The Impact of a Social Robot's Attributions for Success and Failure in a Teachable Agent Framework

Kasia Muldner, Victor Girotto, Cecil Lozano, Winslow Burleson, Erin Walker
Computing, Informatics & Decision Systems Engineering, Arizona State University
{katarzyna.muldner, victor.girotto, cecil.lozano, winslow.burleson, erin.a.walker}@asu.edu

Abstract: Teachable agents foster student learning by employing the *learning by teaching* paradigm. Since social factors influence learning from this paradigm, understanding which social behaviors a teachable agent should embody is an important first step for designing such an agent. Here, we focus on the impact of causal attributions made by a teachable agent. To obtain data on student perceptions of agent attributions, we conducted a study involving students interacting with a social robot that made attributions to *ability* and *effort*, and to the *student*, *itself*, or *both*. We analyzed data from semi-structured interviews to understand how different attributions influence student perceptions, and discuss design opportunities for manipulating these attributions to improve student motivation.

Introduction

Interactive activities can be highly beneficial for learning, because they provide opportunities for knowledge construction through, for instance, contributions to group discussion or guidance from a knowledgeable partner (Chi, 2009). While pedagogical interactions have historically been student-student or student-teacher, as educational technology evolves, an emerging avenue has involved using pedagogical agents to foster learning through agent-student interactions (Woolf et al., 2010). One type of pedagogical agent is a teachable agent, which simulates the collaborative activity of peer tutoring, where it is the student who teaches the agent about the target domain. Prior work has demonstrated that there are cognitive and social benefits to peer tutoring (Roscoe & Chi, 2007), and by extension, to the teachable agent paradigm (Chase et al., 2009). In our work, we are interested in exploring how social interactions between a student and a teachable agent foster student engagement with the agent, and ultimately improve learning and motivational outcomes.

We focus on a specific type of social behavior, namely the *attributions* that a teachable agent makes. In general, attributions correspond to causal explanations one makes for successes or failures. According to attribution theory, the causes that students attribute to outcomes impact their motivation, affect, and reactions, as well as subsequent learning outcomes (Försterling, 1985). Moreover, listening to others' causal attributions influences the overhearing student's affect and social behaviors (Hareli & Weiner, 2002). This prior work, however, has been done in the context of a classroom with students. It is an open question, therefore, as to the impact of a teachable agent that makes causal attributions. We conducted a study gathering data from students' interactions with a social robot that makes different types of attributions to ability and effort. Our general research question was as follows: How do the different types of attributions made by a teachable agent impact students' perceptions of the agent and their desire to teach the agent?

To address this question, we rely on a robotic teachable agent platform that we have developed called TAG (*Tangible Activities for Geometry*) (Muldner et al., 2013). To the best of our knowledge, all related efforts using the teachable agent paradigm have focused on agents in virtual environments. In TAG, students instead walk around a projected space and interact with a physical robot. There may be several advantages to a robotic learning platform. First, a physical presence provided by a robotic agent strengthens users' perceptions of having a social partner more than a virtual agent (Powers et al., 2007). Second, students benefit from learning through embodied, physical interactions, particularly for abstract topics (O'Malley & Fraser, 2004), which robotic platforms naturally support. Thus, in a robotic learning environment, the effects of a robot's social behaviors may be heightened, and so it is a good platform for testing the effects of a teachable robot's attributions. In the remainder of this paper, we present related work on social behaviors in teachable agents, and virtual and robotic agents. Next, we describe the TAG system and its social robot Quinn, and then present a study investigating how students perceive Quinn's attributions. We conclude by identifying implications of the results for designing a teachable robot to engage students and ultimately improve their learning outcomes.

Background

Teachable agent systems have evolved out of personalized learning research and have yielded some early successes. Teaching a computer agent can lead to more learning than being taught by a computer agent (Leelawong & Biswas, 2008) and is nearly as effective as being taught by an expert tutor (Reif & Scott, 1999). Some of the teachable agent effect is due to the deep cognitive processes fostered by teaching: As in peer-to-peer tutoring, peer-to-agent tutors notice their own misconceptions and elaborate on their knowledge as they tutor their teachable agents (Biswas et al., 2005). Another factor responsible for the benefits of learning from teaching is motivational. For instance, students feel responsible for their agent students, and as a result try harder and attend more to subject material (Chase et al., 2009). To capitalize on these motivational aspects, researchers have begun taking steps to build social and affective behaviors into their agents. For example, Gulz et al. (2011) have incorporated off-task social conversation into their teachable agent, and demonstrated that this led students to learn and have a positive attitude. Others have begun to explore how conversational strategies such as teasing between human peers (Ogan et al., 2012b) and human-agent peers (Ogan et al., 2012a) impacts rapport with the human learner. These efforts are at an early stage, and so more work is needed to understand how to capitalize on social and affective elements within a teachable agent platform.

