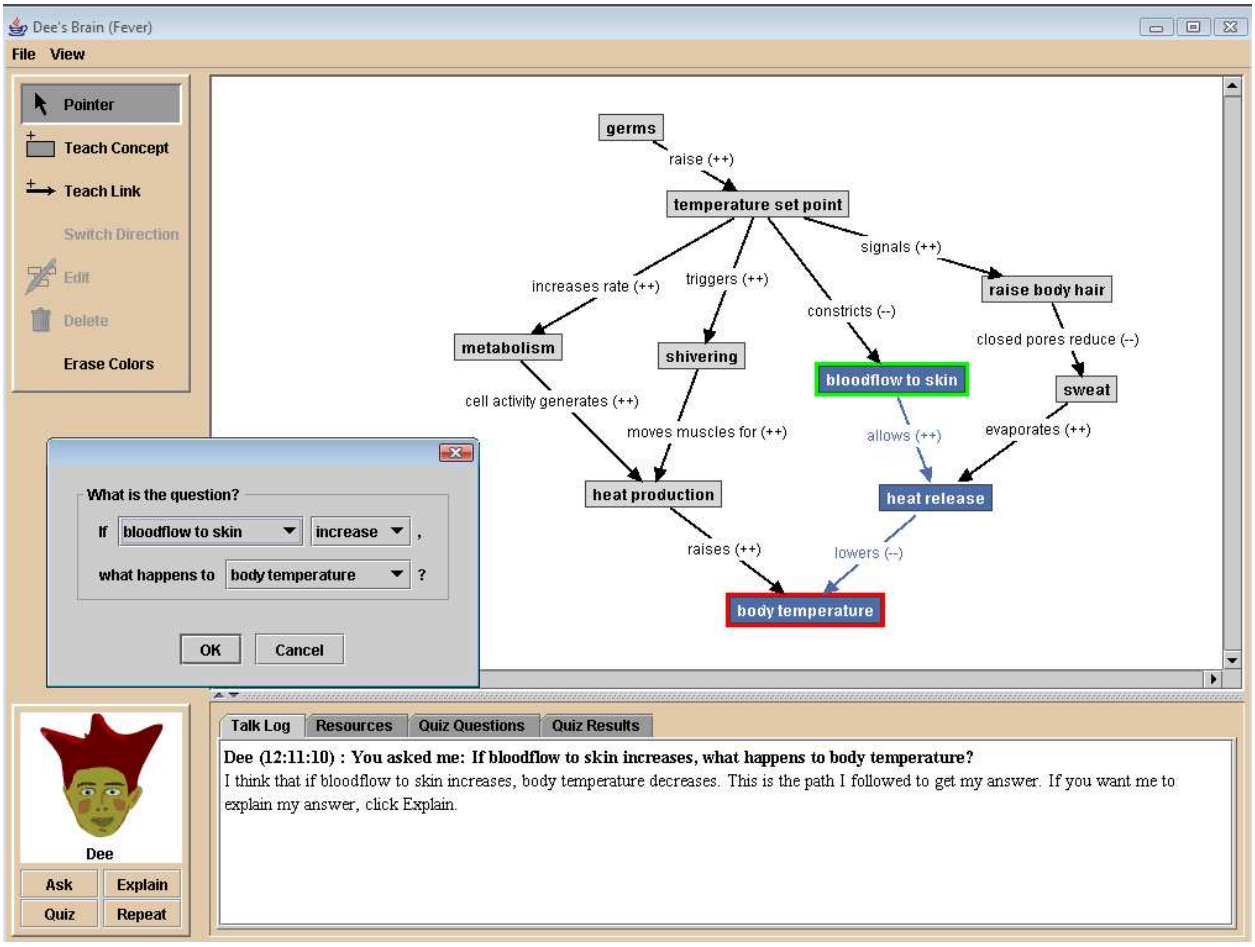


Figure 1.

