Investigating the Effects of Robot Affect and Embodiment on Attention and Natural Language of Human Teammates

Thomas J. Donahue
Human-Robot Interaction Laboratory
Tufts University, Medford, MA, USA
Email: thomas.donahue@tufts.edu

Matthias Scheutz
Human-Robot Interaction Laboratory
Tufts University, Medford, MA, USA
Email: matthias.scheutz@tufts.edu

***Abstract*—HRI studies investigating human-robot interactions in mixed initiative teams typically only look at *macro-level* behaviors. Yet, an investigation of *micro-level* behaviors such as eye gaze fixations, attentional shifts, communicative acts, and others is often necessary in order to determine the exact influence of robot behaviors on human cognitive processes. In this paper, we report the first results from several novel analyses of *micro-level* behaviors obtained from video and audio recordings from previous HRI team studies. The analyses focus on the effects of both robot embodiment and affect expression in the robot's voice on the human's verbal behavior and allocation of attention. The findings show complex relationships among all factors that have to be better understood to improve team performance.**

I. Introduction and Motivation

The prospect of employing autonomous robots in mixed-initiative human-robot teams is increasingly becoming a reality as bold applications of robot technology are envisioned, often in high-risk environments: from search and rescue missions in disaster zones, to bomb disposals in civilian and military contexts, to mapping and exploration in wilderness zones, to NASA's ambitious ideas for robots working on space stations and planetary outposts, and many others. Common to all these scenarios is the hope that humans would be able to rely on their robotic teammate in much the same way that they rely on their human teammates. Yet, enabling human-like interactions in autonomous robots is a major research challenge because many modulating factors can have an influence on team-based interactions, even in human teams (including the team structure and the familiarity of the team members, the specific interaction environment and task, the training levels, background knowledge, and many more. Moreover, in the context of mixed-initiative human-robot teams, additional factors such as the robots' appearance and embodiment, known capabilities and behaviors, and levels of human trust can all have an impact on the kinds of possible interactions and their effectiveness.

An increasing number of HRI studies have attempted to investigate the different modulating aspects of human-robot interactions in team tasks. For example, [1] compared a co-located robot to a video feed of the same robot, and found that the co-located robot was treated in a much more human-like manner (e.g., giving the robot more personal space). [12] carried out a human-robot teaming task which investigated the effect of robots' affect expressions on task performance and found that utilizing affect expressions at key points in the task led to better performance. Moreover, [7] found that such affect expressions, when paired with a co-located robot, lead to higher task performance than when paired with a remote third-person view of the same robot. Similarly, [6], utilizing a human-robot interview scenario, showed that human interviewees were simultaneously more engaged and guarded in their disclosures with a physical compared to simulated robot. [4] tasked participants with investigating a pseudo-bomb threat with either a human or robot teammate and found that participants rated their workload as lower when they worked with a robot teammate as opposed to a human. In a similar vein, [2] investigated how multitasking affected a human teammate's reliance on the automation of a robot teammate in a target recognition task and found that participants' attentional control (ability to shift attention flexibly) was a significant factor in task success.

Most of the studies investigating human-robot team interaction, however, look only at *macro-level* behaviors and the evaluations are thus usually based on subjective measures such as post-experimental surveys (e.g., "How would you rate the robot as teammate?") or overall objective task performance measures (e.g., "Did subjects in condition X have a higher task performance than subjects in condition Y"). However, to understand in detail the temporal dynamics of human-robot interaction and the exact influence of robot behaviors on human cognitive processes (which often occurs at a subconscious level), we need to focus on *micro-level* behaviors such as the dynamics of human attentional shifts over time [14]. We address this problem by applying a methodological paradigm to data from a set of human-robot interaction experiments that allows us to investigate several *micro-level* aspects of team-based human-robot interactions, including: (1) the allocation of attention and attentional shifts based on interactions, and (2) various aspects of natural language exchanges (e.g., when they occur and how frequent they are, how long they are and how many words they contain, etc.). These aspects of interactions have been used extensively in psycholinguistic and developmental research in psychology to understand the dynamics of human-human interactions (e.g., [16]) and we will demonstrate that they can be equally applied to human-robot interaction, yielding insights about the different modulatory effects robot appearance, embodiment, and natural language behaviors can have on human cognitive processes such attention allocation and natural language communication. Understanding these effects is critical, for, as we will show, applying the wrong modulators in the wrong way at the wrong time can have dismal consequences for human cognition, causing high workload and frustration in humans that can ultimately lead to the human resentment of the robotic teammate.