However, in the broader pedagogical agent literature, we do have some information on the impact of social agents, both in the virtual and physical domains. In general, social behaviors can have a positive impact on student perceptions and in some cases learning. As far as virtual agents are concerned, students preferred agents who display facial expressions over ones that do not (Baylor & Kim, 2008). Others have explored the impact of social behaviors in robots. Kanda et al. (2012) had students interact with either a task-oriented robot or one that also provided social support by praising and encouraging students. Although no learning difference was found between the two versions, students preferred the social robot and reported a stronger relationship with it. Saerbeck et al. (2010) showed that students who interacted with a socially-supportive robot (e.g., one with facial expressions) learned more and were more motivated than students who worked with a neutral robot. Leite et al. (2010) found that users reported higher feelings of companionship with a robot that empathized with them during a game of chess, as compared to a robot that did not.

Another promising way of socially engaging students is through the use of attributions. For instance, virtual agents that express attributions emphasizing the utility of perseverance have been shown to improve students' affect during problem solving (Woolf et al., 2010). The teachable agent paradigm gives us a unique platform for exploring the effects overhearing a teachable agent's attributions. We focus on attributions to effort vs. ability, and attributions to the peer tutor, the robot, or both. Our target set of attributions includes both "desirable" and "undesirable" ones, because we want to explore their impact on student perceptions. Some work outside of computational frameworks has highlighted ways that overhearing a student make attributions impacts the observer (Hareli & Weiner, 2002). For instance, for unsuccessful outcomes, hearing a student attribute the failure to low ability can trigger either pity or contempt in the observer, while attributions to low effort can elicit anger in the observer if they are teaching. For positive outcomes, less is known about how a teaching framing influences observers' attributions, but Hareli and Weiner (2002) speculate, for instance, that attributions to effort are perceived as modest and so can contribute to feelings of admiration in observers. While informative, work is needed to extend and refine these findings to computer environments with teachable agents.

Tangible Activities For Geometry (TAG)

The *Tangible Activities for Geometry* (TAG) system that we use as the test-bed for our work aims to help students learn about geometry by providing a projected space that students move in while solving problems and interacting with a robotic agent called *Quinn*. TAG comprises four main components (see Fig. 1, left): *Quinn*, the *problem space*, the *hanging pointer*, and the *mobile interface*. Quinn consists of a LEGO Mindstorms robot with an iPod mounted on top of it representing its face. The problem space is projected on the floor and includes a Cartesian plane and Quinn, which moves autonomously in this space (for details, see Muldner et al., 2013). The hanging pointer is TAG's version of a mouse. It corresponds to a small cylinder attached to the ceiling by a retractable wire, and is used to control a virtual pointer projected in the problem space (i.e., moving the hanging pointer results in the virtual pointer following it). To click on objects in the projected space, students hover the hanging pointer over the desired target (either a point or Quinn), and pull the pointer down towards the ground and back up. The mobile interface is an iPod Touch that lets the student see the current problem, move between problems, check for correctness of the current solution, and issue commands to Quinn (Quinn responds after each student instruction by executing that instruction).

To illustrate student interaction with TAG, suppose a student opens the problem: "*Plot the point (2, 1)*". When a new problem is opened, Quinn moves to the origin and faces east along the x-axis. To solve the problem, the student must walk over to Quinn and click on it using the hanging pointer. Clicking results in the student's iPod showing the list of available commands to give to Quinn. As the first step, the student could tap *move* and specify the distance '2' in his/her iPod, which results in Quinn moving two units along the x-axis. The student must then walk over to Quinn, click, and choose *turn in a direction*, specifying 'N', which results in Quinn facing North. The last two actions correspond to moving Quinn by 1 unit and then telling it to plot a point. When ready, students can tap a button on their iPod and correctness feedback is shown on that iPod.

We chose the current task domain, i.e., geometry, because of its conceptual properties. In theory, as students move over the projected coordinate system and gesture towards aspects of the projection, they can physically encode concepts such as how positive and negative coordinates relate to graphical quadrants.

Quinn's social behaviors. Quinn's social behaviors are based on attribution theory, and are generated after and in response to TAG's feedback for correctness, ostensibly representing the robot's reaction to whether

