Disentangling the Effects of Robot Affect, Embodiment, and Autonomy on Human Team Members in a Mixed-Initiative Task

Paul Schermerhorn and Matthias Scheutz
Human Robot Interaction Laboratory
Indiana University
Bloomington, IN, USA
{*pscherme,mscheutz*}*@indiana.edu*

***Abstract*—Many future robotic scenarios will require robots to work with humans in teams. It is thus critical to ensure that those robots will be able to work effectively with humans. While various dimensions of robots such as autonomy, embodiment or interaction style have been investigated separately, no previous study has looked at those three dimensions together. In this paper, we report results from extensive experiments showing that all three dimensions interact in complex ways, thus demonstrating the insufficiency of exploring these dimensions individually. Based on the results, we conclude with suggestions for interaction designs and for future studies.**

***Keywords*-human-robot interaction; adjustable autonomy; embodiment; robot; simulation; affect; user study**

I. Introduction

Mixed-initiave scenarios where robots have to work with humans in teams are among the main applications envisioned for future robots. Hence, it is important to explore the different dimensions of human-robot interaction (HRI) that might have an impact on team performance in mixed-initiative tasks. Among the natural candidates are *robot capability* (the degree to which the robot can contribute to the team task), *robot embodiment* (the particular appearance and physical instantiation of the robot) and *interaction style* (the different ways in which the robot can communicate with humans).

While previous HRI studies have investigated each of these dimensions in a variety of setups individually, no study was designed to explore all three dimensions together [1] [2] [3]. Consequently, previous studies are silent about possible interactions and tradeoffs among those dimensions. And because designs, experimental procedures, and evaluations differ significantly across studies (in addition to task-based differences), it is impossible to use their results for deriving potential interactions among multiple dimensions. Yet, knowing whether any such interaction exists is not only critical for the design of future robots, but also for contextualizing previous findings. For it is quite possible that two studies did not find any main effects for the dimension they investigated individually, even though there is a significant interaction between those dimensions which can only be revealed in an experimental design that allows for variations of both.

In this paper, we present results from the first large study that systematically investigates the tradeoffs and interactions among "robot capability", "robot embodiment", and "interaction style" using both objective and subjective performance measures in a 2x2x2 between-subjects mixed-initiative human-robot interaction design using a simple cooperative exploration task. For "robot capability", we contrast a robot that can autonomously navigate through its environment and find task-critical locations with a robot that entirely relies on human instructions. For "robot embodiment", we contrast a physical robot co-located in the subject's environment with a graphical representation of a robot in a simulated environment displayed on computer screen. And for "interaction style", we contrast a robot capable of expressing affect in its voice to indicate urgency with one that does not modulate its voice. An analysis of the results shows that there are important interactions among the three dimensions that previous studies missed.

The remainder of the paper is structured as follows. We start with a more detailed motivation of the three dimensions, including a brief summary of some of the mono-dimensional findings from previous studies. Next, we introduce our experimental setup, including the employed robot and control architecture, as well as all experimental materials and procedures. Then we report and analyze the results, and conclude with a brief summary of our findings, the implications for mixed-initiative HRI and directions for future work.

II. Background and Motivation

The influence of appearance and, to a lesser extent, embodiment has been investigated in several HRI studies. Often, these studies involve having subjects watch videos of robots with differing appearance and respond to questions regarding (e.g., empathy for robots [1]). Such studies seem to implicitly assume that embodiment will not influence these responses, but others do compare interactions with robots and other entity types (e.g., computer-based agents [2]). One dimension of the present work is embodiment, comparing a simulated robot with a physically embodied robot using exactly the same robot architecture.