We start with a brief overview of the data annotation and analysis methods together with a summary of the HRI study that generated the dataset we used for applying the framework. We then present five analyses of a subset of the data that demonstrate the complex interactions among human subject gender, robot affect expression, and robot embodiment. The subsequent discussion and conclusion sections highlight the implications of our findings and also briefly discuss future directions for further studies and applications of the framework.

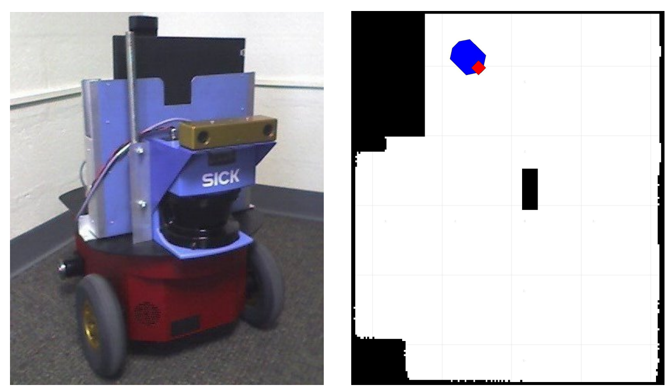
II. Experimental Data and Methods

A critical prerequisite for investigating *micro-level* behaviors and analyzing their effects on a human interactant's cognitive processes is the availability of a richly annotated data set. For team tasks this includes at the very least time-synchronized audio and video recordings of some interactions and activities of some team members during the task, although complete recordings for all team members from multiple perspectives would be better, with additional state-based information directly recorded by the robot (e.g., logging information from various robot sensors, state information from architectural components such as speech recognizers and parsers, planners, working memory, etc.). Even more useful are eye gaze recordings from eye-trackers (e.g., [15]) as they allow for a fine-grained tracking of eye gaze fixations, which, in turn, are often triggered by attentional shifts and can thus be used to measure a person's allocation of attention. Additional recordings from wearable brain sensors (such as EEG or fNIRs) as well as physiological sensors measuring basic bodily parameters (e.g., skin conductance, heart rate, etc.) can be used to make informed inferences about the moment-to-moment state of the human interactants. However, such complex complete data sets are not often available yet, thus necessitating the development of methods that can maximally utilize the currently typical audio and video recordings of experimental trials.

We will use, for all analyses in this paper, such a typical data set from human-robot interaction experiments where only time-synchronized video and audio recordings are available from an indoor human-robot team task. We will first summarize the task and the various conditions, and then describe how we transcribed and annotated the audio and video data to obtain a multi-modal corpus that can be datamined for patterns. We then describe the particular micro-level behaviors we were looking for in the corpus that could be affected by interaction modulators. A full description of the experimental design and overall results based on macro-level analyses procedures can be found in [10], [9].

a) The Team Task: The team task places the human-robot team in a hypothetical space exploration scenario which takes place on a remote planet with the goal of exploring the planetary surface. Exploration of the surface entails measuring rock formations in the environment and transmitting that information back to an orbiting spacecraft which can only be reached in a location in the environment where the signal is strong enough. Measuring a rock formation requires subjects to complete a set of 2-digit by 2-digit multiplication problems (instead of actually measuring rocks and determining their volume). The role of the multiplication was primarily to add cognitive load to the subject. There were 5 of these sets per trial, however subjects only needed to complete and transmit one of these sets to be successful. Finding a location with a strong enough signal for transmission required the subject to direct the robot through a handful of natural language instructions. These instructions included: "Go straight," "Turn left," "Turn right," "Go back" and "Take a reading." Transmission lasted 15 seconds and could only be initiated during the final minute of each 4 minute trial. Each subject completed two sets of 3 trials (referred to as a block) and each block's condition was applied independent of the other.

