The role of physical embodiment in human-robot interaction

Joshua Wainer   David J. Feil-Seifer   Dylan A. Shell   Maja J. Matarić

Interaction Laboratory
Center for Robotics and Embedded Systems
Department of Computer Science, University of Southern California
Los Angeles, CA USA 90089-0781
jwainer | dfseifer | shell | mataric @usc.edu

Abstract— **Autonomous robots are agents with physical bodies that share our environment. In this work, we test the hypothesis that** *physical embodiment* **has a measurable effect on performance and perception of social interactions. Support of this hypothesis would suggest fundamental differences between virtual agents and robots from a social standpoint and have significant implications for human-robot interaction. We measure task performance and perception of a robot's social abilities in a structured but open-ended task based on the Towers of Hanoi puzzle. Our experiment compares aspects of embodiment by evaluating: (1) the difference between a physical robot and a simulated one; (2) the effect of physical presence through a co-located robot versus a remote tele-present robot. We present data from a pilot study with 12 subjects showing interesting differences in perception of remote physical robot's and simulated agent's attention to the task, and task enjoyment.**

Index Terms— **Human-robot interaction, embodiment**

I. Introduction

A growing human-robot interaction (HRI) community is focusing on the social aspects of autonomous robots. This research is critical if robots are to become part of people's everyday lives. Research in HRI is challenging because, in addition to engineering and technological hurdles, the many factors that interact to form a rich social experience must be teased apart for proper study. One factor with possible implications for social interaction is that a robot has a *physical body*. A robot's experience of the physical world and social situation, its perception, computing, and actions are thus all attributable to a tangible artifact. Moreover, such attribution (be it implicit, explicit or both) may have an important role in natural social interaction. This paper addresses the question: "yes a robot's physical embodiment have a measurable affect on its social interactions?"

A greater understanding of the social implications of embodiment will inform design of social robots. In robot systems designed with social interaction as their primary goal (e.g., entertainment robots [1] and socially assistive robots [4]) this study addresses the question of how robots differ, and could complement, alternative technologies like virtual sociable agents [14], or smart spaces [7]. A broad understanding of embodiment is necessary to provide a solid footing for HRI and to contextualize its relationship with other human-factors issues.

HRI seeks constructive constraints that can serve as guidelines to assist in making design decisions. A roboticist constructing a system for use around humans currently has few established design principles. Yet, such principles are particularly critical because of to the number of factors that can influence an HRI system. Insight into the nature of human social interaction, particularly evidence showing fundamental differences between embodied and disembodied communication, would be valuable.

This paper presents pilot data from an empirical study of the role of physical embodiment in a task involving a social robot. In particular, we attempt to distinguish between the effect of physical presence and virtual versus material embodiment. Trials with human participants permit comparison of task performance and perceived social awareness (by the robot) in three conditions: (1) a co-located physical robot, (2) a remotely located (tele-present) robot, and (3) a simulated robot. In each case, the robot engages the participants through a simple Towers of Hanoi puzzle. The robot supervises the task, including explaining the puzzle, setting intermediate goals, directing the person's behavior, and enforcing the rules of the game.

Our results support the belief that physical embodiment is an important factor in social tasks and that it has an effect on the perception of robot's social situatedness. In particular, presence of a physical robot is most enjoyable and believed to be more watchful than either a virtual agent, or even a physical one separated via a video conferencing setup.

II. Related work

The role of embodiment within social robot interactions has not yet been the subject of direct investigation. A number of reasons exist for this, but perhaps most significant is that the relationship between embodiment and situatedness has

not been explicitly articulated within a social context. For example, Fong et al. [5] define *social situatedness* for robots in a manner that could apply to virtual agents. Situatedness (social or otherwise) has been thought of as applicable to virtual agents within a broader AI view [11]. On the otherhand, Pfeifer and Scheier [13] argued that: *"Intelligence cannot merely exist in the form of an abstract algorithm but requires a physical instantiation, a body."* Several authors view the a virtual avatar as adequate embodiment, while others have asserted that a robot may not be adequately embodied; see Ziemke [17] for this debate. We use the term *physically embodied* to refer exclusively to a robot's body, and this study seeks to understand the key impact such embodiment and physical presence have on social interactions.

Frequently, an HRI component is added to an existing system. Furthermore, novelty alone is usually sufficient to justify robots for entertainment purposes, at least initially. More serious and sustained needs (e.g., rehabilitation, elder care. nursing, educating) require a better understanding of the factors involved. Kiesler and Goetz [9] examined the nature of robot personality (terse versus jovial) as one such factor. They evaluated aspects of robot form, like height comparable to a person versus a pet, and discussed human mental models of the robot. Their findings were that, while the appearances were different, the dialog directed toward the robot was the same in each case.