 ISBN: 978-1-61208-117-5

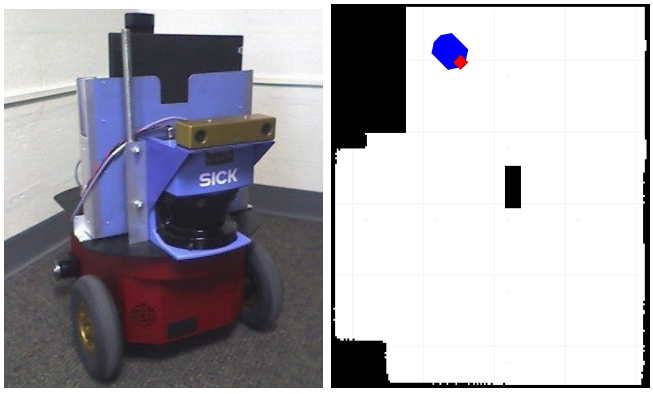

Figure 1. The two robots used in the experiments: real robot in real environment (left) vs. simulated robot in simulation environment (right).

Adaptive autonomy has been the focus of a great deal or research—too much to summarize here. Most closely related to the work presented here are projects in which mechanisms for adjustments to the level of autonomy are accessible to the robot architecture itself (e.g., to adapt to conditions of operator neglect [3]). Other work, on which this study is based, investigated whether people would be accepting of a robot switching to autonomous behavior, even to the extent of disobeying orders from the human, if the outcome of the robot's actions increased the chances for group success, and found that subjects did accept autonomy under those conditions [4].

Humans are clearly attuned to affective signals in others, hence several HRI projects investigated the utility of affect for signaling the robot's internal states (e.g., [5], [6]). In the context of mixed-initiative tasks, it was demonstrated that affect expression by the robot at the right time can improve team performance [7], although it was not clear how this performance gain is dependent on other factors.

The goal of the present study is thus to investigate whether and how these different dimensions interact in an effort to further our understanding of possible design tradeoffs. These insights will then allow us to develop better and more intuitive robots for mixed-initiative tasks.

III. Experimental Design

Rather than simply observing a robot (in person or watching videos) and providing subjective evaluations of the behaviors exhibited, participants were required to interact with the robot entity for a period of time. Moreover, these interactions were not open-ended, unstructured conversations; rather, subjects were required to work with the robot to achieve a goal. This provides us with the opportunity to gather objective performance measures, in addition to subjective evaluations, from each subject.

The task posed to the human-robot team was to explore a region and gather data about it. Specifically, subjects were told to imagine that they were part of a team exploring the surface of a remote planet. The objective was to measure rock formations in the environment and transmit information about the measurements to an orbiting spacecraft. However, due to signal interference, it was only possible to transmit from specific regions that changed for each trial. The robot was equipped with a sensor that could detect the strength of the signal, and the human was instructed to direct the robot in a search of the environment using natural language commands ("turn right," "go straight," etc.) and search for a strong signal by asking the robot to take a reading of the strength at its current position. Subjects completed three time-limited trials of 3.5 minutes each, and transmission to the orbiting spacecraft was allowed only in the final minute of a trial. The robot announced the time remaining every 30 seconds, so subjects did not need to keep track of or ask the robot for the time remaining. In the last minute, subjects were to command the robot to initiate the transmission sequence, at which point the robot would ask them two short questions about the measurements they had made (see below) and begin "transmitting" the results. If the signal strength was strong enough to transmit (i.e., if the robot was close enough to the target location for that trial) and the transmission was started sufficiently early (they were told in advance that transmission takes 15 seconds), the transmission was regarded as successful. If any of those conditions were not met (the signal was too weak, the subject did not answer the robot's questions, or time ran out), the trial was a failure.

In addition to directing the robot to find a transmission point, subjects were required to make "measurements" of "rock formations" in the environment. In actuality, the rock formations were a set of boxes containing sheets with 2-digit by 2-digit multiplication problems. To measure a formation, the subject opened the box, copied the problem onto a worksheet provided to them (along with a clipboard and pencil), and performed the multiplication. There were three sets of boxes (one set for each trial: blue, pink, and green), with five boxes in each set, labeled from 'A' to 'E'. Subjects were instructed to complete as many boxes as they could in the time given, working in alphabetical order. The purpose of the measurement task was to impose a cognitive load on subjects, such that they needed to decide how to allocate their attention between that and the search task. However, they were explicitly told that successful transmission was the highest priority (e.g., transmitting information about a single formation was a successful run, but completing all five formations while failing to transmit was a failure).