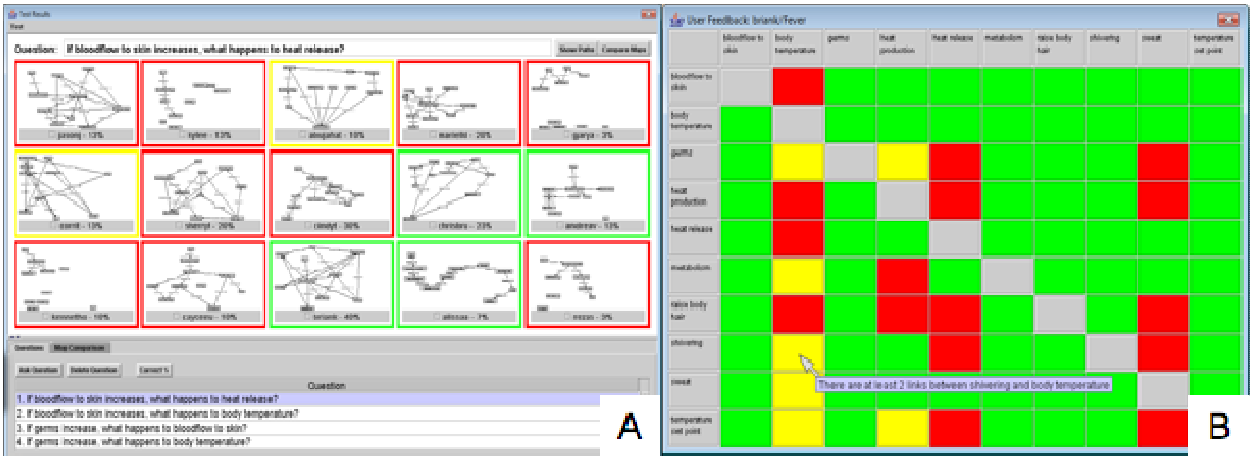
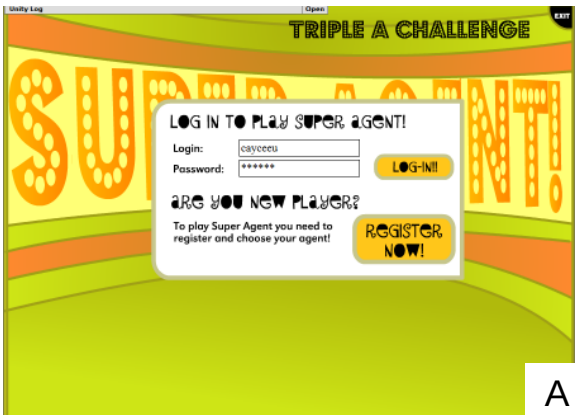
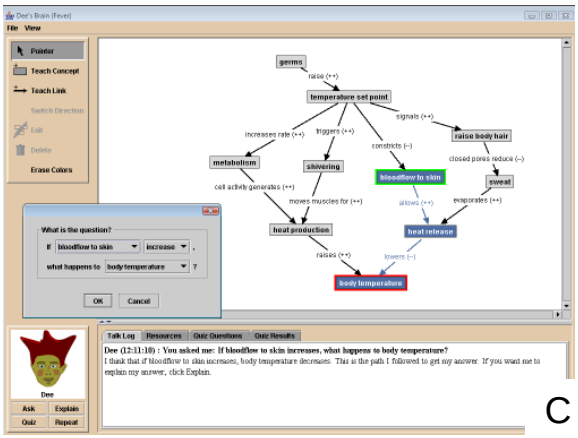


Figure 2.

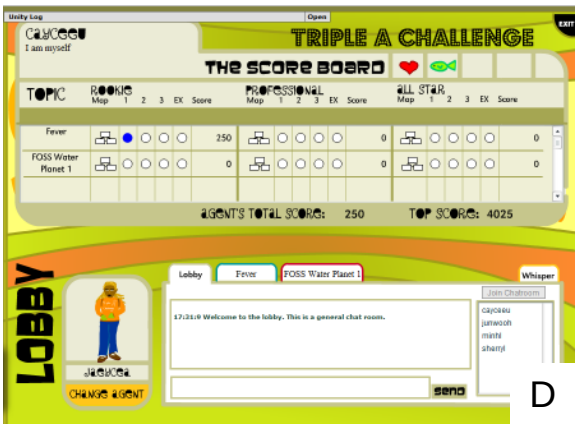


A

B



C



D

E

Triple-A-Challenge Gameshow

A. Login Page
B. Agent Customization Room
C. Betty's Brain Operation or Mapping Window
D. Main Lobby Room
E. Game Room

Figure 3.