The closest existing work to the present study was carried out by Kidd and Breazeal [8] in which three characters: a human, a robot, and an animated character, each verbally instructed participants in a block stacking exercise. Unlike that work, however, our experiments consider scenarios in which the robot is not obscured (only the character's eyes were visible in Kidd and Breazeal [8]). Our work considers a very similar task, but, as famously shown by Zhang and Norman [16], small changes to the Towers of Hanoi task, even isomorphic variations — i.e., changes preserving task structure — can have a marked effect on task performance. Woods et al. [15] also studied perception differences between live and video recorded robot performances. They proposed using video during system development as a complementary research tool, but this becomes less straightforward when considering experiments, like those presented in this paper, in which the robot's behavior depends on the human subject's responses.

Finally, those in a position to investigate aspects of embodiment may have vested interests. Roboticists might favor a study with an outcome that demonstrates the significant positive effects of physical embodiment. (This paper's authors are robotics researchers, but are aware and vigilant against any predisposition.) While we qualify the term embodiment within this paper with "physical", many roboticists may feel that embodiment refers only to physical
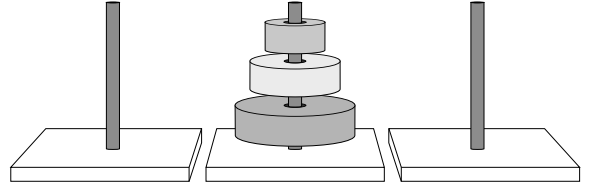

Fig. 1.   The classic Towers of Hanoi puzzle with three rings (slightly exploded for clarity) and three pegs as used in the experiments. In our setup each of the three rings has a different color and mass. Three weight sensors under each peg and a single camera allow the robot to estimate the state of the puzzle.

bodies and that rendering a graphical body for a virtual agent is not equivalent. Our view is teleological: if physical embodiment has unique properties, e.g., producing patient compliance through mutual empathy, then those aspects should be exploited.

III. FORMAL HYPOTHESES

Robots offer uniquely controllable experimental conditions which allow social characteristics that have, until now, not been accessible for controlled study. Social robots also raise new questions. We believe this to be the case in dealing with embodiment. To tease apart the difference between realism and physical situatedness, we formulated the following hypotheses:

H1.A   A co-located physical robot will result in a perception of higher social awareness than a remote physical robot.

H1.B   A co-located physical robot will elicit longer interactions than a remote physical robot in an openended interaction domain.

H2.A   A co-located physical robot will result in a perception of higher social awareness than a simulated (virtual) robot.

H2.B   A co-located physical robot will elicit longer interactions than a simulated (virtual) robot in an openended interaction domain.

H3.A   A remote physical robot will result in a perception of higher social awareness than a simulated (virtual) robot.

H3.B   A remote physical robot will elicit longer interactions than a simulated (virtual) robot.

The term "co-location" refers to the robot being physically located in the same place as a human interacting with it. By "remote physical robot" we mean to a physical robot in a different physical location but which has sensing relayed to it (over a wireless network) and its actions relayed back to the human through a video conferencing system. The simulated robot replaces the physical robot altogether, rendering the
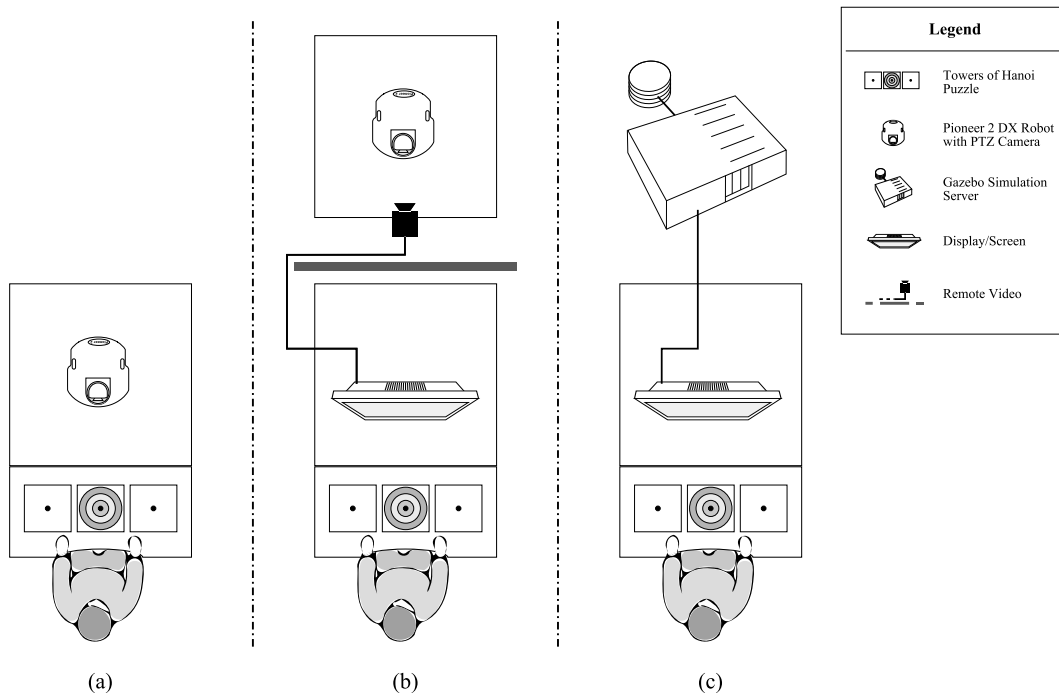
Fig. 2. The three experimental conditions: (a) the participant interacts directly with a physical robot; (b) the participant interacts with a physical robot over a real-time video-conferencing link; (c) the participant interacts with a simulated robot.

