The Interaction of Children's Concepts about Agents and Their Ability to Use an Agent-Based Tutoring System

Alicia M. Hymel[1], Daniel T. Levin[1], Jonathan Barrett[2], Megan Saylor[1], Gautam Biswas[2,3]

[1] Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37240, USA
[2] Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37240, USA
[3] Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN 37240, USA
{Alicia.M.Hymel,Daniel.T.Levin,Jonathan.I.Barrett, Megan.Saylor}@vanderbilt.edu, Biswas@eecsmail.vuse.vanderbilt.edu

Abstract. Computer-based teachable agents are a promising compliment to classroom instruction. However, little is known about how children think about these artificial agents. In this study, we investigated children's concepts about the intentionality of a software agent they had interacted with and tested whether these concepts would change in response to exposure to the agent. We also tested whether individual differences in concepts about agent intentionality would affect children's ability to learn from the agent. After repeated exposure to a teachable agent, students did not make more intentional attributions for the agent than a computer, but a general understanding of agency predicted success in learning from the agent. Understanding basic concepts about agency appears to be an important part of the successful design, implementation, and effectiveness of computer-based learning environments.

Keywords: learning, artificial agents, theory of mind.

1 Introduction

Computer-based learning environments have proven to be a valuable resource for both students and educators. Of particular promise are environments featuring teachable agents – graphical representations of characters that students can teach using speech, text, or visual representations (e.g., [1]). A central assumption underlying these systems is that the agents exhibit behaviors that are similar to human characters, which may invoke cognitive processes for learning and monitoring of one's learning more effectively than less social systems. However, little is known about the specific cognitions that underlie this putative benefit, and almost nothing is known about the relationship between students' specific beliefs about these agents and their more general understanding of concepts such as agency. Although students

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2011, LNCS 6764, pp. 580–589, 2011.

most likely understand that these characters are not actually people and that they do not have human "brains" or a fully functional cognitive system, it is not clear whether students attribute specific cognitive properties to these agents that are similar to those they would attribute to humans. In addition, these agents are clearly designed to invoke broad concepts about intentionality and agency, but an important possibility is that they go beyond invoking these concepts to actually changing them. Conversely, if human-like artificial agents more effectively invoke cognitive and metacognitive processes than agents dissimilar to humans, individual differences in beliefs about artificial agents can affect the use of these systems. Thus, we believe that the link between concepts about agency and utilization of teachable agents is a two-way street, and that concepts about agents will not only affect the use of these systems, but this experience will affect concepts about agents.

To explore children's concepts about teachable agents and the interaction between these concepts and children's use of the systems, we employed a newly developed "behavioral prediction" measure of concepts about agency to understand how students think about a specific teachable agent called Betty. We first measured children's behavioral predictions for multiple agent types to determine whether their beliefs about agency differ from adults. We also tested whether students who used the teachable agent system for extended lessons on climate change and food webs were better able to predict differences in agency between machines and people, and, conversely, whether students who showed a particularly strong understanding were more successful in learning the course content presented in the system. Below, we begin by summarizing basic principles of agency and theory of mind, review a previously utilized measure of intentionality, and discuss reasons why this interplay between concepts about agents and the use of teachable agent systems may occur.

1.1 Basic Concepts About Agency

The body of research on theory of mind development constitutes a good starting point for understanding concepts about software agents. Using theory of mind, individuals are able to comprehend the mental states of other agents, as well as predict their future thoughts, emotions, and actions [2]. These predictions can be guided by how the agent is believed to mentally represent information. Intentional representations are characteristic of human thought, and are closely linked to their referents. One referent cannot be freely substituted for another, as the contents and meaning of that referent are important [3]. Non-intentional representations, on the other hand, are more characteristic of computers. These representations are less closely linked to their referents [4]. By this definition, representations serve as symbolic placeholders that the system acts on with little importance placed on their contents. One way of summarizing this contrast is to suggest that intentional representations reflect truly situated semantic knowledge about the world while non-intentional representations are more like abstract pointers that confer little real knowledge to a representing system that does not really "know" the true meaning of the representations.