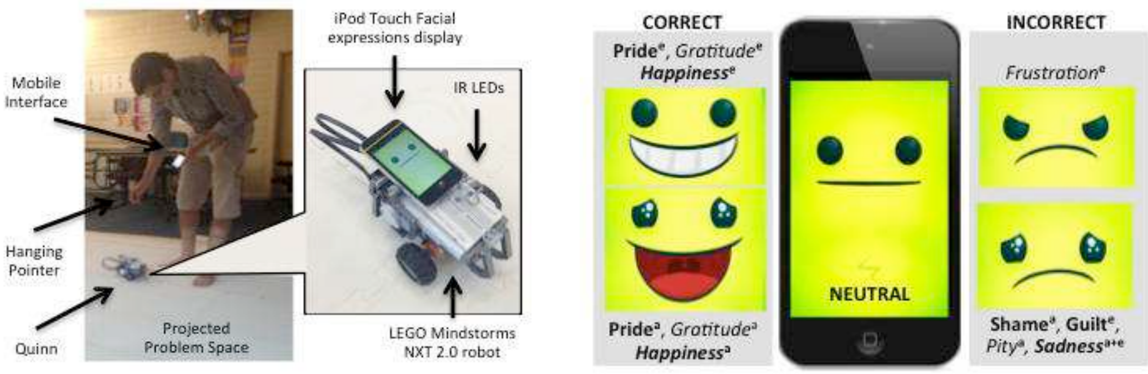

Figure 1: The TAG system (left) and Quinn's facial expressions (right, including text specifying attribution-specific emotions based on the attribution dimension: I (bold), you (italics), we (bold italics); a = ability and e = effort attributions. Students did not see these labels). Note that a single expression is shown at a given time.

it got the correct answer. Specifically, Quinn responds to TAG's feedback by displaying an emotion on its iPod (see Fig. 1, right) and by telling the student how it feels through a message spoken in a gender-neutral voice. In the message, Quinn attributes the outcome to factors along two dimensions: the cause of the outcome (*effort* or *ability*) and the agent responsible for that cause, namely itself (*I*), the student (*you*), or both (*we*). Thus, there are six messages for correct outcomes and six for incorrect outcomes (see Table 1). We focus on ability and effort because they are the most common attributions students use to explain academic outcomes (Hareli & Weiner, 2002). Some of Quinn's attributions are undesirable (e.g., attributing failure to lack of ability). We included the full spectrum of messages because we wanted to comprehensively explore the impact of various attributions on student perceptions in a teachable agent framework. As far as Quinn's facial expressions, attribution theory postulates that primary emotions for outcomes may be refined according to the attribution a student makes for the outcome's cause (e.g., pride if the individual caused the outcome vs. gratitude if a teacher did) (Hareli & Weiner, 2002). Since attribution-related emotions may be only subtly different, Quinn expresses a single primary facial emotion (see Fig. 1, right), shown for about 15 seconds. Quinn can also highlight the attribution-specific emotion in the spoken message (e.g., "*That was right. Oh man, I am smart at math*. **I feel proud").**

Table 1: Sample 6 of the 12 Quinn attribution messages (suffix specifying emotion not shown)

Outcome	***Cause (Agent/Source)***	***Quinn's Message***
correct	I / ability	That was right. Oh man, I am smart at math.
correct	you / ability	Yay! I got that right because you are a good teacher.
correct	we / ability	That was correct! My gosh, We are good at this.
incorrect	I / effort	Oh boy. I got that wrong because I did not try hard to learn.
incorrect	you / effort	You did not put in much effort into teaching me that problem.
incorrect	we /effort	Dang it, that was wrong. We did not work hard to solve that problem.

Students' Perceptions of Quinn: User Study

In order to obtain data on students' interactions with the social robot Quinn, and in particular the different attributions it makes, we conducted a user study. Our specific research questions included:

(1) What are students' reactions to Quinn?
(2) How do students respond to Quinn's causal attributions for failure and success?
(3) Are some attributions more appropriate for fostering social interactions in a teachable agent framework?

To answer these questions we had students solve problems in TAG, because we wanted to afford students the opportunity to interact with Quinn and thus be able to ground their perceptions in their experience. We then used semi-structured exit interviews as our primary source of data. We chose this methodology as it had the potential to provide richer data on students' experiences than, for instance, affective surveys.

Materials

The study involved the following materials: two cheat sheets (domain and a system commands), an attribution questionnaire, the TAG problems, and a set of solution cards. The domain cheat sheet reviewed the target concepts related to plotting and translation; the system cheat sheet described the set of TAG commands. The attribution questionnaire probed student attributions through a series of multiple-choice questions. Eight of the questions proposed a hypothetical situation and asked students to select the choice that best fit their reaction (from six choices representing the *effort/ability* and the *I/you/we* dimensions). Two questions included teaching-centric scenarios (e.g., *A friend that you have been tutoring in math has aced the math test*) and six were student-centric scenarios (e.g., *You have just received an A on your math test at school*). The TAG problems corresponded to two types: *plotting* of a point and *translation* of a point. The solution cards were 8x11 sheets of paper, one for each of the TAG problems; a given sheet was labeled with a TAG problem number on the front and a detailed description of the steps needed to generate that problem's solution in TAG on the back.