robot to a computer monitor and playing the audio feedback directly for the human.

IV. METHOD

Three cases we considered: interaction with a physical robot, interaction with a remote physical robot (through tele-conferencing), and interaction with a virtual (simulated) robot. We measured the user's task-oriented performance relative to the physical presence of the robot. In order to test task-oriented performance, we believe that some minimal task complexity is necessary for the difference between embodiment conditions to be measurable. We thus used a task that, while simple, provides a shared context for the robot and human participant. The task is also relevant to our broader assistive robotics agenda since the physical effort required to move the rings between pegs is useful in a post-stroke rehabilitation setting. The experiments employ a simple robot, as prior experience (see Eriksson et al. [3]) has shown that even simple robots can be engaging.

A. Towers of Hanoi Problem Domain

We designed a task around the classical Towers of Hanoi puzzle [16, pp. 92], in which different sized rings are individually moved from one peg to another (see Figure 1). The three pegs form the focus of the interaction: the robot (remote and co-present as well as the simulated versions) introduces the game and the rules that apply to it. The robot

(details in Section IV-C) provides the user with a particular stacking goal; for example, "Move the rings to the middle peg." The robot can perceive the puzzle state allowing for supervision of the ring movement; feedback is provided based on an estimate of task progress.

Since three rings provide relatively few states, some users are expected to explore the limits of the robot by breaking the rules of the game, or doing nothing to see how the robot will react. The system is sufficiently robust to catch errors and will explain how to put the puzzle back into the last legal state.

B. Experimental Design

The pegs, rings, and robot (when applicable) are placed on tables with height of approximately 1.2m. The human participant faces the robot (placed on an adjoining table) with only the three pegs between them. Figure 2 shows these three experimental conditions diagrammatically. The same set of rings are used in all conditions;this consistency is important since variations, even up to an isomorphism, can affect the performance at the puzzle [16].

(a) This is the co-located physical robot condition, as a typical model of human-robot interaction. The robot is placed on the table in front of the user.

(b) This is the remote physical robot case, in which the audio-video tele-conferencing system provides real-

time playback of the robot, despite being situated in a different room.

(c) This is the virtual robot case. The same screen and audio setup as (b) is used, but without a physical robot; instead output from a simulated robot is used.

Each participant performs the three conditions in a randomly assigned order. A questionnaire is filled out after each condition and the questions are presented before running the condition. Questions asked after each condition are identical and are given in the Appendix. After completion of all three conditions, the participants answer a final questionnaire with comparative rankings, again listed in the Appendix.

The robot provides feedback to the participant through physical movement: the robot moves toward and away from the user, a Pan-Tilt-Zoom (PTZ) camera nods when appropriate, as per previous studies with head gestures [12]. Audio feedback is a pre-recorded female voice, played back by generating segments of sentence-length. In conditions (b) and (c), the audio is played through speakers beside the monitor. In condition (a), the speakers are on the physical robot.

Before each condition, the subject is instructed to follow the robot's instructions for the Tower of Hanoi task, and to press a button when desiring to stop. We recorded the amount of time (in seconds) between start of the task and the indication to stop the task. We assume that this corresponds to how long the robot can encourage the subject to remain on the task.