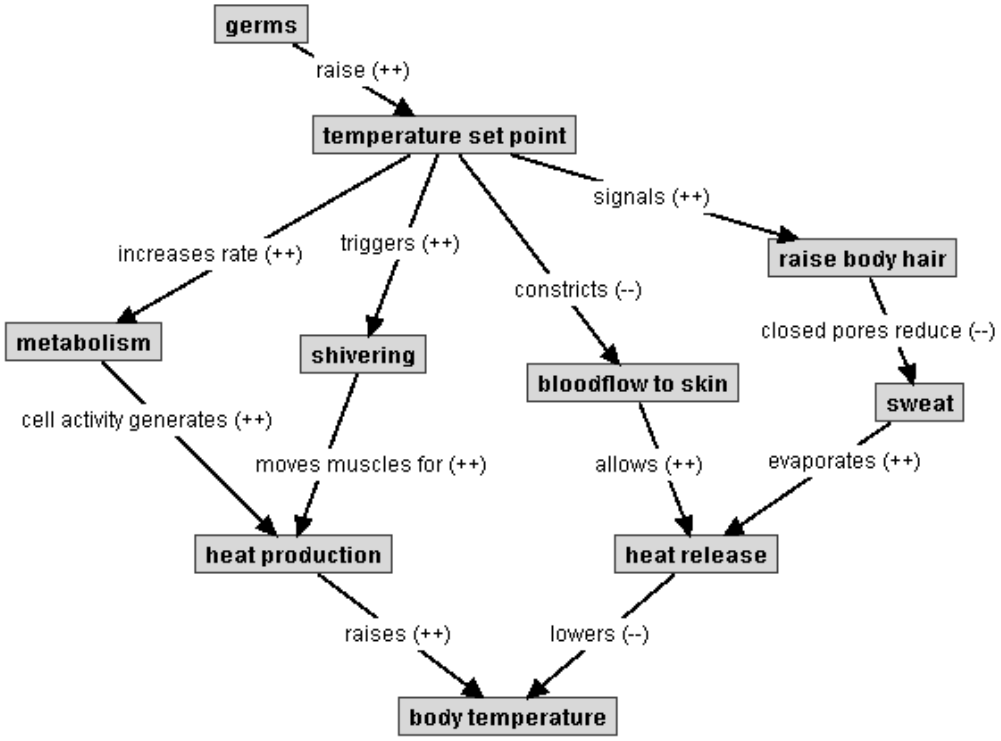


Figure 4.