Participants

The study participants were 19 5th and 6th grade students (8 female) from a middle school in a large southwestern city. Students participated on site at their school, outside of regular classes but during regularly-scheduled classes (i.e., students individually left class to participate), and received $20. We chose students from grade 5 and 6 because these students already had some exposure to our target domain, but were not expert in it, as we identified by checking state standards and confirmed with pilot evaluations and discussions with teachers.

Procedure

Subjects (1) signed an assent form (~5 min.); (2) read the domain cheat sheet (~5 min.); (3) filled in the questionnaires (~15 min.); (4) were trained on how to use TAG (~20 min.); (5) used TAG to teach Quinn (45 min.); and (6) participated in a semi-structured interview (~20 min.). Sessions were conducted individually and were videotaped; two experimenters were present during each session. For the training phase, an experimenter followed a predefined script to go over TAG functionalities with each student. To implement the teaching framing, students were asked to *"tutor Quinn about how to solve geometry problems. The goal is for Quinn to learn enough so that it can solve all kinds of geometry problems. So when you are telling it how to solve a problem, think about what would be most useful for Quinn to know".* Students were also told that they could refer to the cheat sheets and the solution cards and that it was up to them as to how they used these. Since prior work indicated that when peers are friends certain social behaviors are associated with learning (Ogan et al., 2012b), students were asked to pretend that Quinn was a long-time friend; following the teachable agent paradigm, they were also told that it is Quinn who gets the answer correct or not. Students then "taught" Quinn by working through geometry problems. Once a problem was solved, TAG provided feedback for correctness, and Quinn responded by attributing its success or failure to one of the target attribution dimensions. Since we are exploring students' reactions, a given attribution was chosen at random. For a given student, an attribution was never shown twice before all attributions were used. We manipulated whether students also heard Quinn express the attribution-specific emotion in its message (n = 9), to obtain data on students' reactions to receiving this information verbally and explicitly rather than by interpreting Quinn's emotion from its face. As the final step, students participated in a semi-structured interview led by the lead investigator. The interview questions probed students' reactions (e.g., on whether they felt they were teaching), but also focused on Quinn's attributions (both ones they heard in the teaching phase and any remaining ones they did not). To increase realism, during the interview, the experimenter played Quinn's messages for students using Quinn's voice. After the interview, students were debriefed, by being informed that Quinn's interventions were chosen at random and were not directed at them personally.

Results

The interviews were transcribed and the data was analyzed through qualitative description. Specifically, we iteratively derived codes from the data, organized these according to emergent themes, and refined them as needed. Our goal was to provide a qualitative summary of students' perceptions. We also analyzed student responses to the attribution questionnaire, by creating an attribution profile for each student based on frequency counts for the "source cause" (*ability* vs. *effort*, collapsing across *I/you/we)* and the "agent cause" (*I* vs. *you* vs. *we*, collapsing across *ability/effort)*, for positive and negative outcomes.

While we focus our analysis on students' reactions to Quinn attributions, we begin by mentioning overall perceptions of Quinn, its facial expressions, and the teachable agent framing. When asked about Quinn, not a single student expressed dislike for the robot, despite some of its negative attributions. Instead, students had positive reactions, many specifically mentioning Quinn's social behaviors. For instance, students said that the thing they liked the most about TAG was "how Quinn showed his feelings" (s1), Quinn because it "was cute" (12), and how "the robot was talking and stuff to me" (s19); s11 echoed this by stating that "it's cool that he has face emotions". While s14 mentioned that he got frustrated "a little bit […] when Quinn … got it wrong and he was getting mad at me", when asked if a Quinn who did not speak or show faces would be preferable, he responded it would "be worse because then you wouldn't know what he would be feeling". S17 mentioned that Quinn's attributions made him feel "good" and s9 wanted to "take Quinn home" because it was "helpful". As far as Quinn's facial expressions, some students explicitly mentioned liking Quinn's faces (s4, s7, s12, s14, s16, s19). For instance, s16 stated that the faces "make it more fun, cause when you get it right he [Quinn] will be happy", adding that this made him happy. In contrast, some students mentioned no preference for Quinn having facial expressions (s2, s13, s15, s18). Students also did not express a clear preference for having Quinn verbalize its emotions (recall that for some students, Quinn suffixed an emotion following its attribution).

As far as our teaching framing manipulation, the majority of students felt they were teaching Quinn (s1, s3, s5, s7-s11, s13, s16, s18, s19). Some students mentioned that Quinn's attribution messages made them feel like this (s3, s8, s11, s16), e.g., "when she was saying positive things like I'm a good teacher" (s11), and when "he said that I taught him how to do it" (s16). S5 said it was because "I was showing him where ... the coordinates were" and s19 mentioned "by listening and stuff". Other students did not buy into the teachable framing (s2, s4, s12, s15, s17). S2 said this was because "she [Quinn] knew where to go already". The remaining students said that it was because they were "controlling" Quinn and that Quinn initiating actions would make them feel like they were teaching Quinn more. One student felt "the robot was teaching me" (s14). We now present the attribution results organized by incorrect and correct outcomes.