C. Implementation details

Player [6] provides an abstraction layer for programming the robot we used. The robot is a Pioneer 2DX from ActivMedia with a Sony Pan-Tilt-Zoom camera that is used for the ACTS blob-finder (and simulating head-gestures) as well as speakers for audio output (see Figure 3). As an abstraction layer, the software allowed the same code to be executed in all conditions, helping to ensure comparable autonomous behavior in each condition. The simulated robot is rendered using Gazebo [10] (see Figure 4). This simulator contains an approximate physical model of the Pioneer and the PTZ camera used in the physically embodied parts of the experiment. The simulation is also controlled through player using the same control code as for the physical robot situations. The result is a simulation that behaves as exactly as can be approximated to a real-world robot.

The pegs used for the Tower of Hanoi task are equipped with weight sensors. Though not sensitive enough to determine exactly which weights are on the pegs, the sensors are able to detect when a ring has been removed from a peg or placed on a peg. The rings used for the Tower of Hanoi task are covered in brightly-colored paper. When the weight senors have detected a change in state for the pegs, the ACTS color blob tracker is used to determine which rings are

Fig. 3. The ActivMedia Pioneer 2 DX robot used in the experiments, as seen in the remote robot setting. The motions of the PTZ camera provide feedback that supplements the audio.

Fig. 4. The standard Gazebo simulation of the above robot.

placed on which pegs. The result is an accurate observation of the state by the camera-blobfinder combination. Player software allowed virtual sensing to be done transparently for conditions (b) and (c) through the use of a driver called *passthrough* that relays sensor readings across the network.

V. RESULTS

We tested a series of human participants as described above ($n = 11$). The gender spread was 9M-2F. Most subjects were experienced with computers and some were experienced with robots. While the sample size is fairly small and more uniform than we would like, it was adequate to evaluate the experimental design and determine areas for improvement.

We examined the survey results in order to determine if the embodiment of the moderator has any effect on the subject's reported perception of it as a guide for the task. For the analysis, we used a pairwise t-test on the comparative rankings used in Form 2 (see Figure 7).

Comparison ($n = 11$)	Significance
Co-located more watchful than remote-located	$p < 0.001$
Co-located more watchful than simulation	$p < 0.001$
Co-located more enjoyable than remote-located	$p = 0.012$
Co-located more enjoyable than simulation	$p < 0.001$

Fig. 5. Significance for survey data

While most of the survey data were discarded as not relating to the hypotheses, we were able to use the post-hoc questions: "(The moderator) I enjoyed the most " and "(The moderator that) Watched me the most closely" to address hypotheses H1.A, H2.A, and H3.A. As shown in Figure V, the co-located physical robot was seen as both more watchful and more enjoyable than either the simulation of a robot or the remote robot.

We also collected data on how much time a user spent on the task for each moderator. The difference between the time spent on each moderator did not vary significantly. Thus no conclusions regarding H1.B, H2.B and H3.B could be drawn.

Since none of the recorded times for the participants indicate a significant difference of result, we conclude that the embodiment of the moderator does not affect the time spent on the task we tested.

VI. DISCUSSION

Mean time spent on each of the conditions did decrease, as might be expected due to novelty for the first trial and a satiation effect thereafter. The open-endedness of the task contributed to a particularly wide variance in task times: several participants' minimum time (across conditions) exceeded other participants' maximum times. Time spent with the puzzle is not an ideal measurement of social interaction. A more structured task would make measurement of time spent on task more meaningful, as task performance is particularly critical for domains in which social cues must be used to steer interactions toward particular outcomes (e.g., in socially assistive robots [4]). This highlights the observations of Kiesler and Goetz [9] showing that a less friendly (or as we interpret in our results, fun) robot may result in lower task performance.

The effect of novelty is difficult to avoid. If three different tasks with different scenarios for interaction are used for each condition, then the door is opened to questions about task comparability. An alternative is to perform a suitable number of pre-experiment tasks and wait for the novelty to wear off. The idea of capturing only rote actions seems to measure a factor of HRI that may show the least difference between the three conditions. Our future work will try to obtain statistical significance for task-related measures by considering only participant's first trials. This will require a large number of participants.

Our experience with the experimental implementation has suggested several additions and improvements. We believe that the task interaction can be improved by grounding interactions more directly with the puzzle. For example, the PTZ camera gestures should make better reference to the state rings on each peg. Greater turn-taking and interleaved interactions would make the interactions more natural for all three conditions. As per the previous discussion, a better metric for task performance would quantify intermediate steps (looking at the state of the rings) within the puzzle state.

