

# The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains

Daniel Leyzberg (daniel.leyzberg@yale.edu)

Samuel Spaulding (samuel.spaulding@yale.edu)

Mariya Toneva (mariya.toneva@yale.edu)

Brian Scassellati (scaz@cs.yale.edu)

Department of Computer Science, Yale University

51 Prospect St., New Haven, CT 06511, USA

## Abstract

We present the results of a 100 participant study on the role of a robot's physical presence in a robot tutoring task. Participants were asked to solve a set of puzzles while being provided occasional gameplay advice by a robot tutor. Each participant was assigned one of five conditions: (1) *no advice*, (2) *robot providing randomized advice*, (3) *voice of the robot providing personalized advice*, (4) *video representation of the robot providing personalized advice*, or (5) *physically-present robot providing personalized advice*. We assess the tutor's effectiveness by the time it takes participants to complete the puzzles. Participants in the *robot providing personalized advice* group solved most puzzles faster on average and improved their same-puzzle solving time significantly more than participants in any other group. Our study is the first to assess the effect of the physical presence of a robot in an automated tutoring interaction. We conclude that physical embodiment can produce measurable learning gains.

**Keywords:** Robotics; Computer Science; Tutoring

## Introduction

What kinds of human-robot interactions benefit from the physical embodiment of a robot? For human-robot interactions that require manipulating the physical world, a physical robot is a necessity, but for those interactions where physical embodiment is optional, when is an embodied robot more useful than an on-screen agent?

In this study, we explore the differences in task performance of participants engaged in a cognitive learning task in which a robot acts as a tutor. Participants were asked to play a puzzle game while receiving strategy advice from either: a physically-present robot, a video of the same robot, its disembodied voice, a robot giving randomized advice, or no agent at all. We use the resulting data to draw conclusions about the effect of embodiment in robot tutoring tasks.

Previous work has investigated the social influence of a robot's embodiment. Does a robot engender more trust, more compliance, more engagement, or more motivation by its physical presence, more so than an on-screen agent or a video representation of a robot would? Such questions have been explored via two methodologies: self-report measures and task-performance measures. Using self-report measures, Kidd and Breazeal (2004) found that a physically-present robot was perceived as more enjoyable, more credible, and more informative than an on-screen character in a block-moving task. In Wainer, Feil-Seifer, Shell, and Mataric (2007), an embodied robot was rated as more attentive and more helpful than both a video representation of the robot

and a simulated on-screen robot-like character. Tapus, Tapus, and Mataric (2009) found that individuals suffering from cognitive impairment and/or Alzheimer's disease reported being more engaged with a robot treatment than a similar on-screen agent treatment.

Kiesler, Powers, Fussell, and Torrey (2008) used task-performance measures to find that participants who received health advice from a physically-present robot were more likely to choose a healthy snack than participants who received the same information in robot-video or on-screen agent conditions. In Bainbridge, Hart, Kim, and Scassellati (2008), a physically-present robot yielded significantly more compliance to its commands than a video representation of the same robot.

No previous work has investigated whether learning outcomes are affected by a robot's physical presence. The closest related work is in Intelligent Tutoring Systems (ITSs), which are educational computer programs that produce individualized lessons, advice, and questions usually in a workbook-style or quiz-style environment (Nkambou, Bourdeau, & Psyché, 2010). A parallel notion of embodiment called "the persona effect" exists in ITS research. (See Dehn and Van Mulken (2000) for an overview.) The persona effect is the impact, if any, that an on-screen character has on students using an ITS. The majority of research on the persona effect has shown no significant learning gains produced by on-screen agents, although many studies note that students find an ITS with an on-screen more engaging than one without (Moundridou & Virvou, 2002).

Our study is the first to assess the effect of the physical presence of a robot in an automated tutoring interaction. We use the task-performance measure of puzzle solving time in this work as well as several self-report measures.