Student Perceptions of Quinn's Attributions for Incorrect Outcomes

For the *I/we* dimensions for incorrect outcomes, students recognized the utility of effort over ability because "she [Quinn] knows she should try harder and she might get it" (s11). Students also commented on the fact that the effort attributions made them feel like a teacher (e.g., "next time like try to learn while I'm teaching" s18). In contrast, for the *you* dimension we did not see clear differences between ability and effort - details are below. The interview data for effort attributions is aligned with the attribution questionnaire data, in that the majority of students selected effort for the student centric (n = 12) and teaching centric (n = 15) questions. As far as the "agent cause" dimension (*I/you/we*) in the questionnaire data, for the student-centric questions, students attributed to themselves (*I, n* = 14) or to *we* (n = 3); for the teaching-centric questions, students attributed to themselves (*I, n* = 11), to *we* (n = 5) or to *you* (n = 3) (minor variations in N are due to missing data).

***I* (Quinn) dimension / *ability* + *effort*.** Most students did not appreciate Quinn's attribution to its ability upon an incorrect outcome. S14 said it was not realistic "because he is not really dumb at math". Other students had been taught to not attribute failure to ability, and transferred this to Quinn (s4, s5, s7, s12, s13, s17, s18). S5 said that "you shouldn't call yourself dumb", while S7 said that it's "not nice saying that to himself"; S13 echoed this, i.e., "he's putting himself down". Some students learned these sentiments from their parents (e.g., "my mom told me not to say this", s4). S8 felt that Quinn shouldn't take all the blame, because "it is both our fault". Some (s11, s17) didn't like the message "because she [Quinn] is putting too much pressure on herself - she thinks that she can't do it but we all know she can do it" (s11). Students also mentioned feeling sad for Quinn when they heard the message (s12, s17). When asked how the attribution would make them feel in terms of teaching Quinn, s12 answered "Quinn shape up!", while s17 wanted to "make it feel better" by giving it "an easy problem". In contrast, s3 liked the message, saying that it would make her want to "teach him more"; s10 echoed this sentiment. S6 said the message was funny, empathizing with it more than the effort attribution because "you can try hard and still not get it".

In contrast to Quinn's attribution to ability, more students appreciated Quinn attributing the negative outcome to effort (s4, s7, s9, s14, s18, s19), for instance "because he's being honest that he was not really listening" (s4). S5, who got the message appended with Quinn's attribution-specific emotion (guilt), also liked the attribution, but did not think Quinn should feel guilty, so as to not "sound hopeless". In contrast, s1 mirrored Quinn's guilt, i.e., "I would also feel guilty because I am teaching him", and that this would make him want to teach Quinn more. Several students, however, did not buy into the message (s1, s8, s13, s16). S1 did not believe that Quinn failed due to lack of effort, because "I'm doing everything for him to get him the right answer"; likewise s16 said it was "who controlled it didn't try hard [instead of Quinn]". Along a similar vein, s8 stated that "it wasn't her fault, it was mine", but also added that Quinn "could try a little bit harder".

***You* (student) dimension / *ability* + *effort*.** When Quinn attributed the negative outcome to its teacher (i.e., the student), we did not see a clear difference in student perceptions between ability and effort and so collapse them here. Some students had a negative reaction (s1, s6, s9, s10, s14). S9 said he did not think Quinn was teasing him and that the message was "hurtful … [Quinn] needs to get … respectful"; s2 felt "sad that like I'm not teaching Quinn hard enough to get the problem right", but he also expressed that it "would make me try teach the robot more". Some students expressed frustration: S14 said the message made him "mad at him cause I'm trying to help him" and so might feel less like teaching Quinn, while s1 mentioned it hurt his "feelings a little" and as a result he "felt a little frustrated with [Quinn]", because Quinn "didn't say I'm sorry for anything"; likewise, s10 stated "I wouldn't help [Quinn]". S6 also said the attribution made him "a little mad", because "he is just blaming me". S11 had mixed feelings: Although she expressed preference for other attributions, she said this one "makes me feel like she doesn't think that I can do better but I think I can show her that I can", adding that it would both make her "a little mad" and also make her try harder.

Other students had more positive reactions (s2, s4, s8, s16-s19). S4 wanted "to help him more so that he would succeed"; this was echoed by s17. Some students said the message was fine because it came from a robot: S8 responded with "it was just a robot", and found the message amusing. Likewise, s16 said "it's a small robot, so I don't know how it could hurt your feelings", also mentioning the message was fair game if "you did badly". S2 said that if it was coming from an "actual person … then I'll probably get mad but it's coming from a robot that doesn't really mean it". S17 agreed with Quinn that he didn't teach well because "I couldn't do translation" and while its message did not make him feel "bad", it did make him feel "guilty".