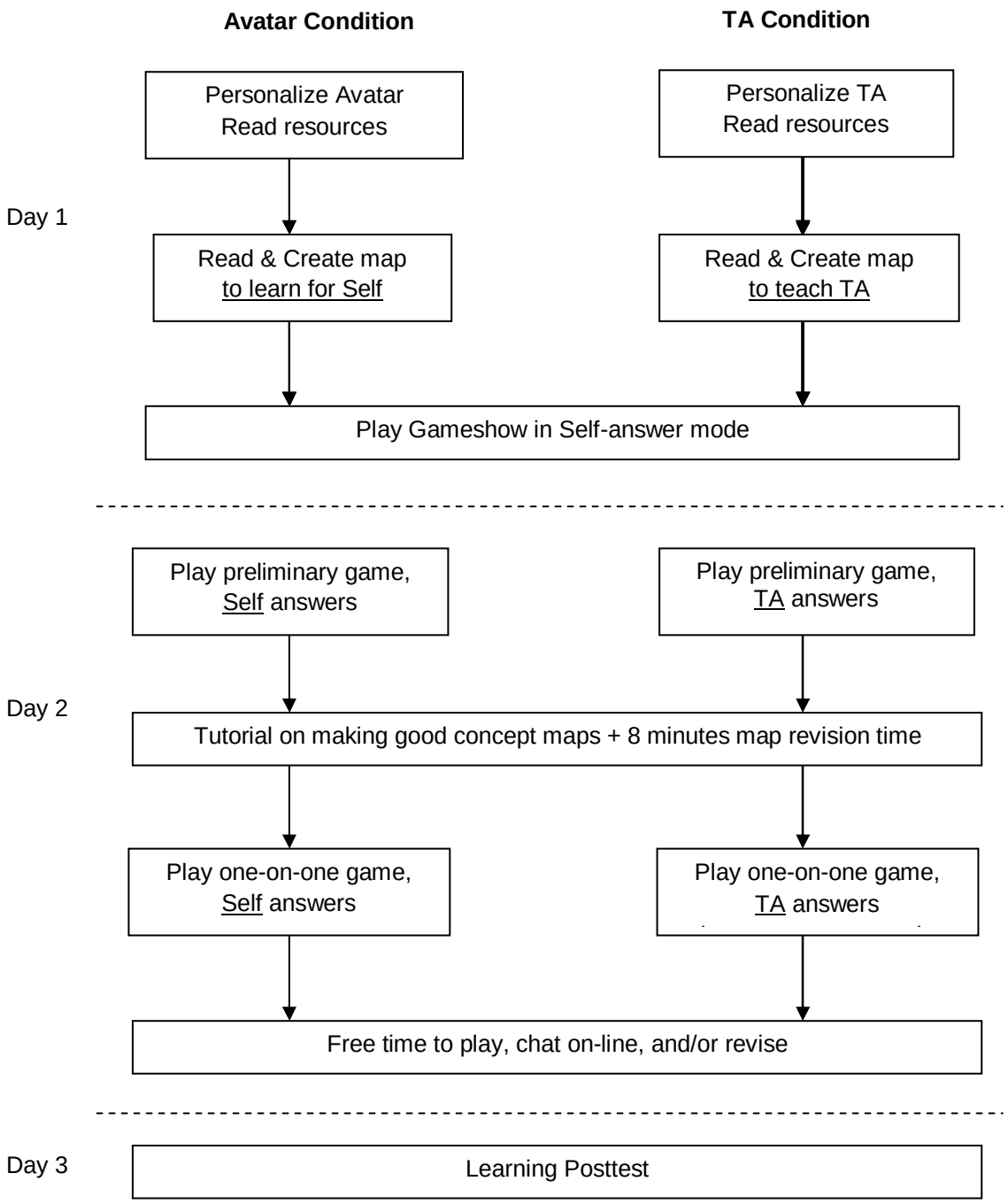


Figure 5.

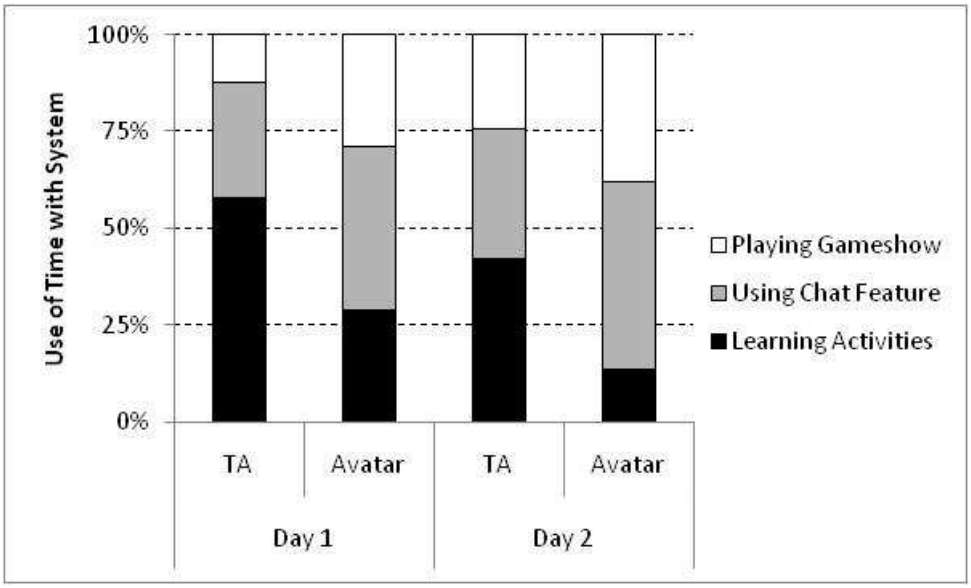


Figure 6.