## Methodology

### Participants

There were 100 participants in this study, between 18 and 40 years of age. The study was conducted in New Haven, Connecticut. Most participants were undergraduate and graduate students of Yale University. Each participant was assigned to one of five groups: (1) *no lessons*, (2) *randomized lessons from a physically-present robot*, (3) *personalized lessons from a disembodied voice*, (4) *personalized lessons from a video representation of the robot*, and (5) *personalized lessons from*

1	1	2	4	3	1
2	1	1	1	1	4
1	3	3	2	3	2
1	1	3			
1	1	2			
1	1	2			
2	2	1			
1	1				
2	1	2			
7					
6	1				

(a) Sample nonogram puzzle, blank.

1	1	2	4	3	1
2	1	1	1	1	4
1	3	3	2	3	2
1	1	3			
1	1	2			
1	1	2			
2	2	1			
1	1				
2	1	2			
7					
6	1				

(b) Sample nonogram puzzle, solved.

Figure 1: A sample nonogram puzzle. The objective of nonograms is, starting with a blank board (see left figure), to find a pattern of shaded boxes on the board such that the number of consecutively shaded boxes in each row and column appear as specified, in length and order, by the numbers that are printed to the left of each row and above each column (see right figure). For a more detailed explanation see the **Domain** section.

a physically-present robot. There were approximately 20 participants in each group. Exclusion criteria for participants were lack of English fluency or prior academic experience with robotics or artificial intelligence.

## Apparatus

In this experiment, participants were asked to solve a series of logic puzzles. In the four experimental conditions with a tutor, the tutor interrupted participants several times per puzzle to deliver puzzle-solving strategy lessons. The lessons themselves were pre-recorded audio and synchronized visual aids, between 21 and 47 seconds in length, that explained and gave examples of the use of a single puzzle-solving strategy. In the experimental conditions with *personalized lessons*, the order of the lessons was determined by a skill assessment algorithm that identified skills in which participants were weak; see the **Skills & Lessons** section. In the *randomized lessons* condition, the tutor chose a random lesson among the same ones used in the *personalized lessons* conditions, such that it was immediately applicable to the current state of the game-board. We compare the puzzle solving time performance between participants in these groups to evaluate the effect of the robot’s physical presence on the effectiveness of the tutoring.

**Domain** To minimize the influence of prior experience, we chose a test domain to which participants likely had little previous exposure: a grid-based fill-in-the-blanks puzzle game called “nonograms” (or “nonogram puzzles”) that resemble crossword puzzles or Sudoku. Nonogram puzzles are a difficult cognitive task, one that requires several layers of logical inferences to complete. Solving a nonogram puzzle of arbitrary size is an NP-complete problem (Nagao, Ueda, Ueda, Sato, & Watanabe, 1996), meaning that no efficient computational solution is known.

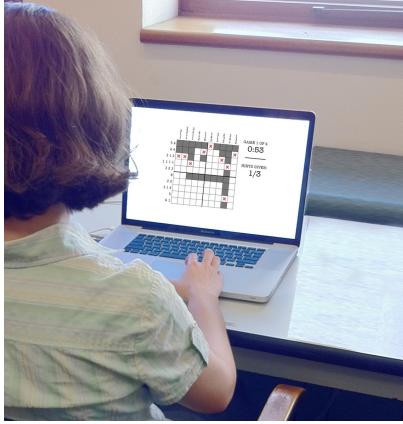
The objective of nonograms is, starting with a blank board, to shade in boxes on the board such that the number of consecutively shaded boxes in each row and column appear as

specified, in length and order, by the numbers that are printed to the left of each row and above each column. (See Figures 1(a) and 1(b) for a sample puzzle and solution.) For instance, a row marked as “4 2” must have 4 adjacent shaded boxes, followed by 2 adjacent shaded boxes—in that order, with no other boxes shaded, and with at least one empty box between the sets of adjacent shaded boxes. We refer to these contiguous sets of shaded boxes as “stretches” in this paper. For instance, the row described above requires two stretches, one of length 4, the other of length 2. One solves the puzzle when one finds a pattern of blank and shaded boxes such that all of the requirements for each row and column are satisfied.