***We* dimension / *ability + effort*.** When Quinn attributed failure to itself and the student (*we*), some students mentioned liking the "we aspect" without responding specifically to agent causing it (s9, s11, s16, s12, s19). S12 said "it's okay with me because he said *we*, so it's not only him [and] it's not only me that did all the work". Likewise, s11 felt the message was appropriate because it "partly my fault and its partly her fault" qualifying with "I don't know if its all of her fault because she is new [..] and she hasn't done this a lot". S16 said the message made sense as Quinn "said *we,* because he followed me and I did it wrong, so me and his fault". In contrast, s3 said that Quinn was "kind of blaming" her for the outcome.

Reactions to the ability attribution in the *we* dimension were mixed. S15 felt it meant "that I have to be better at math" and make him try harder; when asked why him and not Quinn, he responded with "because I'm making it". S2 thought the message was "fine" but cautioned that "other people might not". In contrast, s11 did not like the attribution because she knew "she was fine at math". S17 felt the message unfairly blames both parties: "if you're saying you're bad, it's not like the one who's working with you is bad too". Although s18 found the attribution "funny", he also thought it felt "weird". Thus, in general, the majority of students explicitly mentioned preferring the effort attribution (s1, s2, s4, s5, s7, s9-s14). S9 empathized with the effort aspect, by saying that "sometimes at tests cause I'm so nervous I just guess". Others said this was because the ability one made them feel "something bad" (s4) or was "hurtful" (s12). S10, s13 and s14 elaborated on this sentiment by saying the effort attribution "wouldn't let the other person down […] and could help each other work harder" (s10); "well he's not putting anyone down he's just saying we did not try very hard on that" (s13) and "it's not like you suck at math […] you're just not working hard enough" (s14). S7 added that it made him feel like "if we try it again together we could figure it out". S4 said that the effort attribution would make him "try harder" over the ability one. S1 thought it was more realistic to attribute to effort "because well we've done a few problems, but it's not like we're really bad at it", a sentiment echoed by s5, i.e., "everyone is ok when working on math - no one is perfect". In contrast, s17 commented on the fact that Quinn did not know how hard she worked ("it's not like he can tell by looking") and so felt the effort attribution was "a little mean spirited".

Student Perceptions of Quinn's Attributions for Correct Outcomes

In contrast to incorrect outcomes, for correct outcomes the majority of students did not specify a preference for ability over effort for the *I* and *you* dimensions, but did prefer effort for *we* dimension. While students responded positively to all three of the *I/you/we* dimensions, they were especially enthusiastic about *we*. The student profiles extracted from the attribution questionnaire indicate that the vast majority of students selected effort for the student-centric (n = 18) and teaching-centric (n = 17) questions. As far as the "agent cause" dimension (*I/you/we*) in the questionnaire data, for the student-centric questions, students attributed most often to themselves (*I,* n = 12), followed by *you* (n = 2) and *we* (n = 1); for the teaching-centric questions, attributions were similar for *I* and *we* (n = 7 and n = 8) and less common for *you* (n = 3).

***I* dimension (Quinn) / *ability + effort*.** Many students responded positively when Quinn attributed success to either its ability or effort (s1, s4, s5, s7, s11, s12, s14, s16, s18, s19). For instance, s1 said Quinn "felt proud of me making him make the right choice" and s11 agreed with the attribution because "she [Quinn] is smart at math". Reactions to students who heard Quinn expressing the emotion (pride) in the message were mixed. S16 mirrored the pride ("I taught him and made him good at it") while s5 felt Quinn was "boasting a little"; this was echoed by s17 who said Quinn was "a little too proud". S12 who got the attribution without the emotion also felt like Quinn was bragging, adding "think of the other students". Thus, some students preferred the effort attribution (s4, s7, s11, s12), recognizing the utility of effort over ability, e.g., "if you try hard you usually succeed" (s4). S7 empathized with the effort attribution "cause that's like me, I work so hard until I finally get it", while s11 felt the attribution made her feel like she was teaching Quinn "because she tried harder to do that problem ... and got it right the second time". Several students could not pick between effort and ability attributions (s1, s5, s18), with s1 saying that "both have good feelings not against others". Students who did not like the attribution said it was because Quinn was taking all the credit. S14 responded by saying "*we* are smart at math"; s8 explained that "it makes me feel a little sad cause she's saying that she's the only one that did it". Likewise, s9 felt that Quinn was mean because "he didn't put me in that sentence".