Over the first few years of life, children come to understand people's actions as being driven by intentional mental representations that underlie beliefs, desires, and goals. A key component of this understanding is illustrated in a study conducted by Woodward [5]. Infants were shown either a human hand or an inanimate object (e.g.,

a mechanical claw) moving towards one of a pair of objects. After several repetitions of this event, the locations of the two objects were switched, and the hand or inanimate object either moved toward the previously reached-for object in its new location, or the new object, now in the previously acted-upon location. Woodward hypothesized that the hand that moves towards the previously reached for object in the new location is behaving consistently with intentional thought processes. The action is explainable based on a goal that is supported by an intentional representation of a real-world object, and because this goal has remained consistent across all trials infants should not be surprised to see the hand reach to the previous object even though it is in a new location. Alternatively, the hand that moves towards the new object in the previously used location is behaving consistently with non-intentional thought processes. Rather than acting upon an object, this agent is repeatedly acting on a location, meaning that the goal object can be freely substituted across trials without consequence.

Woodward found that when a human hand was reaching, infants looked longer when it reached towards the new object at the old location. In contrast, when the actions were performed by the stick-like inanimate object, there was no difference in looking time. This finding suggests the infants interpreted the reach by the hand as a goal-driven intentional action on an object, but did not do so for the inanimate stick. We used the underlying logic of this paradigm to construct a questionnaire to investigate adults' understanding of the mental representations and actions of human and non-human agents. Participants were asked to make predictions about the behavior of multiple agent types (e.g., a human, an ambiguous mechanical agent, and a computer) after reading a description of the agent's previous actions. This description was held constant across all agents. Participants then selected one of two actions as the agent's next step, one of which was consistent with intentional reasoning, while the other indicated non-intentional representations.

1.2 Measuring Attributions of Intentionality

In previous experiments using our measure (see Methods section for questionnaire details), we found that adults are able to distinguish goal-directed human representations from non-intentional mechanical representations and believe that ambiguous agents, such as robots, are similar to computers [6]. This is likely influenced by individual differences in beliefs about ambiguous agents, as individuals who believe that non-human agents are good at understanding human goals do not demonstrate as large of a disparity in their behavioral predictions for the two agent types [6]. These results do not seem to be driven by perceived limits in current technology [7] or by perceived intelligence of the agents [6].

Follow-ups to these basic findings have demonstrated that adults' expectations can be modified by experience. For example, when participants were asked to track a robot's focus of attention as it looked at objects and to remember which object the robot preferred, participants believed the robot more likely to engage in intentional thought than a computer [6]. However, simpler manipulations such as giving the robotic agent anthropomorphic labels (e.g., a human name and describing it as having goals) or showing participants a video of the robot did not lead subjects to differentiate it from computers. These results suggest that interactivity and continuous

monitoring of a nominal mechanical agent's mental state may begin to generate expectations for intentional representations in these agents.

1.3 Causes and Effects of Intentional Attributions

While previous experience likely influences beliefs about agents, it is not clear whether any given experience will increase or decrease perceived differences among agents. Some findings imply that extended experience with a mechanical agent is associated with attributions of intentionality to that agent, which would lessen differences in attributed intentionality between humans and machines over time. For example, Nass and Moon [8] found that even expert computer users with extensive email and word processing experience act as though computers have inner states and the ability to feel emotional pain. Conversely, experience may increase the apparent difference between humans and machines. Even infants are experienced enough to generate differing expectations about the movement and goals of people and inanimate objects [9, 10] and are able to distinguish between the object-directed actions of a person and location-directed actions of a machine [5].

While experience affects concepts about agents, it is also possible that concepts about agents affect one's experience with them. For example, it is possible that users who are able to fully distinguish teachable agents from real intentional agents have a learning advantage. Such users may have a more mature understanding of theory of mind, and would therefore be more prepared for the ways such artificial agents differ from humans, lessening the likelihood that unexpected differences cause misunderstandings that interfere with learning. Such increased understanding could ensure metacognitive monitoring techniques promoted by a teachable agent are successfully utilized.