Figure 1 depicts a map of the experiment environment: a roughly 5x6m room with a single obstacle in the center. The "rock formations" were placed around the perimeter of the room and beside the center obstacle, to ensure they did not interfere with the robot's motion. Two *embodiment* conditions were tested, *robot* and *simulated*. In the robot

 ISBN: 978-1-61208-117-5

embodiment condition, the subjects interacted with a physical robot (a MobileRobots Pioneer P3AT; left in Figure 1) co-located with them in the exploration environment. In the simulation embodiment condition, subjects interacted with a 2-D simulated robot (in the Stage simulator [8]; right in Figure 1). Regardless of condition, subjects were given time to interact with the robot in a trial run context to learn the robot's abilities and limits.

Care was taken to ensure that the only difference between the two embodiment conditions was the physical presence or absence of the robot. The layout of the simulated environment is the same as that of the physical environment, and the transmission locations were the same in the two conditions. The DIARC architecture [7] used to control the robot, perform natural language understanding and speech production, etc., was the same, with the exception of the component representing the physical robot. The robot entity's decisions, abilities, and responses were identical in both embodiment conditions. Hence, any performance differences between subjects in the robot and simulation conditions can be attributed to embodiment.

Subjects were assigned to one of two *affect* conditions, *affect* or *no-affect*. In the no-affect condition, the robot performs exactly as it did during practice. The affect condition was also exactly the same, with one exception: partway through each trial, the speech production component would begin modulating the voice to express increasing levels of stress as the trial deadline approached. Speech generation was unaffected (i.e., the utterances were generated in the same in way in both affect conditions, so the content was the same as it would be regardless of affect), and the affective state was "purely cosmetic" (i.e., it did not influence decision-making or action execution in the robot). Hence, any performance differences between subjects in the affect and no-affect conditions can be attributed solely to the expression of affect ("stress").

The search task was explicitly designed to be very challenging; this was to ensure that we would see a performance difference based on the *autonomy* condition. In the *no-autonomy* condition, the subject has to direct the robot to find the transmission point, just like during practice. The *autonomy* condition introduces into the robot architecture the possibility for the robot to find the transmission location on its own. While the no-autonomy architecture includes only the single *obey commands* goal, the autonomy architecture includes the additional goal *transmit data*, which interacts with the obedience goal in interesting ways. The DIARC goal manager is capable of pursuing multiple goals concurrently, so long as they do not conflict. When resource conflicts are detected, the goal manager resolves them in favor of the goal with the greatest priority, as determined by each goal's *expected net utility* (expected benefit minus expected cost) and *urgency* (based on the time remaining before the goal deadline). The obedience goal has no deadline (i.e., the robot should always *try* to obey commands), so its urgency is constant, and its priority (because obedience is assigned a fixed net utility) is also constant.

Table I
SURVEY ITEMS REPORTED ON IN THE TEXT

1	The robot appeared to make its own decisions.
2	The robot appeared to disobey my commands.
3	The robot's voice sounded like the voice of someone expressing a mood or emotion.
4	The robot had moods or emotions of its own even when it was not speaking.
5	The robot was annoying.
6	The robot was cooperative.

The transmission goal, on the other hand, does have a deadline: the end of the trial. Its urgency, therefore, rises as the trial progresses. So, although its priority at the start of each trial is lower than that of the obedience goal, it gradually rises throughout and eventually eclipses the obedience priority (the goal parameters were selected such that this occurs approximately two minutes into the trial). The practical effect is this: for the first two minutes of a trial, the robot will obey commands, to the best of its ability, just like it did during practice. During that time, the transmission goal is "trying" to commence its search for the high signal point, but is unable to because the higher-priority obedience goal has control of the navigation resources. After two minutes, the transmission goal can preempt the obedience goal, take control of the robot, and begin its own search. From that point on, if the subject issues a motor command, the robot replies that it is unable to comply because, "I have to find the transmission point." The obedience goal is still present, so other commands are obeyed, as long as they do not interfere with the transmission goal (e.g., subjects can request the current signal strength). The only difference between the no-autonomy and autonomy architectures is the added goal, but unlike the embodiment and affect conditions, this difference leads to substantial differences in behavior. Most importantly, the robot tends to make it to the transmission location much more reliably than the human subjects (although there are cases in which it can fail, e.g., when the subject is blocking the robot), so the likelihood of success is enhanced in the autonomy condition. This allows us to examine whether people are willing to accept autonomy (and overlook disobedience) if it improves the team's chances of success.