***You* dimension (the student) / *ability + effort*.** When Quinn attributed its success to the teacher (i.e., student), the majority of students did not specify a preference between the effort and ability, and most reacted positively to both. To illustrate, students said "that is really, really – it's nice and kind of him to say" (s1), "that's very nice" (s12), "that's a good one" (s14), "I feel really good about that" (s4), that it felt "like I've accomplished something" (s13). Students also described how the attribution made them want to teach Quinn. S4 said "because she [Quinn] said I was a good teacher and I didn't want to say like - no I don't want to do this anymore". In response to the effort attribution, s1 said that he liked the message because "I taught him … it's like we did it together." In contrast, s13 wanted Quinn to acknowledge that they were working hard and so preferred the effort attribution, while s3 did not like the effort attribution as it meant Quinn was not putting in effort itself, i.e., "that would say that he just listen to me for the answer and he didn't do anything".

***We* dimension / ability + effort.** S14 liked attributions for success that included them and Quinn (i.e., *we*) because "if I was the only one who did it he'd feel bad … so when it's both of us it's like we're both getting the same share". S1 said *we* was best because "he was really happy that we both did it"; s6 "I like that out of all of them". S18 stated that "expresses a lot because the person controlling him feels great because he was teaching him and it makes them feel good to keep trying". S10 said in response "she [Quinn] was right that we are good at this, and that she tried harder at the question and she knew what she was doing". Likewise, students also liked the effort attribution (s1, s6, s7, s17, s19), because it made them "feel good" (s6). When asked to compare effort to ability attributions, the majority preferred effort (s3, s6, s9-s12, s17 - s18). S3 said this was because the effort attribution "means we learned… if we were to say we are good at this, then that means we already knew it." S18, s6 and s9 liked that the message acknowledged their effort (e.g., "geometry is hard and I had to help Quinn… so we both worked hard", s6), with s9 adding that because they worked hard Quinn "is happy". S10 picked the effort attribution, but also added that the ability one "gets my hopes up". S17 said that both attributions were cheerful, but also picked the effort one. Other students did not have a preference between the two attributions (s5, s8, s14), saying, for instance, that they both "sounded good".

Discussion and Future Work

Weiner's theory proposes that students should attribute to high ability or high effort for correct outcomes, and to low effort for incorrect outcomes (Försterling, 1985). The majority of our results for the type of ability and effort attributions students feel *Quinn* should make mirror this finding. In the future, we plan to use a computational model that assesses a student's effort and takes into account prior history of success and failure to select Quinn's attributions. For instance, the robotic agent could attribute success to effort using the *you* dimension ("you tried hard to teach me") when the student struggles though an especially challenging problem. We now explore how Quinn might take on these adaptive attributions.

For incorrect outcomes, our analysis suggests that attributions to effort in the *I* (Quinn) and *we* dimensions are safe for a social robot to express, in terms of maintaining positive student affect and motivation to teach. Prior to the study, we anticipated students may have preferred to shift the responsibility of an incorrect outcome on the agent, as found in non-educational settings (Groom et al., 2010). However, the pre-study questionnaire data indicated that for incorrect outcomes, for both student-centric and teaching-centric questions, students most often attributed failure (and success) to themselves. Interestingly, interview data revealed that when actually teaching, the *we* dimension got the most enthusiastic responses for incorrect outcomes – this finding is aligned with the fact that teaching is inherently a social activity. Students appreciated spreading the blame attribution between themselves and Quinn, which suggests students felt responsible for the robot, and so did not want it to take all the blame. When Quinn instead attributed failure to the student teaching it (the *you* dimension), most students responded with frustration, stating that the message was not motivating them to teach Quinn. Prior work with students in peer-to-peer interactions showed face threatening moves of this type were in some cases associated with learning (Ogan et al., 2012b), and in fact we did find that some students had a positive reaction, even saying that the message made them want to teach more. A promising avenue, proposed by one of the students, is to have Quinn exhibit playful affect when stating such messages (this student suggested Quinn should stick out its tongue). Thus, our results suggest that this *you* dimension should only be used for students who will interpret it as a playful challenge, rather than as unproductive rudeness. As far as the *I* dimension, students reacted negatively when Quinn attributed incorrect outcomes to its own low ability. Prior work suggests that this should elicit either sympathy or contempt in the observer (here, the student teaching Quinn). While one student did express sympathy for Quinn, for the most part students instead expressed irritation (and possibly contempt) that Quinn was putting itself down. When Quinn attributed failure to low effort, we did not see the reaction forecast by prior work, namely anger. Instead, most students reacted positively, some adding that the attribution would want to make them teach Quinn more. *In sum, these results suggest that for incorrect outcomes: (1) students prefer we attributions that share the blame between them and Quinn, but (2) you attributions might be motivating for students who see them as a playful challenge, and (3) I attributions could motivate students to try teach Quinn the relevant concepts*.