1.4 Exploring the Interaction between Concepts and Experience Using a Teachable Agent System

The relationship between concepts and experience with agents is complex and could affect one's ability to successfully use a system featuring an artificial agent, such as a teachable agent system. The ambiguity introduced by working with an agent that neither appears completely human nor completely mechanical may have a particularly strong effect on children, whose theory of mind is not fully developed. In this experiment, we utilized a teachable agent system ("Betty's Brain") to investigate whether kids differentiate agents in a manner similar to adults, whether concepts about an agent affect children's ability to learn from the agent, and whether concepts about artificial agents are affected by extended experience with them. Students interacted with a teachable agent (Betty) through a series of classroom lessons in which they were asked to teach Betty and give her quizzes to assess her understanding. Betty engaged with the students and expressed a desire to learn. We used our previously developed questionnaire, described below, to measure students' ability to predict the mental representations and actions of a human, a computer, and Betty.

If students' experience affects their concepts about agents, this will be reflected in their behavioral prediction questionnaire responses. If experience with an agent is associated with intentional attributions, repeated exposure to the Betty agent should

lead to increased attributions of intentionality for Betty, with no changes in attributions for other agents. However, if experience with a mechanical agent helps to differentiate between humans and machines, students may expect Betty and the computer to perform non-intentionally after interacting with Betty. For example, students may discover that the artificial agent is not capable of reasoning in a manner similar to their own. Alternatively they may draw upon the experience to deepen their understanding of the differences between machines and people, and predict both more intentional behaviors for people, and less intentional behaviors for the computer and Betty. Finally, if concepts about agency affect students' ability to learn from the teachable agent, there should be a correlation between behavioral predictions and quality of students' concept maps in the Betty's Brain system.

2 Methods

2.1 Participants

Participants were recruited from five classrooms in a Nashville, Tennessee public middle school. A total of 108 7^{th} graders (57 experimental and 51 control) were enrolled in the study, and 74 students (69%) completed both the pretest and the posttest. Age and sex were not collected from the participants. Informed consent was obtained from all students and at least one legal guardian of each student.

2.2 Materials

Students completed our previously developed behavioral prediction questionnaire in which they were told they would be asked questions about Betty, a computer, and a person. They were given pictures and a short description of each agent. For example, the description of Betty stated: "Betty is a part of the teaching exercise you just completed. Think of how Betty acts and what makes her think." Next, participants made behavioral predictions. The first prediction scenario drew closely from Woodward's infant paradigm. Participants saw an image of a pair of objects on a labeled grid and were instructed to imagine that the agent chose one of the objects (the toy duck at location A-1). The second object (toy truck at location C-3) was not mentioned. They were then prompted to imagine that the agent repeated the action. Finally, subjects were shown an image of the toy duck and toy truck after their locations had been switched, and were asked to choose which of the two objects the agent would select (the toy truck at A-1, or the toy duck at C-3). Following from Woodward's results, if participants believe the agent to be acting in an intentional and goal-directed manner, they should predict that the agent would select the same object, that is, maintain the same goal. However, if participants believe the agent to be acting in a rote or non-intentional manner, they should predict the agent would maintain its movement pattern without regard to goal state by reaching to the new object at the old location. After completing the behavioral prediction for the first agent, participants made predictions for each of the remaining two agents.

Next, participants completed a second set of behavioral predictions. In this scenario, participants saw seven objects arranged in a horizontal line. The first, third, fifth, and sixth objects were writing instruments, while the second, fourth, and

seventh were of varying object type (spoon, scissors, and screwdriver). Participants were told that the agent reached towards the first, third, and fifth items, all of which were writing instruments. They were then asked which object the agent would choose next: the writing instrument located in the sixth position, or the non-writing instrument located in the seventh position. Responses were coded as intentional if the participant selected the writing instrument, as it serves a similar purpose, and helps achieve the same goal, as the previously selected objects. Selecting the non-writing instrument in the seventh position was considered non-intentional as while the spatial pattern of reaching was maintained, the object's purpose was different from the previously selected items.