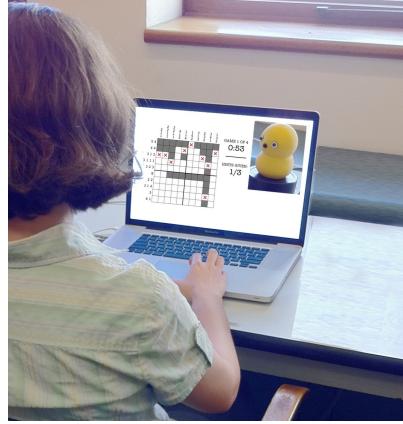
In a typical puzzle, one cannot solve many rows or columns independently. One must infer the contents of parts of rows or columns and use previous inferences as the basis of subsequent inferences. To that end, when a player has reasoned that in some box or boxes there should not be shading, they can mark such boxes with an ‘X’ for reference.

We created a full-screen nonograms computer program that participants used via mouse and keyboard. The user interface provided a timer and a count of how many lessons (called “hints” in the interface) the participant had received and how many they would receive; see Figure 2.

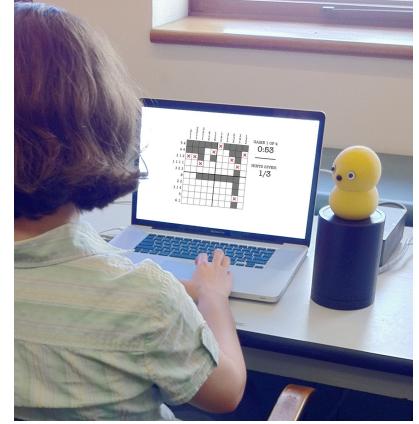
Participants were asked to play four puzzles on ten-by-ten grids with a time limit of fifteen minutes per puzzle. The puzzles themselves were the same across all participants. The fourth puzzle used the same board as the first, although disguised in the fourth puzzle by rotating the board 90° (such that the column stretch requirements were swapped with row stretch requirements). This means that the first puzzle and the last puzzle are of the exact same difficulty and require knowledge of the exact same set of skills to solve. This manipulation enables us to make within-subjects comparisons about the extent to which each participant improved their skills over the course of their participation in the study. There was no indication that any participant was aware of this manipulation.



(a) Experiment apparatus in the *no lessons* condition and the *personalized lessons from a disembodied voice* condition.



(b) Experiment apparatus in the *personalized lessons from a video representation of the robot* condition.



(c) Experiment apparatus in the *randomized lessons from a physically-present robot* condition and the *personalized lessons from an embodied robot* condition.

Figure 2: Experiment apparatus by condition.

**Skills & Lessons** In the four conditions with lessons, the tutor interrupted the participant three times per puzzle, paused the puzzle, and delivered a short lesson about nonograms. The lessons ranged from 21 seconds to 47 seconds in length and consisted of a voice recording and a set of animations presented on screen during the lesson as well as a set of coordinated robot motions specific to each lesson.

When beginning a lesson the tutor would turn to face the participant (in the *video* and *physically-present robot* conditions) and say “I have an idea that might help you,” or “Here’s another hint for you.” During the lesson, the tutor bounced subtly and looked back at the screen whenever, in the course of the lesson, it would make reference to the example presented on screen. For instance, when in the audio of the lesson the robot would say “Like in this example...” or “As you see here...,” the robot would turn briefly to the screen and then back to the participant.

Ten nonogram puzzle-solving skills were identified based on the subjective experience of the authors; they are not universally identified skills or rules for nonograms. Each skill is a set of row or column states in which one can logically fill in some of the remaining empty boxes. For example, a stretch of length 9 can fit in a blank row or column of 10 boxes in only two ways. Either it fills the first box and 8 more, or it fills those same middle 8 boxes and the last box. In either case, the middle 8 boxes are shaded. One of the ten skills in this experiment is that, for an empty row or column with just one stretch requirement of  $n$  where  $n > 5$ , the middle  $(2n - 10)$  boxes are shaded. See Figure 3 for examples and explanations of this skill and two others.

There was one recorded lesson for each skill. Three lessons were delivered per puzzle, for each of four puzzles. The number of lessons was constant for all participants regardless of how long they needed to finish the puzzle. Lessons were triggered either when a participant made no moves for 45 seconds or as he or she filled the 25<sup>th</sup>, 50<sup>th</sup> or 75<sup>th</sup> box on the board

(of 100). The user interface displayed the number of lessons remaining for each puzzle at all times.