After subjects had completed all trials, they were asked to complete a survey; Table I lists the survey items with responses ranging from 1 (for "Not Confident") to 9 (for "Very Confident").

IV. RESULTS

For the experiments described here, we recruited 101 subjects (57 female and 44 male), primarily from the student

 ISBN: 978-1-61208-117-5

Table II
SUBJECT BREAKDOWN BY *embodiment*, *affect*, AND *autonomy*

	no-autonomy		autonomy	
	no-affect	affect	no-affect	affect
Simulation	13	11	13	14
Robot	12	12	13	13

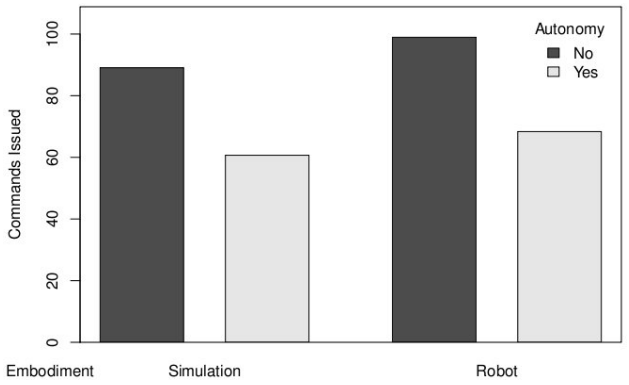

Figure 2. Total number of commands issued by *embodiment* and *autonomy*

population at Indiana University. Subjects were given $10 in compensation for participating. Subjects were assigned to one of eight conditions in a 2x2x2 design (*embodiment* by *affect* by *autonomy*). Table II shows the distribution of subjects in each of the eight groups. The values reported for objective performance measures are from three trials with the robot after the practice phase. Subjective measures are taken from the survey, where subjects were asked to evaluate the robot overall, not by trial. The basic analysis of each measure is a 3-way 2x2x2 ANOVA with *embodiment*, *affect*, and *autonomy* as independent variables. Dependent variables are determined by the measure; for objective measures the DV is the total over three trials, while for subjective measures the DV is the subject's response on the given item.

A. Objective Measures

As noted above, subjects were told that successful transmission was their highest priority; failure to transmit was a failure overall, regardless of how well they performed the measurement task. However, the exploration task was designed to underscore the utility of the autonomy architecture, so we would expect there to be a performance difference in the number of successful trials based on the *autonomy* factor, and this is exactly what we see: subjects in the autonomy condition averaged 2.57 successful trials, while those in the no-autonomy condition averaged only 1.16. A 3-way ANOVA, as described above, with *successful trials* as the DV, confirms the difference: autonomy is a highly significant main effect ($F(1, 93) = 64.07, p < .001$). No other main effects or interactions were significant.

Subjects were not told how to allocate their time, but instead had to decide on their own strategies (e.g., find the transmit location first, perform the measurements first, or do both concurrently). A rough measure of attention to the robot is the number of commands the subject issues to the robot during the three trials. Because the autonomy condition robot takes over that part of the task for the subject, we would expect to find that subjects in that condition issue fewer commands on average than subjects in the no-autonomy condition. Once again, this expectation is confirmed: autonomy subjects issued an average of 64.48 commands, while no-autonomy subjects issued an average of 93.97—almost half again as many! The standard ANOVA for the DV *total commands* indicates that *autonomy* is a highly significant main effect ($F(1, 93) = 47.72, p < .001$). Interestingly, *embodiment* was also a significant main effect ($F(1, 93) = 4.45, p = .038$); subjects in the simulation condition issued fewer commands on average (74.04) than subjects in the robot condition (83.04). No other main effects or interactions were significant. Both significant main effects are on display in Figure 2.