For correct outcomes, our data indicates a slightly different picture in terms of design recommendations for social teachable agents, in that all six attribution types across the ability/effort and I/you/we dimensions were positively received. Thus, having a robotic agent attribute to all dimensions, with varying degrees of frequency, could increase believability and student bonds with the agent. In particular, the *we* attribution was very popular, with students appreciating Quinn, recognizing their role in the successful outcome (a correct solution). However, the *you* and *I* dimensions also fared well. For the *you* dimension, students appreciated hearing compliments about their teaching, suggesting that they bought into the teaching framing to some extent. Likewise, both *I* attributions were well received, and compared to incorrect outcomes, for correct ones there was not as strong a negative response to Quinn's ability attributions, with some students mirroring Quinn's pride expressed in this attribution. While the effort attributions for the *I* dimension may have been more positively received than ability ones, having Quinn attribute to both ability and effort could add to the realism of its design. It is interesting to note that students responded enthusiastically to the effort attributions made by Quinn upon a correct solution, for instance empathizing with Quinn because they also had to try hard to learn. These reactions are in opposition to prior work indicating that when students attribute success to effort, they may be seen as less capable (Försterling, 1985). *To summarize, upon a correct answer: (1) students perceive all of Quinn's attributions positively, but (2) you attributions might be particularly effective for students who need to boost confidence in their teaching abilities, and (3) I attributions to effort might encourage students to take on effort attributions for their own problem-solving.*

To conclude, prior work has shown that when students make appropriate attributions with respect to themselves, they persist more and learn better. There is much less work exploring the impact of others' attributions has on observers who overhear these. Here, we took the first steps in filling this gap by investigating students' perceptions of a teachable robot that attributes success and failure to different causes, showing that students' perception of such a robot is influenced by various attributions it makes.

References

Baylor, A., & Kim, S. (2008). The Effects of Agent Nonverbal Communication on Procedural and Attitudinal Learning Outcomes. In *Proc. of Intelligent Virtual Agents*, 208-214.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at, V. (2005). *Learning by teaching: a new agent paradigm for educational software. Applied Artificial Intelligence*, 19(3-4), 363-392.

Chase, C., Chin, D., Oppezzo, M., & Schwartz, D. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, 18(4), 334-352.

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

Försterling, F. (1985). Attributional retraining: A review. *Psychological Bulletin*, 98(3), 495-512.

Groom, V., Chen, J., Johnson, T., Kara, F. A., & Nass, C. (2010). Critic, compatriot, or chump? Responses to robot blame attribution. In *Proc. of Human-robot interaction*, 211-218.

Gulz, A., Haake, M., & Silvervarg, A. (2011). Extending a Teachable Agent with a Social Conversation Module - Effects on Student Experiences and Learning. In *Proc. of AIED 2011*, 106-114.

Hareli, S., & Weiner, B. (2002). Social Emotions and Personality Inferences: A Scaffold for a New Direction in the Study of Achievement Motivation. *Educational Psychologist*, 37(3), 183-193.

Kanda, T., Shimada, M., & Koizumi, S. (2012). Children learning with a social robot. In *Proc. of HRI*, 351-358.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain System. *International Journal of Artificial Intelligence in Education*, 18(3), 181-208.

Leite, I., Mascarenhas, S., Pereira, A., Martinho, C., Prada, R., & Paiva, A. (2010). "Why Can't We Be Friends?" A Game Companion for Long-Term Interaction. In *Proc. Of Intelligent Virtual Agents,* 315-321.

Muldner, K., Lozano, C., Girotto, V., Burleson, W., & Walker, E. (2013). *Designing a Tangible Learning Environment with a Teachable Agent.* In Proc. of Artificial Intelligence in Education , 299-308.

O'Malley, C., & Fraser, D. S. (2004). *Learning with Tangible Technologies*, Literature Review series, report 12.

Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., & Cassell, J. (2012). "Oh, dear Stacy!" Social interaction, elaboration, and learning with teachable agents. *In Proc. of CHI 2012*, 1381-1390.

Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012). Rudeness and Rapport: Insults and Learning Gains in Peer Tutoring. *In Proc. of ITS 2012*, 11-21.

Powers, A., Kiesler, S., Fussell, S., & Torrey, C. (2007). Comparing a computer agent with a humanoid robot. *In Proc. of Human-robot interaction*.

Reif, F., & Scott, L. A. (1999). Teaching scientific thinking skills: Students and computers coaching each other. *American Journal of Physics*, 67(9), 819-831.

Roscoe, R. D., & Chi, M. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4), 534-574.

Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. *In Proc. of CHI'10*, 1613-1622.

Woolf, B., Arroyo, I., Muldner, K., Burleson, W., Cooper, D., Dolan, R., & Christopherson, R. (2010, June 2010). The Effect of Motivational Learning Companions on Low Achieving Students and Students with Disabilities. *In Proc. of ITS'10*, 327-337.

Acknowledgements

This research was funded by NSF 1249406: EAGER: A Teachable Robot for Mathematics Learning in Middle School Classrooms and by CAPES Foundation, Ministry of Education of Brazil- DF 70040-020.