b) Experimental Conditions: Here we focus (for space reasons) on only two conditions of the original study, *embodiment* and *affect*. Embodiment has two levels: a physical robot and a simulated robot. Figure 1(a) shows the robot used in the physical robot-embodiment condition, a MobileRobots Pioneer P3AT, while Figure 1(b) shows the view of the simulated robot-embodiment condition (both the robot and the environment), utilizing the Stage simulator [3]. In order to limit the differences between the levels of embodiment to physical presence, the layout of the real and simulated environments were identical relative to the robot. Furthermore, both the real and simulated robots utilized the DIARC architecture [12] for control, ensuring each responded identically to both the commands of the subject and its environment.



(a) Real robot (b) Simulated Robot

Fig. 1. Comparison of real and simulated conditions

The affect condition is composed of three levels: no affect, medium affect and high affect. Affect is expressed through the modulation of the robot's speech production. Unlike the embodiment condition, the expression of affect is related to the progression of the trial. When the trial is the final 2 minutes, the medium level of affect begins, causing the robot's speech to be modulated to express increased urgency/stress. In the final minute, the modulation of the robot's speech changes to express an even more urgent/stressed state. Importantly, the expression of affect is limited purely to the speech production and has no effect on content or the robot's behavior.

c) Participants: Participants in the experiment were primarily comprised of students from Indiana University. Individuals in the analysis are limited to 2 conditions from the original 8 of the study. We selected only participants from conditions which had no experimental manipulation applied to block 1, in order to limit any ordering effects from the analyses. Fifty-five individuals (32 female, 23 male) were present in all but one analysis (see Section III-E), which was comprised of 31 individuals (17 female, 14 male).

A. *Annotations and Corpus Development*

In order to get at the attentional and natural language effects as a result of affect and embodiment, each subject trial was transcribed and annotated across a number of factors which centered around: the robot, the human subject, and the task. For the robot, both the robot's speech (i.e., what it said) and the tone (i.e., presence of affect) were recorded. For human subjects, everything they say during the trial (including both commands to the robot and miscellaneous speech), where they are looking (e.g., at the robot or their clipboard), what they are doing (e.g., a calculation) as well as any changes in tone or the presence of gestures are annotated. For the annotation data from two video recordings was used: from the main camera (which was mounted in a corner at the ceiling, providing a view of the entire experimental space) and from a head-mounted camera on the subjects' heads (which provided a better view of the subject's current viewpoint).

The software used to create the annotations was the open-source EXMARaLDA Score Editor [13]. In order to ensure the validity of the annotations, each subject was annotated once and then independently verified by another experimenter. The same experimenter verified each annotation in order to maintain consistency and validity across the annotations.

B. *Micro-Level Behaviors and Measures*

Attention allocation is determined a variety of cognitive and non-cognitive factors. Here we are specifically interested in visual attention as determined by eye gaze and shifts in visual attention that are manifested in eye saccades, or gaze shifts. For example, if a subject were working on a calculation while looking at the clipboard and then changes the gaze to the robot, this would count as an attentional switch and marked as such in the annotation. Attentional switches are categorized by the catalyst of the switch (cp. to [15]). A catalyst categorizes what was happening around the time of the attentional switch (e.g., was the robot speaking when the switch occurred?). We isolated six catalyst categories: *Preceding Human Speech*, in which the attentional switch occurs immediately prior to human speech beginning; *During Human Speech*, in which the attentional switch occurs as the human is speaking; *Following Human Speech*, in which the attentional switch occurs immediately after human speech has finished; *During Robot Speech* and *Following Robot Speech*, both of which are identical to their human counterpart and *No Catalyst*. Catalyst categorizations are made based on the proximity of the onset of the change in attention to another event. For every attentional switch, all possible catalysts within a second of the onset of the switch are compiled and whichever catalyst is closest to the switch within that window is chosen. If no catalyst is found within the onset window, the attentional switch is categorized as having no catalyst.