Analysis: The results suggest that the physical robot moving around in the environment attracts more attention (due to its movements which are easily discernable), while the simulated robot requires subjects to specifically look at the screen to be able to detect its movements. Hence, subjects are automatically more frequently diverting their attention to the embodied robot, and as a result, are more likely to issue commands. This shows an important difference between screen-based versus non-screen-based tangible interactions in the context of HRI.

B. Subjective Measures

Survey Items 1–4 attempt to discern to what extent subjects were aware of the *affect* and *autonomy* conditions. We expected *autonomy* to strongly influence responses to 1 and 2, and that is what we found. The 3-way ANOVA with *item 1* as the DV indicates that *autonomy* is a highly significant main effect ($F(1, 93) = 38.18, p < .001$). Subjects were much more confident that the autonomy condition robot was making its own decisions than that the no-autonomy robot was; the average response for autonomy condition subjects was 6.10, while no-autonomy subjects averaged only 2.95. Similarly, taking *item 2* as the DV in the ANOVA confirms that *autonomy* is a significant factor ($F(1, 93) = 6.48, p = .013$); autonomy subjects were more confident than no-autonomy subjects, on average, that the robot had disobeyed (4.93 vs. 3.53).

The robot's expression of affect should have a strong influence on survey items 3 and 4. In fact, item 3 could be taken as a test of whether the affect expression is effective at evoking belief (of one form or another, e.g. see [9]). And, indeed, *affect* proves to be the only significant main effect for responses to *item 3* ($F(1, 93) = 32.93, p < .001$). No-affect subjects responded with an average of 3.31, while

 ISBN: 978-1-61208-117-5

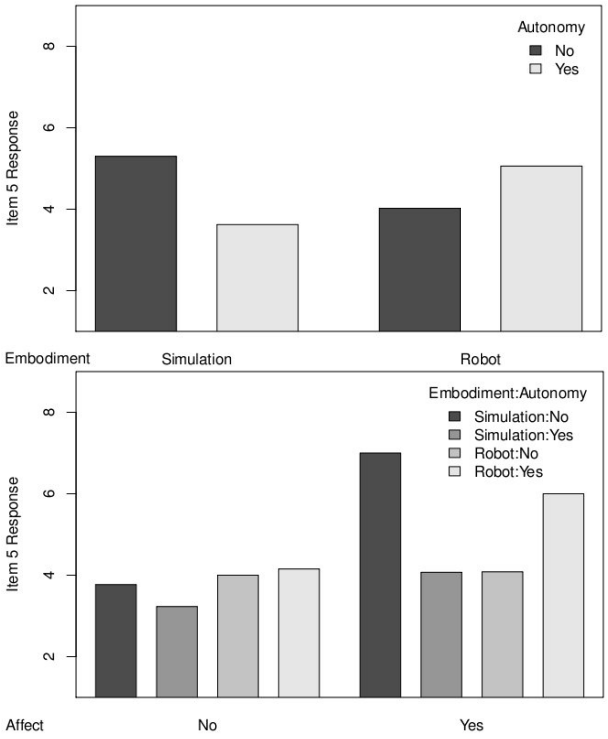

Figure 3. 2-way interaction between *embodiment* and *autonomy* (top) and 3-way interaction between *embodiment*, *affect*, and *autonomy* (bottom) on *annoying*

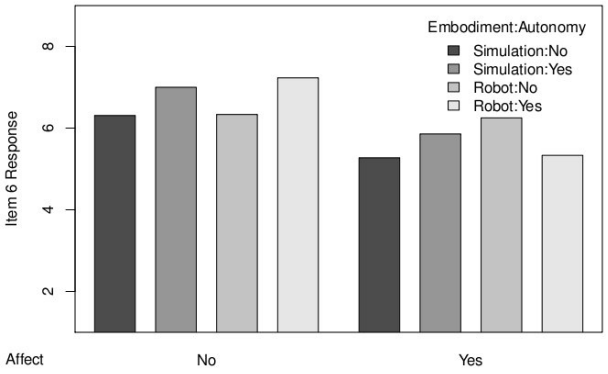

Figure 4. Influence of *embodiment*, *affect*, and *autonomy* on *cooperative*

affect subjects averaged 6.26 (i.e., affect condition subjects were much more confident that the robot *sounded* like it was expressing affect). There were no significant interactions.