In the *personalized lesson conditions* the lessons were chosen based on a skill assessment algorithm. For each skill, a weighted sum was calculated internally consisting of: (1) the number of recent demonstrations of that skill (weighted positively) and (2) the number of recent gameboard states in which a skill could have been applied but no action was taken (weighted negatively). These assessments were updated for each skill separately throughout the game, and the skill with the lowest assessment that was applicable to the current gameboard was the skill for which a lesson was selected. In this way, participants in the *personalized lesson* conditions received lessons based on their individual performance on the puzzles.

Alternatively, in the *randomized lesson condition*, lessons were chosen among the same ten lessons at random each time, such that the lesson chosen could be applied to the current state of the gameboard. This ensures that although the lessons were randomized, they would provide actionable information every time.

**Robot** The robot we used, Keepon, is a small yellow snowman-shaped robot; see Figure 2(c). Keepon has previously been used as an emotive non-threatening communication tool (Kozima, Nakagawa, & Yasuda, 2005; Leyzberg, Avrunin, Liu, & Scassellati, 2011).

The robot operated in one of three modes. First, it refereed the puzzle game: it welcomed participants when they started, told them when they had finished or when they had run out of time, and told them when the experiment was over. Second, it “observed” the board during gameplay: the robot frequently turned its head to face the location of the mouse cursor. Third, it delivered short gameplay lessons three times per puzzle: it “spoke” to the participant by turning to face him or her and “bouncing” its body subtly while playing one of several pre-recorded spoken messages. If a lesson needed to be repeated,

the robot would first apologize for repeating itself (i.e., “I’m sorry to repeat this hint but I think it might help.”).

To simplify the potential perception problems inherent in real-world measurements, the robot in this study received perfect knowledge of the state of the game. We did not use a robotic vision system to detect state changes.

## Procedure

Participants were first asked to watch a five-minute instructional video and read a two-page instruction manual describing the rules of nonograms and how to use the computer interface. In the video and in the text, participants were encouraged to use logical reasoning to make moves in the puzzle rather than making moves by guessing. Potential questions about the rules of the puzzle game were answered by the experimenter after the instructions.

During the experiment, participants were alone in a room with the computer, the robot in conditions including the robot, and a video camera positioned behind them; see Figure 2. Participants would choose when they were ready to start each new puzzle; each round would end either when the participant solved the puzzle or when fifteen minutes had elapsed, whichever happened first.

After the conclusion of the final puzzle, participants were asked to complete a survey consisting of five Likert-scale questions with open-ended follow-up questions for each. The questions were designed to assess whether the lessons were helpful, clear, and influential, as well as the user’s perceptions of the robot. We asked participants to rate: how relevant the lessons were, how much the lessons influenced their gameplay, how well participants understood the lessons, and how “smart/intelligent” and “distracting/annoying” they perceived the robot to be.

## Results

This study investigates the role of physical embodiment in a robot tutoring system. The behavioral measure is the time in which participants were able to solve each of the four puzzles. For the purposes of calculating a mean, puzzles in which participants ran out of time were evaluated as having solved the puzzle when time ran out, fifteen minutes from the start of each puzzle. This occurred in 12.4% of all puzzles.

Table 1: Mean Solving Time

	Puzzle 1	Puzzle 2	Puzzle 3	Puzzle 4
<i>None</i>	$13.6 \pm 2.2$	$13.0 \pm 2.3$	$12.3 \pm 2.5$	$11.6 \pm 2.7$
<i>Rand.</i>	$13.8 \pm 1.4$	$12.5 \pm 2.0$	$11.4 \pm 2.3$	$10.3 \pm 2.9$
<i>Voice</i>	$12.6 \pm 2.4$	$10.7 \pm 2.7$	$10.3 \pm 3.3$	$9.1 \pm 3.0$
<i>Video</i>	$12.8 \pm 2.1$	$11.1 \pm 2.6$	$9.9 \pm 2.6$	$8.7 \pm 2.4$
<i>Robot</i>	$12.7 \pm 2.6$	$10.0 \pm 3.5$	$9.4 \pm 3.0$	$7.6 \pm 3.1$

Participants in the *robot group* performed better, on average, on the second, third, and four puzzles than participants in any other group. See Table 1 for means and standard deviations and see Figure 4(a). In the forth puzzle, the



(a) In this row, there must be one long stretch. By the process of elimination one can infer that this stretch must occupy at least the middle six boxes, no matter where in the row it is placed.