The natural language effects we consider for these study are more straightforward. During the trial, each command to the robot is characterized by its duration and the number of words that compose the command. Given the nature of the task, the speech productions are often filled with aborted words and disfluencies. In terms of the word count of a communication, disfluencies were not counted, while aborted words were. Further, though the commands available to the subject are limited, it was common for subjects to modify the commands with extraneous speech (e.g., "Would you... uh... take a reading"). Additionally, the number of commands given to the robot overall, across each trial.

C. *Hypotheses*

Based on our expectations of how attentional switches and instructions should interact with the autonomy and embodiment conditions, we formulated five main hypotheses:

H1: *Total attentional switches should be higher overall in the physical-embodiment condition than in the simulated-embodiment condition.* We expect that the physical nature of the robot in the room will draw more gaze with its movement. Furthermore, we expect participants will view the physical robot as a more credible teammate, which will translate to more human-like treatment and regard in terms of attention allocation.

H2: *Total attentional switches should decrease with affect as the trial progresses (i.e., from phase to phase).* We expect this result due to affect serving as a reminder that time is running out and that the participant needs to focus on their primary task.

H3: *Attentional switches occurring During Robot Speech should be higher overall in the physical robot condition than in the simulated robot condition.* We expect the physical robot to cause more switches because it will talk from different locations in the environment and robot speech can serve as a reminder for subjects to track the robot's position (which they need to do in order to give it instructions). This is different in the simulated condition where the robot can always be seen in the same location (i.e., on the monitor) and sound also always originates from the same location.

H4: *Attentional switches During Robot Speech when affect is present should be higher in the physical robot condition than in the simulated robot condition.* We expect that the lack of physical presence in the simulated robot condition will lead to participants taking the expression of affect less seriously.

H5: *Utterances should get shorter, by both word count and duration, as affect increases.* We expect that the expression of affect will serve as a reminder for subjects that time is running out and that they need to focus more on the primary task.

III. Results and Analyses

A mixed-design type-2 ANOVA was conducted on all attentional and natural language data with between-subjects factors *Gender* (male and female), *Embodiment* (real and simulated) and *Affect* (with and without) and, when affect was present, within-subjects factors *Phase* (with three levels corresponding to the time cut-offs – 1/no affect, 2/medium affect, and 3/high affect) and *Trial* (1, 2 and 3). All post-hoc analyses were computed using pairwise t-tests without correction. All dependent measures are normalized by phase length (i.e., Phase 1 is longer than Phases 2 and 3, which are the same length). Of the six attentional switch catalysts described in Section II-A, our analysis was focused on those which occurred During Robot Speech (DRS) as well as the Total number of attentional switches overall, as these two measures give the most concise perspective into the role of attention. Moreover, all three of the natural language features were used in our analysis, i.e., the total number of communications with the robot (NC), average communication length (CL), and average word count (WC) (as described in Section II-A).

A. Analysis 1

Collapsing over factors Phase and Trial, *Analysis 1* focuses on the overall trends. For attentional measures, there was a significant interaction of Gender and Affect on DRS attentional switches ($F(1,47) = 4.40, p < .05$). Post-hoc analysis revealed significant differences between the Affect/Male ($M = 0.84, SD = 0.81$) interaction and both the Affect/Female ($M = 1.45, SD = 0.82$) and No-Affect/Male ($M = 1.37, SD = 0.55$) interactions. Further, the difference between Affect/Female and No-Affect/Female ($M = 1.16, SD = 0.70$) was nearly significant ($p = .059$).