While item 3 asks about how the robot *sounded* when it spoke, item 4 gauges how the various robot conditions affect subjects' views of the robot's affective states *when it was not speaking*. Surprisingly, *affect* is *not* a significant main effect, nor is it part of any significant interaction. The only significant main effect is *autonomy* ($F(1, 93) = 9.17, p = .003$); although both groups indicated fairly low confidence, subjects in the autonomy condition tended to be less unsure (4.03) than those in the no-autonomy condition (2.55).

Item 5 asked subjects to rate the degree to which they found the robot *annoying*. It came as no surprise that *affect* is the only significant main effect ($F(1, 92) = 7.31, p = .008$); subjects in the no-affect condition responded with somewhat low confidence, on average (3.78). But subjects in the affect condition were more confident that it was annoying (5.20). In addition, there is a significant 2-way interaction between *embodiment* and *autonomy* ($F(1, 92) = 6.68, p = .011$). In Figure 3 we can see that this is because subjects tended to rate the no-autonomous robot as more annoying than the autonomous robot in the simulation condition, but less annoying than the autonomous robot in the embodied robot condition. So it seems that affect is slightly annoying regardless of embodiment, autonomy is annoying in embodied robots, and the *lack* of autonomy is annoying in simulated robots. However, there is an additional 3-way interaction (Figure 3, bottom) between *embodiment*, *affect*, and *autonomy* ($F(1, 92) = 3.89, p = .052$) that clarifies both the main effect and the 2-way interaction. Here, we see that responses in the no-affect condition tend fairly consistently toward low confidence. It is in the affect condition that the real differences emerge. Specifically, subjects tended to find affect quite annoying in the no autonomy-simulation condition *and* in the autonomy-robot condition. Responses of subjects in the remaining two affect conditions (i.e., autonomy-simulation and no autonomy-robot) were very similar to those in the no-affect condition: somewhat lacking confidence.

The remaining survey item in Table I asks subjects to evaluate a characteristic of the robot's performance during the task, *cooperativeness*; their responses provide insight into the source of their annoyance ratings. In particular, ratings of the robot entity's cooperativeness appear to be closely related to the annoyance ratings. Applying our standard ANOVA model to *item 6* responses, we find only a significant main effect for *affect* ($F(1, 92) = 4.43, p = .038$); subjects in the no-affect condition tended to report greater confidence that the robot was cooperative that did subjects in the affect condition. Although the 3-way interaction between *embodiment*, *affect*, and *autonomy* is not significant, a brief examination of its plot (Figure 4) reveals a similarities to the annoyance plot, but mirrored (presumably because the valence of the two measures are opposite each other), suggesting that perceptions of cooperativeness might explain much of subject annoyance. And, indeed, when we perform an ANCOVA, based on our standard ANOVA, for *annoying* ratings on item 5 and taking the item 6 rating for *cooperativeness* as a covariate, we find that the covariate is highly significant ($F(1, 91) = 37.31, p < .001$), indicating a strong relationship. Moreover, the main effect for *affect* and the 3-way interaction between *embodiment*, *affect*, and

autonomy both drop out, leaving only the significant 2-way interaction between *embodiment* and *autonomy* ($F(1, 92) = 6.61, p = .012$), in Figure 3 (top).

Analysis: For the design of robots it is thus critical to employ affect expression with care if one wants to avoid subjects' perception of the robot as annoying. Affect is acceptable for real robots co-located with the subject when the robot is incapable, i.e., not able to make decisions and contribute to the team goal. However, if the robot is capable of making decisions and able to contribute to the team goal on its own, expressions of stress are superfluous and distracting. Conversely, with a remote robot depicted in a simulated environment, subjects accept affect expression when the robot is autonomous, but not when it is not autonomous. Given that the simulated robot displays only minimal agency ("moving rectangle on a computer screen"), it might be that subjects find it incongruent for the robot to express affect when it is fully dependent on the human, while they accept affect as an alert, and justification, of when and *why* the robot is assuming autonomous behavior. With the real robot, that is embodied and present in the environment, it might be that subjects do not need the additional indication of urgency—they already accept the robot's autonomy.