Participants then completed the final set of behavioral predictions. Participants saw a picture of six items and were told that these items were shown to the agent. The objects could be categorized in one of two ways: taxonomically (office supplies and candy), or perceptually (large, dark, rectilinear objects and small lightly-colored objects). Participants were shown each of these two groupings and were asked to select which way the agent would organize the objects. The use of a taxonomic, or category-based, organizational strategy was considered to be representative of intentional thought, as the objects are grouped according to their function and the intent of their creators [11]. Perceptual categorization relying on the examination of surface features was considered non-intentional, as it requires no knowledge of object meaning or goal states.

2.3 Procedure

Students were assigned by classroom into either the experimental or control condition. Students in the experimental condition used Betty's Brain to complete lessons, and students in the control group completed traditional classroom assignments to learn the same material. Students in both groups learned about arctic climate change and food webs.

Betty's Brain is an agent-based learning environment in which students create causal concept maps to teach Betty, an interactive agent represented by an animated face. The software was designed to promote and reinforce metacognitive techniques, such as knowledge state monitoring, as students must ensure that Betty understands the material sufficiently for her to perform well on quizzes. Students use the Betty's Brain program by reading provided resources and identifying the causal relationships present among concepts described in the text. Concepts and their causal relationships are later entered by each student into a concept map. Students are able to ask Betty questions about the concepts or direct her to take quizzes from a mentor agent (Mr. Davis) to assess her learning. Betty's answers are always logically drawn from the student's concept map and may be incorrect if the concept map is erroneous or incomplete. An example of the Betty's Brain user interface is provided in Figure 1.

Betty is programmed to interact with students as they construct their concept maps. Betty encourages students to read the resources and learn new information so they can teach it to her. She is able to initiate conversations by restating recently taught knowledge and its effects on established causal chains. Betty also requests that students ask her questions to ensure she understands new causal relations. Students

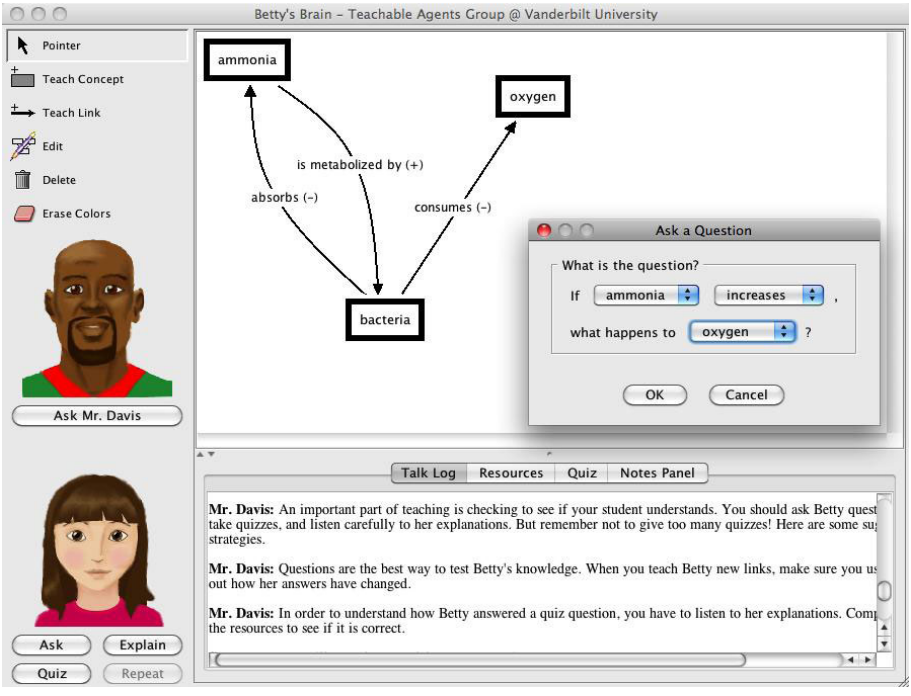

Fig. 1. The Betty's Brain user interface, featuring the Betty and Mr. Davis agents, student-created causal concept map, pop-up window used to ask Betty questions, and sample conversation with Mr. Davis

can ask Betty to explain the logic behind her answers, and she responds using speech, animation, and text. Betty expresses desire to improve her scores on quizzes and disappointment if this goal is not met.