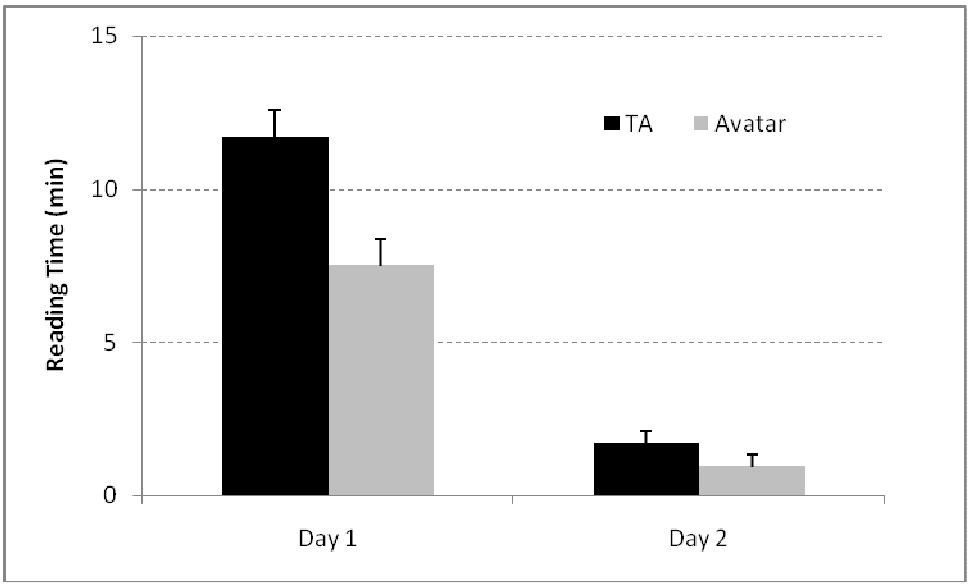


Figure 7.

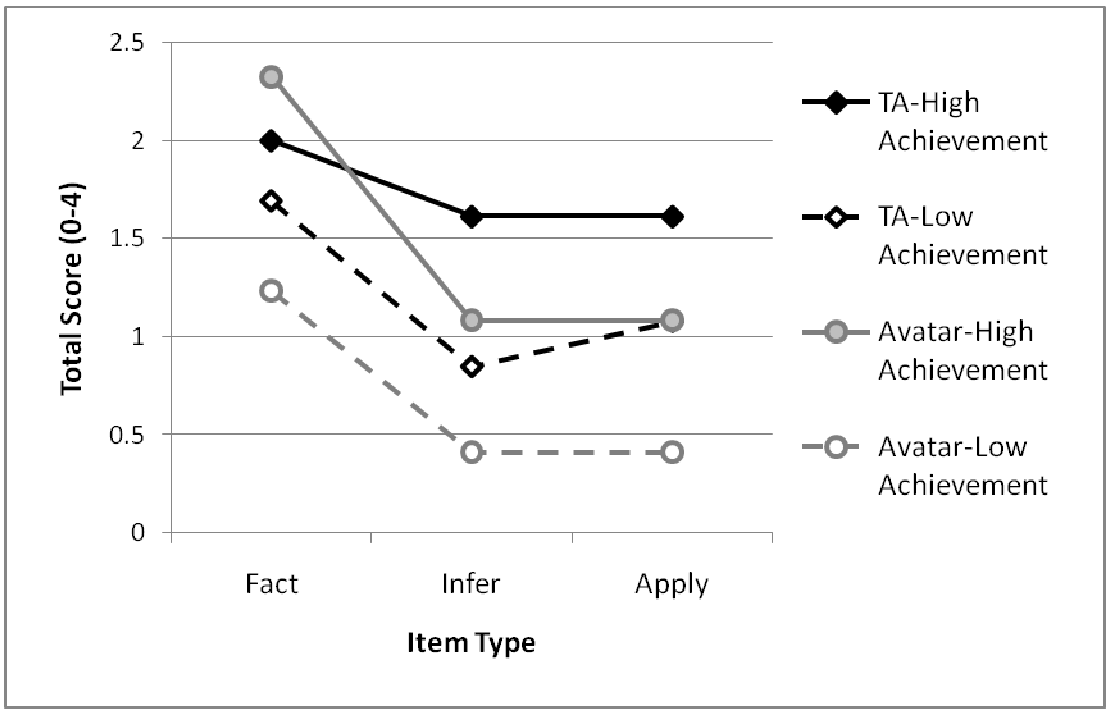


Figure 8.