In terms of the natural language measures, there was a significant interaction of Affect and Embodiment on the average word count per communication ($F(1,47) = 6.77, p < .05$). Post-hoc analyses revealed significant differences between: Affect/Simulated interaction ($M = 1.79, SD = 0.50$) and the No-Affect/Simulated ($M = 2.14, SD = 0.19$), No-Affect/Physical ($M = 1.98, SD = 0.34$) and Affect/Physical ($M = 2.10, SD = 0.24$) interactions; No-Affect/Physical interaction and No-Affect/Simulated.

Discussion. The significant interaction of Embodiment and Affect on average word count per communication shows a stark contrast between real and simulated robots in terms of the effects of affect. When affect is present, participants use significantly more words per communication with a physical robot than a simulated robot. For simulated robots, the presence of affect significantly negatively affects communication, in terms of word count per communication with the robot, while affect has the opposite effect on physical robots. We predicted in **(H5)** that utterances would decrease in word count as affect increases, regardless of embodiment, because it would serve as a reminder to focus on the task. This result suggests that affect has a more complex role than merely as a reminder that time is running out, and that the embodiment of the robot plays a key role in how human-teammates communicate.

Moreover, the significant interaction of Gender and Affect attentional switches made During Robot Speech suggest that males and females respond to affect modulated speech in robots in opposite ways. Female participants were significantly more likely to switch their attention to the robot compared to male participants in the affect condition, on par with males in the no affect condition, where they were significantly more likely to switch their attention to the robot than when no affect was present. It is possible that this difference is due to task-based differences, with males performing differently in Phase 1 where no affect is present; however, the lack of a significant interaction of Gender and Phase in *Analysis 2* (see Section III-B) makes a task-based explanation unlikely, lending support to more basic gender differences (see also [8], [11]).

B. Analysis 2

To better understand the role of Affect across the trial, *Analysis 2* incorporated Phase as within subjects factor. In terms of the attentional switch measures, there was a significant 3-way interaction of Affect, Embodiment and Phase with both DRS attentional switches ($F(2,94) = 5.11, p < .01$) and Total attentional switches overall ($F(2,94) = 7.37, p < .01$). Post-hoc pairwise analyses revealed an identical pattern of significant interaction differences across both measures in Phase 1. Affect/Simulated (DRS: $M = 1.20, SD = 0.97$; Total: $M = 4.02, SD = 2.61$) was significantly different from No-Affect/Simulated (DRS: $M = 2.13, SD = 1.05$; Total: $M = 6.93, SD = 2.75$) and Affect/Physical (DRS: $M = 2.12, SD = 1.21$; Total: $M = 5.81, SD = 2.15$), while No-Affect/Physical (DRS: $M = 1.25, SD = 0.94$; Total: $M = 4.50, SD = 2.60$) was significantly different from No-Affect/Simulated and Affect/Physical. Additionally, there was a significant difference in DRS attentional switches between participants in the No-Affect/Physical ($M = 0.87, SD = 0.89$) and No-Affect/Simulated ($M = 1.54, SD = 0.86$) conditions in Phase 2.

On the natural language measures, a significant 3-way interaction of Affect, Embodiment and Phase was found with average word count per communication ($F(2,94) = 3.56, p < .05$) and number of communications overall ($F(2,94) = 3.59, p < .05$). In Phase 1, across both measures, a significant difference was found between Affect/Simulated (AWC: $M = 1.79, SD = 0.89$; NC: $M = 6.77, SD = 3.83$) and Affect/Physical (AWC: $M = 2.32, SD = 0.25$; NC: $M = 9.41, SD = 2.29$). Further significant pairwise differences with word count per communication were Affect/Simulated and No-Affect/Simulated ($M = 2.37, SD = 0.31$), as well as a significant difference between No-Affect/Physical ($M = 2.05, SD = 0.54$) and No-Affect/Simulated. In Phase 2, significant differences were found between No-Affect/Simulated ($M = 2.22, SD = 0.38$) and both Affect/Simulated ($M = 1.83, SD = 0.56$) and No-Affect/Physical ($M = 1.94, SD = 0.51$). Finally, in Phase 3, significant differences were found between Affect/Simulated ($M = 1.74, SD = 0.32$) and Affect/Physical ($M = 1.87, SD = 0.21$).