The lessons were split into two units: arctic climate change and arctic food webs. Students in both conditions first took a pretest to establish their baseline knowledge of arctic climate change. After a brief introduction to the arctic climate, students in the experimental condition underwent one class period of training in the Betty's Brain program, while students in the control condition continued with normal lessons. Experimental students then spent four full class periods constructing their concept maps and teaching Betty. After completion, students took a post-test identical to the pre-test. Both groups of students then repeated the series of activities for the arctic food web lessons.

After completing the learning sessions and the second post-test, students were given the behavioral prediction questionnaire asking them to make predictions about a human, a computer, and Betty. Control participants with no previous exposure to Betty were given a brief description of the teachable agent before completing the questionnaire.

3 Results

Students' responses to the behavioral prediction scenarios were very similar to those given by adults [6]. They made similar proportions of intentional predictions for the anthropomorphic agent (Betty; 41% intentional responses) and the computer (43% intentional responses) and gave more intentional responses to the human agent (63%) than for either Betty ($t(72)=4.437$, $p<.001$), or the computer ($t(72)=3.687$, $p<.001$).

We also tested whether behavioral predictions were related to better performance on the learning task, as measured by the quality of the concept maps that the students created. Intentional behavioral predictions for humans and pretest scores were entered into a regression model predicting the average number of correct concepts and links in the concept maps that students created to teach Betty. The overall regression model was significant ($F(2,38)=8.59$, $p=.001$) and accounted for 27.5% of the variance in concept map quality. Both behavioral predictions and pretest scores contributed significantly to the model (standardized $\beta=.278$, $t(37)=2.05$, $p=.048$; and standardized $\beta=3.301$, $t(37)=.45$, $p=.001$, respectively). Thus, an advanced understanding of human agents was associated with increased learning from Betty's Brain.

Two similar regression models were created by substituting the human behavioral predictions for behavioral predictions for the computer and for Betty. While both models were significant ($F(2,38)=5.87$, $p=.006$ and $F(2,38)=5.86$, $p=.006$, respectively) neither the behavioral predictions for the computer (standardized $\beta=-.021$, $t(37)=-.145$, $p=.889$) nor for Betty (standardized $\beta=.003$, $t(37)=.023$, $p=.982$) contributed significantly to their respective models. While an understanding of human intentional thought is related to effective use of Betty's Brain, there appeared to be no link between beliefs about Betty's or the computer's mental representations and the ability to learn from the teachable agent.

Overall, students gave very similar predictions for Betty and the computer. However, there was a nonsignificant trend for students who interacted with Betty to give more dissimilar ratings of intentionality to humans and machines, implying that interacting with Betty highlighted the differences between human and mechanical agents. The person-machine difference (the intentionality assigned to the person minus the mean of the intentionality assigned to the computer and to Betty) was larger in the experimental condition (26%; $t(40)=4.008$, $p<.001$) than in the control condition (14%; $t(31)=2.103$, $p=.044$), although this contrast was not significant ($t(71)=-1.215$, $p=.228$).

4 Conclusion

We found that 7th grade students did not make more human-like intentional predictions for a teachable agent than they did for a computer. These findings mirror patterns of intentional attributions previously found in adults, but differ from adults who were asked to monitor the mental states (i.e., preferences) of a mechanical agent [6]. One possibility is that the human-agent interaction involved in using the Betty's Brain system did not sufficiently encourage students to actively monitor Betty's mental states. The students may not have considered her to be human-like (an assertion reinforced by the behavioral prediction results) but rather considered her to

be a simple tool for use in completing their lessons. Many students in this study may have either required more prompting to consider Betty's mental state or alternative tasks that would have encouraged them to consider Betty's thought processes more deeply.