VII. CONCLUSION

Out results suggest that our current experimental setup is inadequate for fully addressing questions that deal with embodiment for the purposes of the task, not for social interaction itself. We have suggested some improvements for such direction of inquiry to be tackled.

Our initial data suggest that physically embodied interactions are favored over virtual ones and remote teleconference ones. Dautenhahn et al. [2] discussed the nature of embodiment, but left the question of "embodiment without a physical ontology" as future work, thus leaving open the question of what makes "material embodiment" special. We conclude from our results that physical or "material" embodiment in a task-oriented setting can make a difference in perception of a social agent's capabilities and the user's enjoyment of a task.

VIII. FUTURE WORK

We will continue to explore the nature of embodiment as it relates to HRI. As part of an ongoing research program into socially assistive robotics, we intend to compare a physical robot to a non-embodied agent for the purposes of education in a special education classroom and physical therapy for patients post-stroke.

The focus of this future work will be to examine the factors that inform a proper design for a robot system for socially assistive applications. We plan to focus on the task-oriented nature of the system to gain insight into proper embodiment and interaction design.

REFERENCES

[1] C. Breazeal, A. Brooks, J. Gray, M. Hancher, J. McBean, D. Stiehl, and J. Strickon. Interactive robot theatre. *Communications of the ACM*, 46(7):76–85, 2003.

[2] K. Dautenhahn, B. Ogden, and T. Quick. From embodied to socially embedded agents – implications for interaction-aware robots. *Cognitive Systems Research*, 3(3):397–428, 2002.

[3] J. Eriksson, M. Matarić, and C Winstein. Hands-off assistive robotics for post-stroke arm rehabilitation. In

International Conference on Rehabilitation Robotics, pages 21–24, Chicago, IL, USA, June 2005.

[4] D. Feil-Seifer and M. Matarić. Defining socially assistive robotics. In *Proceedings of the International Conference on Rehabilitation Robotics*, Chicago, IL, USA, July 2005.

[5] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3–4):143–166, 2003.

[6] Brian P. Gerkey, Richard T. Vaughan, and Andrew Howard. The player/stage project: Tools for multi-robot and distributed sensor systems. In *Proceedings of the International Conference on Advanced Robotics*, pages 317–323, Coimbra, Portugal, Jul 2003.

[7] S.S. Intille. Designing a home of the future. *IEEE Pervasive Computing*, 1(2):76–82, April-June 2002.

[8] C. D. Kidd and C. Breazeal. Effect of a robot on user perceptions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3559–3564, Sendai, Japan, Sep 2004.

[9] S. Kiesler and J. Goetz. Mental models and cooperation with robotic assistants. In *Proceedings of Conference on Human Factors in Computing Systems*, pages 576–577, Minneapolis, Minnesota, USA, April 2002. ACM Press.

[10] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2149–2154, Sendai, Japan, Sep 2004.

[11] P. Maes. Modeling adaptive autonomous agents. *Artificial Life*, 1(1–2):135–162, 1994.

[12] T. Ono, T. Kanda, M. Imai, and H. Ishiguro. Embodied communications between humans and robots emerging from entrained gestures. In *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pages 558–563, Kobe, Japan, July 2003.

[13] R. Pfeifer and C. Scheier. *Understanding Intelligence*. MIT Press, Cambridge, MA, 1999.

[14] J. Rickel and W. Lewis Johnson. Task-oriented collaboration with embodied agents in virtual worlds. In *Embodied conversational agents*, pages 95–122. MIT Press, Cambridge, MA, USA, 2000.

[15] S. Woods, M. Walters, K. Lee Koay, and K. Dautenhahn. Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach. In *Proceedings the 9th International Workshop on Advanced Motion Control*, Istanbul, march 2006.

[16] J. Zhang and D. A. Norman. Representations in Distributed Cognitive Tasks. *Cognitive Science*, 18(1): 87–122, 1994.

[17] T. Ziemke. Are robots embodied? In Balkenius, Zlatev, Brezeal, Dautenhahn, and Kozima, editors, *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, volume 85, pages 75–93, Lund, Sweden, 2001.

APPENDIX

We include the two questionnaires filled out by the experiment participants. Figure 6 shows the questions answered after each condition. Figure 7 has the final questions, completed after the third condition.

Fig. 6. Form 1: Completed after each case.

Fig. 7. Form 2: Comparative rankings of the three conditions completed at the end of the experiment.