(b) In this row, the first box is already shaded. Given that, and that the first stretch must be 3 boxes long, one can infer that the first three boxes must be shaded and the fourth must be crossed out.



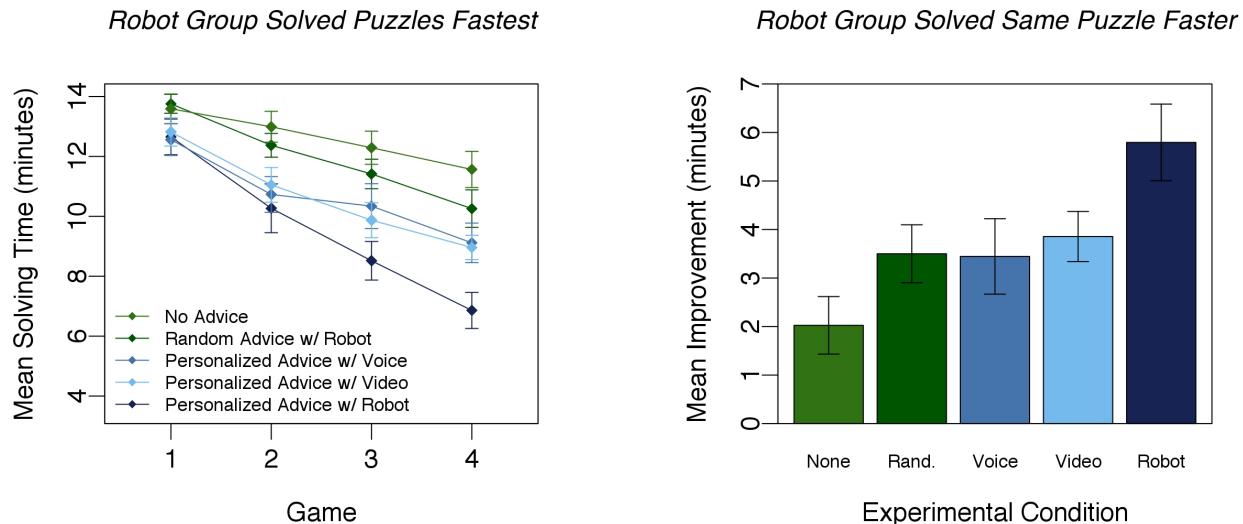
(c) In this row, there is only one short stretch and some boxes are already shaded. One can infer that regardless of where that one stretch is placed, it cannot occupy the first two or the last two boxes in that row.

Figure 3: Examples of nonograms skills. Displayed are the contents of a row before and after each skill is applied. Although only rows are shown here, all nonograms skills apply to columns as well.

mean solving time in the *robot group* ( $M = 7.6$  minutes,  $SD = 3.1$ ) is significantly better than in the *video group* ( $M = 8.7$  minutes,  $SD = 2.4$ ),  $t(36) = 0.03$ , and in the *voice group* ( $M = 9.1$  minutes,  $SD = 3.0$ ) as well,  $t(36) = 0.02$ . These data indicate that the robot’s physical presence made a significant learning impact on participants greater than that of an disembodied voice and a video representation of a robot.

In this experiment, the first and fourth puzzles were  $90^\circ$  rotated variations of the same board. Thus they required exactly the same skills to solve and the difference in their solving time is a measure of the participants’ acquired knowledge over the course of the study. Participants in the *robot condition* improved ( $M = 5.8$  minutes,  $SD = 3.5$ ) their same-puzzle solving time significantly more than those in both the *video condition* ( $M = 3.9$  minutes,  $SD = 2.3$ ),  $t(36) = 0.048$  and *voice condition*, ( $M = 3.4$  minutes,  $SD = 3.5$ ),  $t(36) = 0.04$ ; see Figure 4(b). This data indicates that participants who received lessons from the robot learned more effectively than those who received only voice- or video-based lessons.