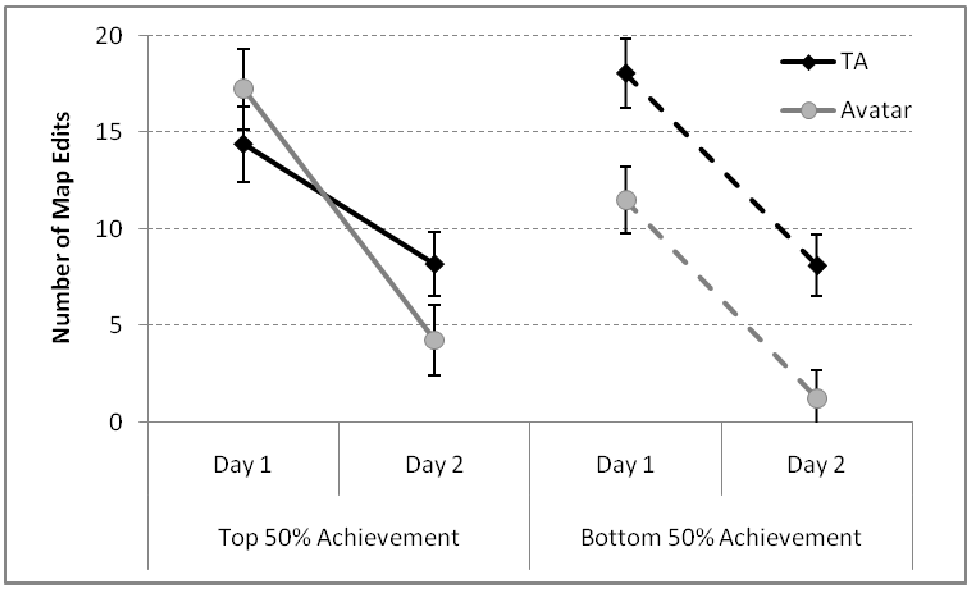


Figure 9.

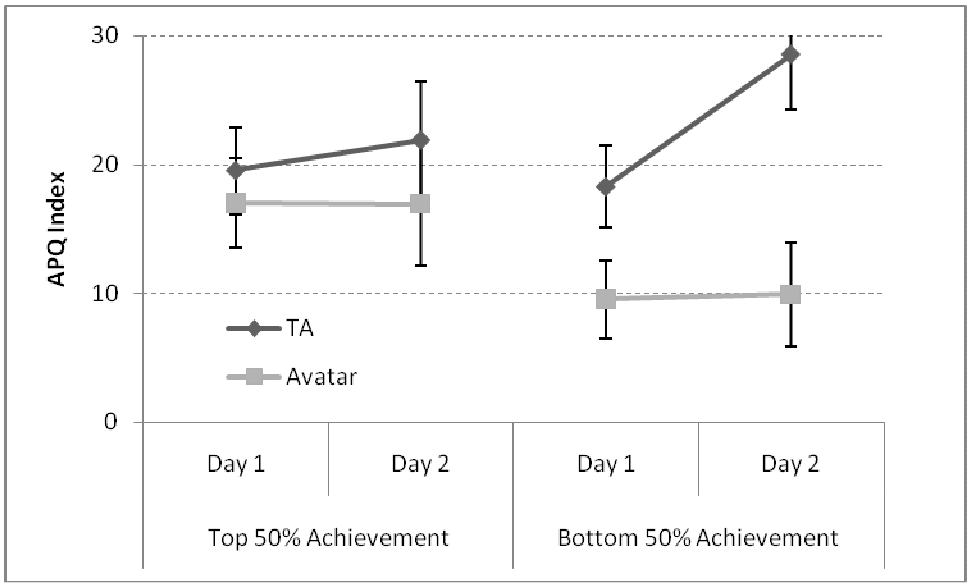


Figure 10.