An alternative explanation is that children and adults are differentially influenced by mental state monitoring when attributing intentionality to a non-human agent. The students may have sufficiently monitored Betty's mental states but did not go on to infer that Betty's behavioral motivations should be any different than a computer's. By this explanation, while children and adults may both be capable of considering a nonhuman agent to have human-like mental states, adults are more likely to generalize these beliefs to other tasks (i.e., questionnaire responses).

Additionally, we found that a general understanding of agency did predict success in learning domain content in the Betty's Brain system, perhaps because students with a more elaborate theory of mind were more likely and better able to monitor the agent's knowledge states. However, this relationship only holds when considering the behavioral predictions made for the human, not the behavioral predictions made for Betty or the computer. It is possible that only the human predictions are a good general measure of the kind of social skills, or intelligence, that theory of mind depends on. The students have had more practice attributing mental states to humans, and this skill may have made them more sensitive to the metacognitive activities promoted by the Betty agent, even if students did not directly attribute intentional representations to the agent.

In this study, we found not only that children's and adults' beliefs about agent intentionality are similar, but that these beliefs can be used to predict success when learning from an artificial teachable agent. This suggests that training about agency and even ideas about intentionality could help children use teachable agent systems. Of course, it is possible that the children who gave highly intentional predictions for humans did so because of a relatively broad social intelligence that would not be strongly modified by any specific limited-duration experience. On this view, training might not be particularly effective. Instead, it might be helpful to modify the teachable agents to provide more, or fewer, cues about intentionality, depending on children's basic social-cognitive skills. In either case, we believe that understanding basic concepts about agency will be an important part of the successful design and implementation of interactive agent-based learning environments.

Acknowledgments. This material is based upon work supported by the National Science Foundation under grants #0826701, #0904387, and #0633856.

References

1. Biswas, G., Leelawong, K., Schwartz, D., Vye, N.: Learning by Teaching: A New Agent Paradigm for Educational Software. Appl. Artif. Intell. 19(3), 363–392 (2005)
2. Gopnik, A., Wellman, H.M.: Why the Child's Theory of Mind Really is a Theory. Mind Lang. 7(1-2), 145–171 (1992)
3. Dennett, D.C.: Consciousness Explained. Little Brown & Co., Boston (1991)
4. Searle, J.: Minds, Brains and Science. Harvard University Press, Cambridge (1986)

5. Woodward, A.L.: Infants Selectively Encode the Goal Object of an Actor's Reach. Cognition 61(1), 1–34 (1998)
6. Levin, D.T., Saylor, M.M., Killingsworth, S.S., Gordon, S.M., Kawamura, K.: Tests of Concepts About Different Kinds of Minds: Predictions About the Behavior of Computers. Robots, and People (under Review)
7. Levin, D.T., Killingsworth, S.S., Saylor, M.M.: Concepts About the Capabilities of Computers and Robots: A Test of the Scope of Adults' Theory of Mind. In: HRI 2008 Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, pp. 57–64. ACM, New York (2008)
8. Nass, C., Moon, Y.: Machines and Mindlessness: Social Responses to Computers. J. Soc. Issues. 56, 81–103 (2000)
9. Kuhlmeier, V.A., Bloom, P., Wynn, K.: Do 5-month-old Infants See Humans as Material Objects? Cognition 94(1), 95–103 (2004)
10. Spelke, E.S., Phillips, A., Woodward, A.L.: Infants' Knowledge of Object Motion and Human Action. In: Sperber, D., Premack, D., Premack, A. (eds.) Causal Cognition, pp. 44–76. Clarendon Press, Oxford (1995)
11. Bloom, P.: Intentionality and word learning. Trends Cogn. Sci. 1(1), 9–12 (1997)