Discussion. In Phase 1, No-Affect/Simulated and Affect/Physical yield significantly more attentional switches overall and During Robot Speech than both Affect/Simulated and No-Affect/Physical. A similar trend is seen in the natural language measures, which echoes results found in *Analysis 1*. In terms of **(H4)**, this result is difficult to interpret. While the results show that affect, when coupled with a physical robot, yields significantly more attentional switches During Robot Speech than a physical robot without affect, it does so in Phase 1, where no affect manipulation is present. A possible explanation of this may be that experiencing affect has lasting effects on participants after the first trial, altering their behavior on subsequent trials even when affect is not present (i.e., in Phase 1) – we will investigate this further in *Analyses 3* and *4* below.

We predicted in **(H2)** that Total attentional switches would decrease with affect (from phase to phase) as the trial progressed and this prediction holds true. *Analysis 1* yielded a complex picture in terms of **(H5)**. The results of this analysis add to that complexity, with average word count per communication following our prediction, but the number of communications overall trending in the opposite direction. The results in terms of both **(H2)** and **(H5)** suggest that more work is necessary in determining affect's ability to, in general, instill focus/urgency in a human teammate, as well as it's effect on natural language.

C. Analysis 3

As a result of the significant interactions with Affect found in Phase 1 (despite the absence of the Affect manipulation), we now limit the analysis to Trial 1 data to control for changes in experiencing affect. Across both attentional switch and natural language measures, no significant interactions of Phase and Affect were found. A significant main effect of Phase was found on DRS attentional switches ($F(2, 94) = 10.31, p < .001$), Total attentional switches ($F(2, 94) = 14.48, p < .001$) and average word count per communication ($F(2, 94) = 3.91, p < .05$). All results followed the trend seen in previous analyses (i.e., attentional and language measures highest in Phase 1 and tapering off).

Discussion. The results of *Analysis 2* suggest that experiencing affect has effects that carry on even when it is no longer being directly experienced. *Analysis 3* is limited to Trial 1, thus the lack of any significant effect of affect in Phase 1 suggest the effects of affect in Phase 1 seen in *Analysis 2* are likely due to lasting effects of affect. The results of *Analysis 4* (see Section III-D) corroborate this interpretation.

D. Analysis 4

Given the results of *Analyses 2* and *3*, *Analysis 4* was limited to Trials 2 and 3 in order to investigate the differences by Phase after participants had experienced the affect manipulation once (effectively treating Trial 1 as a practice). With attentional switch measures, there was again a significant 3-way interaction of Affect, Embodiment and Phase with both DRS attentional switches ($F(2, 94) = 4.74, p < .05$) and Total attentional switches overall ($F(2, 94) = 7.37, p < .01$). Post-hoc analyses revealed an identical pattern of significant interaction differences across both measures in Phase 1. Affect/Simulated (DRS: $M = 1.35, SD = 1.06$; Total: $M = 4.35, SD = 2.71$) was significantly different from No-Affect/Simulated (DRS: $M = 2.25, SD = 1.09; Total$:M=7.30, SD=2.94) and Affect/Physical (DRS: $M = 2.16, SD = 1.22$; Total: $M = 5.96, SD = 2.24$), while No-Affect/Physical (DRS: $M = 1.22, SD = 1.01$; Total: $M = 4.33, SD = 2.80$) was significantly different from No-Affect/Simulated and Affect/Physical. In Phase 2, significant differences in both DRS attentional switches and Total attentional switches between No-Affect/Physical (DRS: $M = 0.81, SD = 1.09$; Total: $M = 2.54, SD = 1.69$) and No-Affect/Simulated (DRS: $M = 1.67, SD = 0.95$; Total: $M = 3.97, SD = 2.18$).