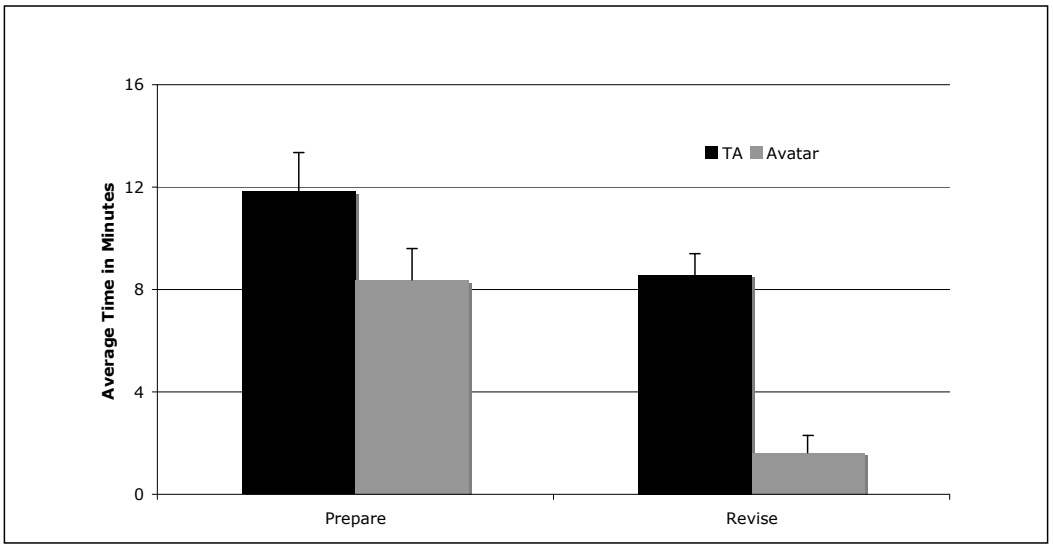


Figure 11.

APPENDIX

Fever Passage (Studies 1 and 2)

Many people worry when they get a fever. But, a fever can be a good thing. It's part of your body's defense system and means your body is working to kill an infection. A fever means the body is hot, and the heat helps to kill germs like bacteria and viruses.

How does the body increase its temperature? The brain has a set point that determines how hot the body gets. When the set point rises, it causes the body to get hotter. The set point rises when germs invade the body. When this happens, your brain tells the body that the temperature must be raised a few degrees to kill the germs.

There are four different ways the set point causes the body temperature to increase. One way is to decrease blood flow to the skin, by shrinking the veins (blood vessels). When less blood gets near the skin, the blood cannot release as much heat through the skin. This explains why people can have a fever but still feel cold in their hands and feet. There is less blood near the skin.

A second way is shivering. Shivering makes the muscles move. When muscles move, they produce heat. Shivering can make the body produce more heat than normal.

A third way is to raise body hairs. When the small hairs on the body stand up, pores (small holes) in the skin close. This means less heat can escape through the pores. It also means that less sweat can escape through the skin. When you have a fever, you sweat less, because sweating cools the body. Raised hair explains why a fever causes a person's skin to feel tender. The little hairs get rubbed and irritate the skin.

A fourth way is to increase the body's metabolism. A higher metabolism means that the body burns energy faster, and this causes it to produce more heat. Higher metabolism explains

why people have faster breathing and a faster heart rate when they have a fever. A body with high metabolism needs more blood and oxygen.

If the body gets too hot, it will begin to kill its own cells. How does the body stop from getting too hot? When the body temperature reaches the set point, all the processes reverse. Blood goes to the skin, shivering stops, the hairs lie down, and metabolism decreases. Aspirin and Tylenol help reduce a fever by bringing down the set point, so the body stops trying to heat up. The good thing about aspirin is that it makes you feel better. The bad part is that there is less fever to help kill the germs.

Posttest Questions (Study 1)

Factual

(a) Even though a fever feels bad, it can still be good for you. Why?

(b) If you hold hands with someone who has a fever:

The person's hand feels (circle one):

(a) DAMP (b) DRY

The person's hand feels (circle one):

(a) HOT (b) COLD

Integration

(c) Explain what body hairs have to do with causing a fever. If there are many steps in the process, be sure to describe all of them clearly.

(d) Why is shivering not enough to cause a fever?

Application

(e) Here is a common situation. People wake up all sweaty, and their flu is gone. Why are they sweaty?

(f) Why does a dry nose mean a dog might have a fever?

Posttest Questions (Study 2)

Factual

1. Why do your hands and feet get cold when you have a fever?

2. What does Aspirin or Tylenol do?

Integration

3. How does the body stop having a fever?

4. When do you know that your body is recovering, and why?

Causal Reasoning

5. If raised body hair increases, what happens to heat release? Why?

6. If bloodflow to the skin decreases, what happens to heat production? Why?

7. If temperature set point increases, what happens to heat release? Why?

8. If germs decrease, what happens to sweat? Why?

9. If shivering increases, what happens to body temperature? Why?