Survey results verify the following manipulation: participants in the three *personalized advice conditions* rated the lessons significantly more relevant ( $M = 6.0$ ,  $SD = 1.4$ )



(a) Mean solving time per puzzle. Participants in the *robot condition* solved each puzzle faster than participants in any other condition. In the fourth puzzle, significantly faster ( $p \leq 0.03$ ). See Table 1 for means and standard deviations.

(b) Mean improvement in solving time between puzzles #1 and #4. These two puzzles were variations of the same gameboard, disguised in the fourth puzzle by a 90° rotation. Participants in the *robot condition* improved their solving time significantly more than those in any other condition ( $p < 0.05$ ).

Figure 4: Behavioral measure results: (a) participants who received personalized lessons from an embodied robot solved every puzzle faster on average, the fourth significantly so ( $p \leq 0.03$ ) than participants in all other conditions; see Table 1. (b) *robot condition* participants also improved on their same puzzle solving time significantly higher more than participants in all other conditions ( $p < 0.05$ ).

than participants in the *randomized advice condition* ( $M = 3.9, SD = 1.1$ ),  $t(33) < 0.001$ . There was no significant difference in how highly participants rated their understanding of the lessons between groups: ( $M = 6.0, SD = 1.4$ ) in the *random condition*, ( $M = 6.6, SD = 1.2$ ) in the *voice condition*, ( $M = 6.6, SD = 1.5$ ) in the *video condition*, and ( $M = 6.4, SD = 1.2$ ) *robot condition*; see Figure 5. These data indicate that whatever social effect physical embodiment has on this interaction, it does not influence the participants' perception of their understanding of the lessons, despite the fact that the behavioral measure indicates better learning in the *robot condition*.

## Discussion

Our results indicate that a physically-present robot tutor produces better learning gains than on-screen or voice-only tutors. Further work is needed to identify the underlying social factors and mechanisms that cause this effect.

One such factor may be the novelty of the stimulus. Robots are an uncommon stimulus in the present day; we may expect participants to be more attentive to the agent in the physically-present condition. However, novelty can also be a distraction. The physical presence of the robot during the game may divert the participant's attention from the puzzle solving task. More work is needed to identify what effect, if any, a robot's novelty has on interactions such as these.

Physical presence may imbue the robot with more per-

ceived authority than an on-screen agent. Earlier work in this area indicates that people are more likely to comply with commands given by a physically-present robot than an on-screen video of the same robot (Bainbridge et al., 2008). Embodiment may cause participants to take the robot tutor's advice more seriously. We are accustomed to receiving lessons from teachers and authority figures who have physical bodies. Perhaps a robot's physical presence increases its authority or social standing.

Participants, however, did not report having significantly more difficulty understanding the lessons in any of the three advice conditions. In fact, all four groups rated their level of understanding of the lessons fairly highly ( $M = 6.3, SD = 1.3$ ); see Figure 5(b). This may indicate that the embodiment effect is so subtle that the participants did not notice its effect on their learning.

Another social factor is the potentially increased sense of peer pressure during the performance of the task itself. The distinction between physically-present robot and on-screen agent may parallel the way we perform tasks when we think of ourselves as alone rather than in view of another person. In person-person interactions, social presence can lead to significantly worsened task performance, especially in cognitively-demanding tasks (Short, 1976). More work is needed to compare the potential effect of peer pressure caused by a physically-present robot tutor to the peer pressure exerted by a human tutor who observes as participants perform