A significant 3-way interaction of Affect, Embodiment and Phase was found on the natural language measures number of communications overall ($F(2, 94) = 3.76, p < .05$) and average word count per communication ($F(2, 94) = 5.88, p < .01$). Post-hoc analyses revealed in Phase 1, for both the number of communication overall and average word count, significant differences between Affect/Simulated (NC: $M = 6.85, SD = 3.97$; AWC: $M = 1.78, SD = 0.89$) and both No-Affect/Simulated (NC: $M = 8.97, SD = 1.88$; AWC: $M = 2.46, SD = 0.16$) and Affect/Physical (NC: $M = 9.51, SD = 2.25$; AWC: $M = 2.35, SD = 0.25$). Also in Phase 1, significant differences by average word count were found between No-Affect/Physical ($M = 1.89, SD = 0.79$) and both Affect/Physical and No-Affect/Simulated. In Phase 2, a significant difference in average word count was found between No-Affect/Simulated ($M = 2.24, SD = 0.51$) and No-Affect/Physical ($M = 1.82, SD = 0.59$). In Phase 3, a significant difference in number of communications overall was found between No-Affect/Simulated ($M = 6.81, SD = 3.49$) and No-Affect/Physical ($M = 9.62, SD = 5.96$).

Discussion. *Analysis 4* is limited to Trials 2 and 3, thus the reappearance of significant effects of affect in Phase 1 suggest that experiencing affect does have lasting effects. Indeed, taken together, the results of *Analyses 2, 3* and *4* provide strong evidence that once participants experience affect in Phases 2 and 3, it altered their behavior toward the robot later, in Phase 1, when affect was no longer present.

E. Analysis 5

Finally, to isolate effects of embodiment from affect, *Analysis 5* was limited to participants who never experienced the affect manipulation. In terms of attentional switch measures, there was a significant main effect of embodiment of those which occurred During Robot Speech ($F(1, 27) = 8.90, p < .01$) with significantly more attentional switches with simulated embodiment ($M = 1.50, SD = 0.57$) than physical embodiment ($M = .91, SD = 0.58$). Additionally, there was significant interaction of Embodiment and Phase for Total attention switches ($F(2, 54) = 4.28, p < .05$). Post-hoc analyses revealed significant differences in Phase 1 between physical embodiment ($M = 4.50, SD = 2.60$) and simulated embodiment ($M = 6.93, SD = 2.75$).

Discussion. The results of *Analysis 5* were perhaps the most surprising. We predicted in **(H1)** and **(H3)** that attentional switches would be higher overall, and more specifically During Robot Speech, in the physical robot condition than in the simulated robot condition. *Analysis 5* directly investigated this by removing all participants in the affect condition from the analysis. Reviewing **(H1)** first, we see significantly more attentional switches overall in the simulated embodiment condition than in the physical embodiment condition in Phase 1 and maintaining the general trend across trials. Moving to **(H3)**, attentional switches During Robot Speech were significantly higher in the simulated embodiment condition overall. One possible explanation of such differences may be that participants viewed the simulated robot as less competent [5] and thus felt they needed to check in with it more often (much like they would with a less experienced human teammate). Another explanation may be that it was easier to keep an eye on the physical robot through their peripheral vision, while they needed to avert their gaze to the computer screen to check in with the simulated robot.

IV. General Discussion

Analysis 1 showed that affect expressions, when paired with simulated robots, negatively affect communication with the robot, while they improve communication when paired with physically embodied robots. Furthermore, it was shown that female and male participants respond in opposite ways to expression of affect, with females more likely to change attention to the robot when its speech is modulated with affect. In terms of future human-robot teaming scenarios, these results make clear that affect expressions must not be applied in a one-size-fits-all manner, but rather carefully utilized when the aspects of the given scenario are a good fit.

Analyses 2, 3 and *4*, taken together, suggest that experiencing affect with a robot has lasting effects on future behavior/interactions with the robot, even when affect is no longer present. Further research into this result is necessary to understand how these lasting effects may manifest themselves, and for how long, in the human-robot team dynamics. This result is especially important because much contemporary work in HRI often modulates affect expressions without consideration to such effects.