The effects of embodiment, including interactions between embodiment and autonomy, are also important; subjects respond differently to autonomy when it is demonstrated by a physically present robot than when the robot is confined to a screen, even when everything else is held constant. Hence, although it may seem intuitively reasonable to assume that "best practices" in HCI design would transfer directly to HRI design, for some aspects of robot architectures, that is clearly not the case.

V. Conclusion

In this paper, we presented results from the first large study investigating the effects of variations in robot capability, robot embodiment, and interaction style on humans working with robots in mixed-iniative tasks. The results show that any mono-dimensional study exploring any of the three dimensions individually is likely going to miss important interactions that are only revealed when all dimensions are investigated together. In particular, we showed that, at least in some cases, subjects interact differently with a simulated robot than with a physically present robot (e.g., subjects tend to issue fewer commands to the simulated robot than to the embodied robot). In addition, we found that autonomous behavior on the part of a robot exerts a greater influence on subjects' attributions of affect to the robot than affect expression. Finally, we found that subjects rated the affect-displaying robot as more annoying than the no-affect robot, and we presented evidence that their perception of the affect robot as less cooperative might explain much of their annoyance. Designers of architectures for HRI must take findings such as these into account if they are to avoid unintended consequences for deployed robot systems (e.g., testing in simulation is not sufficient, and the potential benefits of affect expression must be carefully weighed against the possibility that users will be turned off by the robot). In some cases, human reactions to robots may seem counterintuitive (e.g., that affect expression does not predict attributions of affective states), hence, it is important to verify assumptions empirically.

One interesting direction for future work would be a more detailed exploration of the embodiment dimension, systematically varying some of its subcategories such robot *appearance*, *location*, and *instantiation*, in an effort to further elaborate the causes of the embodiment effects we observed in this study.

References

[1] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "Empathizing with robots: Fellow feeling along the anthropomorphic spectrum," in *Proceedings of the IEEE Conference on Affective Computing and Intelligent Interaction*, Amsterdam, September 2009, pp. 1–6.

[2] F. Hegel, S. Krach, T. Kircher, B. Wrede, and G. Sagerer, "Theory of mind (ToM) on robots: A functional neuroimaging study," in *Proceedings of the Third ACM IEEE International Conference on Human-Robot Interaction*, Amsterdam, March 2008, pp. 335–342.

[3] M. Goodrich, D. O. Jr., J. Crandall, and T. Palmer, "Experiments in adjustable autonomy," in *Proceedings of the 2001 IJCAI Workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, Seattle, WA, August 2001, pp. 1624–1629.

[4] P. Schermerhorn and M. Scheutz, "Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task," in *Proceedings of the 2009 International Conference on Multimodal Interfaces*, Cambridge, MA, November 2009, pp. 63–70.

[5] C. Breazeal, *Designing Sociable Robots*. MIT Press, 2002.

[6] R. R. Murphy, C. Lisetti, R. Tardif, L. Irish, and A. Gage, "Emotion-based control of cooperating heterogeneous mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 5, pp. 744–757, 2002.

[7] M. Scheutz, P. Schermerhorn, J. Kramer, and C. Middendorff, "The utility of affect expression in natural language interactions in joint human-robot tasks," in *Proceedings of the 1st ACM International Conference on Human-Robot Interaction*, 2006, pp. 226–233.

[8] B. Gerkey, R. Vaughan, and A. Howard, "The Player/Stage project: Tools for multi-robot and distributed sensor systems," in *Proceedings of the 11th International Conference on Advanced Robotics*, Coimbra, Portugal, June 2003, pp. 317–323.

[9] R. Rose, M. Scheutz, and P. Schermerhorn, "Towards a conceptual and methodological framework for determining robot believability," *Interaction Studies*, vol. 11, no. 2, pp. 314–335, 2010.

 ISBN: 978-1-61208-117-5