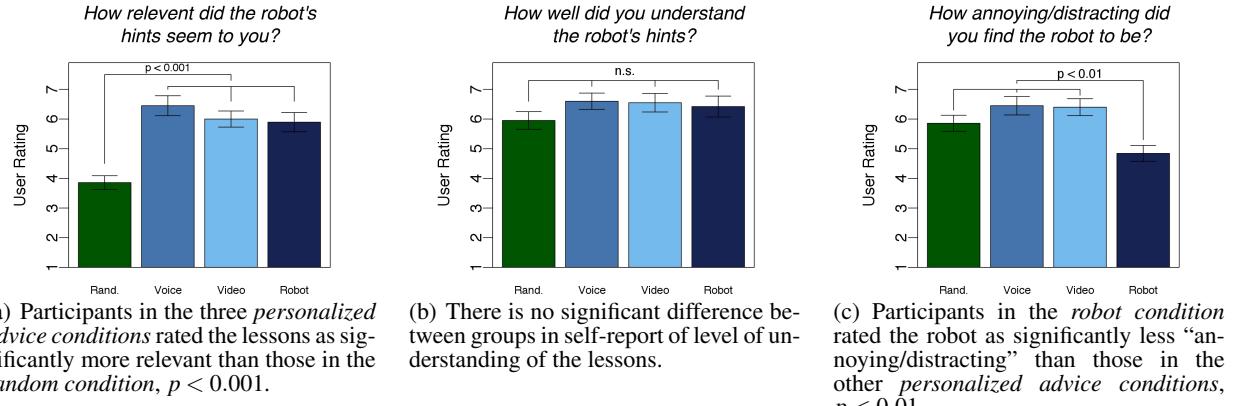


Figure 5: Results of self-report questionnaire measures completed after the interaction.

cognitively-demanding tasks.

Social presence effects may also be responsible for the survey result in which participants in the *physically-present robot condition* rated the robot ( $M = 4.7, SD = 1.8$ ) as significantly less “annoying/distracting” than participants in the *other advice conditions* ( $M = 6.1, SD = 1.3$ ),  $t(33) < 0.01$ ; see Figure 5(c). This may indicate that physical embodiment produces a significantly greater sense of social acceptance than an on-screen agent does.

Participants in the *robot condition* became better puzzle solvers than those in the other conditions. Further research is needed to identify the underlying social factors that contribute to this empirically-observed effect.

## Conclusion

This study investigates the role of physical embodiment of a robot tutor in a cognitive skill learning task. Participants who received personalized lessons from a physically-present robot outperformed participants who received the same kind of advice from a video representation of the same robot as well as participants who received the same kind of advice from a disembodied voice on the last three puzzles. Participants in the *robot condition* also improved their same-puzzle solving time significantly more than those in any other group, which is a direct measure of learning gains over the course of the experiment. From these data we conclude that physical embodiment can yield measurable learning gains in robot tutor interactions.

## Acknowledgments

This material is based upon work supported by grants from the National Science Foundation under contracts No. 1139078, No. 1117801, and No. 0835767.

## References

Bainbridge, W., Hart, J., Kim, E., & Scassellati, B. (2008). The effect of presence on human-robot interaction. *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 701–706.

Dehn, D., & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies, 52*(1), 1–22.

Kidd, C., & Breazeal, C. (2004). Effect of a robot on user perceptions. *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, 4, 3559–3564.

Kiesler, S., Powers, A., Fussell, S., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition, 26*(2), 169–181.

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication-care: a case-study in autism therapy. *IEEE International Symposium on Robot and Human Interactive Communication, 341 - 346*.

Leyzberg, D., Avrunin, E., Liu, J., & Scassellati, B. (2011). Robots that express emotion elicit better human teaching. *6th International Conference on Human-Robot Interaction, 347–354*.

Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning, 18*(3), 253–261.

Nagao, T., Ueda, N., Ueda, N., Sato, C. P. T., & Watanabe, C. P. O. (1996). *Np-completeness results for nonogram via parsimonious reductions* (Tech. Rep.). Tokyo Institute of Technology.

Nkambou, R., Bourdeau, J., & Psyché, V. (2010). Building intelligent tutoring systems: An overview. *Advances in Intelligent Tutoring Systems, 361–375*.

Short, W. E. . C. B., J. (1976). *The social psychology of telecommunications*. London, England.

Tapus, A., Tapus, C., & Mataric, M. (2009). The role of physical embodiment of a therapist robot for individuals with cognitive impairments. *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 103–107.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. *IEEE Proceedings of the International Workshop on Robot and Human Interactive Communication, 872-877*.