In *Analysis 5*, we found that simulated robot teammates garner more attention than physically embodied teammates. We suspect that the simulated embodiment condition resulted in significantly more attentional switches than the physical embodiment condition due to a lack of confidence in the simulated robot (see [5]) or that the physical robot was easier to keep track of in participant's peripheral vision. Confirming or disconfirming these alternatives is important to allow robots to get the attention of human teammates, especially in high-risk scenarios.

V. Conclusion

In this paper, we used audio and video recordings from a previous human-robot interaction study to build an annotated corpus that could be datamined for micro-level behaviors in human-robot interactions. We specifically focused on affect expression and robotic embodiment as modulators that can affect human attention allocation toward, and natural language dialogues with, a robot teammate. We found that affect and simulated embodiment pairs are, in general, a detriment to the team dynamic, and that affect is only a desirable trait in physically embodied robots. In addition, we found affect to be extremely polarizing in terms of attention allocation along gender lines, drawing significantly more attention from female participants. Finally, we found that simulated robot teammates garner more attention than physically embodied teammates, however, we suggest this may be the result of a lack of confidence in simulated robots. Overall, our analysis reveals the complex relationship affect and embodiment play on key *micro-level* behaviors that are vital to attaining successful human-robot team dynamics.

References

[1] W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati. The effect of presence on human-robot interaction. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. 17th IEEE International Symposium on*, pages 701–706, 2008.

[2] J. Y. Chen. Individual differences in human-robot interaction in a military multitasking environment. *Journal of Cognitive Engineering and Decision Making*, 5(1):83–105, 2011.

[3] B. Gerkey, R. T. 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*, volume 1, pages 317–323, 2003.

[4] C. E. Harriott, G. L. Buford, T. Zhang, and J. A. Adams. Assessing workload in human-robot peer-based teams. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 141–142. ACM, 2012.

[5] C. Kidd and C. Breazeal. Effect of a robot on user perceptions. In *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 4, pages 3559–3564 vol.4, 2004.

[6] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey. Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition*, 26(2):169, 2008.

[7] R. Rose, M. Scheutz, and P. Schermerhorn. Towards a conceptual and methodological framework for determining robot believability. *Interaction Studies*, 11(2):314–335, 2010.

[8] P. Schermerhorn, J. Benton, M. Scheutz, K. Talamadupula, and R. Kambhampati. Finding and exploiting goal opportunities in real-time during plan execution. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, MO, October 2009.

[9] P. Schermerhorn and M. Scheutz. Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task. In *Proc. of the 2009 International Conference on Multimodal Interfaces*, Cambridge, MA, November 2009.

[10] P. Schermerhorn and M. Scheutz. Disentangling the effects of robot affect, embodiment, and autonomy on human team members in a mixed-initiative task. In *Proceedings of the 2011 International Conference on Advances in Computer-Human Interactions*, pages 236–241, Gosier, Guadeloupe, France, February 2011.

[11] P. Schermerhorn, M. Scheutz, and C. R. Crowell. Robot social presence and gender: Do females view robots differently than males? In *Proceedings of the Third ACM IEEE International Conference on Human-Robot Interaction*, pages 263–270, Amsterdam, The Netherlands, 2008. ACM Press.

[12] M. Scheutz, P. Schermerhorn, and J. Kramer. The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 226–233. ACM, 2006.

[13] T. Schmidt and K. Wörner. Exmaraldacreating, analyzing and sharing spoken language corpora for pragmatics research. *Pragmatics-Quarterly Publication of the International Pragmatics Association*, 19(4):565, 2009.

[14] C. Yu, P. Schermerhorn, and M. Scheutz. Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems*, 1(2):13, 2012.

[15] C. Yu, M. Scheutz, and P. Schermerhorn. Investigating multimodal real-time patterns of joint attention in an hri word learning task. In *Proceedings of the 2010 Human-Robot Interaction Conference*, March 2010.

[16] C. Yu and L. B. Smith. Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, 119(1):21–39, 2012.