The presence of a humanoid robot can be detrimental to human performance in an attentional task

Davide Ghiglino  
Italian Institute of Technology

Davide De Tommaso  
Italian Institute of Technology

Adam Wojciech Lukomski  
Italian Institute of Technology

Agnieszka Wykowska (agnieszka.wykowska@iit.it)  
Italian Institute of Technology

Article

Keywords:

Posted Date: May 25th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2957390/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. 
Read Full License
Abstract

Being surrounded by others has enabled humans to optimize everyday life tasks, as the mere presence of others can improve performance in some daily tasks. At the same time, the presence of an audience can also be detrimental to an individual's performance. Still, it is unclear what happens when these “others” include artificial agents, such as robots. Literature has shown mixed results in understanding whether robots can be facilitators or distractors in joint tasks. To understand the impact that the presence of a robot might have on human attentional mechanisms, we designed a visual-search-based game that participants could play alone, under the surveillance of a humanoid robot, or in collaboration with it. Thirty-six participants completed this experiment (age = 26.44 ± 6.35, 10 males). Attentional processes were assessed using metrics of performance (i.e. search times), and eye-tracking (i.e. fixation duration and time to first fixation). Results showed that the presence of the robot negatively affected participants’ performance in-game, with longer search times and time to first fixation when the robot was observing them. We hypothesize that the robot acted as a distractor, delaying the allocation of attentional resources to the task, potentially exerting monitoring pressure.

Introduction

Living in social groups has been claimed to be beneficial for many animal species\(^1\). The increased survival rate accelerated brain development, and enhancement of cognitive abilities are just some of the many advantages of coexisting with other individuals\(^2\). Throughout the history of humankind, being surrounded by our conspecifics did not only grant us a higher survival rate but allowed for the optimization of everyday life tasks. From cultivating fields to industrial assembly lines, collaboration with others has been crucial for boosting productivity and reducing the workload of an individual.

When it comes to motor tasks, such as coordination or athletic performance, the mere presence of “others” seems to increase an individual's physiological (and psychological) arousal. If the task is familiar, a boost in physiological arousal tends to cascade and result in better performance, through an increase in the frequency of dominant responses (i.e. responses with the greatest habit strength)\(^3\). Initially, this effect has been studied within the drive theory of social facilitation framework\(^4–5\), giving rise to several different models and hypotheses\(^6–7\). Despite this theory stemming from the observation of facilitation during motor behavior, the presence of others seems to affect individuals’ performance also at a cognitive level.

In an fMRI work, Chib, Adachi, and O’Doherty (2018) investigated the neural correlates of social facilitation using a monetarily incentivized reward paradigm\(^8\). The authors found increased connectivity between participants’ ventral striatum (vSTR) and dorsomedial prefrontal cortex (dmPFC) in trials in which social facilitation occurred (i.e. when an audience was watching them). The authors claim that the increased brain activity in the vSTR reflects motivational encoding occurring when other individuals are watching participants during the task. This might indicate that when we are “watched” by someone while
performing a task, we are more motivated to perform better or to focus more on the task, rather than when we perform the same task alone.

Providing a systematic review of social facilitation theories goes beyond the scope of the current paper, but it is important to highlight that according to most social facilitation accounts, the (social) presence of an audience can improve individuals’ performance. Still, there are times in our daily life in which the presence of another individual is not necessarily beneficial. Every one of us experienced the (sometimes unpleasant) feeling of being observed by someone while performing a casual task, such as cooking or writing an essay. Indeed, even during daily activities, people tend to behave differently when they think they are being watched by someone else, a phenomenon known as the Hawthorne effect. Whether it is simply walking down the street or giving an athletic performance, having an audience observing you can be either facilitating or detrimental, depending on task demands, context, and personality traits. The Hawthorne effect has been studied and debated within several theoretical frameworks, highlighting the complex interplay between social presence, contextual factors, and individual differences.

Several studies examined the potentially detrimental effects of social presence on individuals’ performance. Indeed, elevated arousal levels can improve individuals’ performance, but only up to a certain point, especially if the task (or the environment) is unfamiliar to the individual. Therefore, in conditions under which the arousal becomes excessive, performance might dramatically deteriorate. This effect is also known as the Yerkes-Dodson law, which postulates that an optimal level of arousal can help individuals focus on a task, while a high level of arousal can impair the ability of an individual to concentrate. Under certain circumstances, the increase in an individual’s arousal can be connected with feelings of anxiety and stress. In particular, the phenomenon of “choking” (i.e. performing worse than one’s actual skills would allow) under monitoring pressure presents a fascinating counterpart of the social facilitation theory.

In an fMRI study, Yoshie and colleagues (2016) found out that the presence of an evaluative audience can worsen participants’ fine motor performance. While engaged in a motor task (i.e. feedback-occluded isometric grip task), participants reported higher subjective anxiety in the “observed” condition, compared to the “unobserved” one. Interestingly, in the “observed” condition, the authors reported increased activation of the posterior superior temporal sulcus (pSTS), when compared to the “unobserved” condition. As the pSTS is claimed to be a key neural substrate for social perception based on visual information, the authors claim that individuals performing a task needed to allocate additional attentional resources to monitor external observers. Such costs (i.e. the reallocation of attentional resources) might conflict with the execution of the task. In another study, Belletier and colleagues (2015) investigated the same phenomenon, which they define as “monitoring pressure”. The authors observed that being watched by an evaluative audience leads individuals to choke on executive control tasks. This is in line with the distraction-confict theory stating that, when an individual is performing a task, the mere presence of others generates an attentional conflict between attending to the observers and attending to the task.
The impact of social presence on performance has been widely studied in the context of human-human interaction\textsuperscript{22}. However, despite a growing body of literature investigating the impact of artificial agents’ presence on performance in collaborative tasks\textsuperscript{23–24}, it remains to be understood whether collaborative artificial agents, facilitate or impair individuals’ performance. This question is becoming of great importance, as interaction with social robots is becoming increasingly present in our lives. From industrial to clinical applications, robots are becoming part of our everyday life activities.

Past research demonstrated that the presence of a robot gazing at individuals while they perform a task modulates attentional orienting\textsuperscript{25–26}, social decision-making\textsuperscript{27}, and engagement\textsuperscript{28}. For example, in two studies, Spatola et al.\textsuperscript{24,29} found that individuals’ performance in a Stroop task\textsuperscript{30} improved when participants were observed by a social robot rather than when they were observed by a non-social robot or when they were not being observed at all\textsuperscript{24}. These results speak in favor of the social facilitation effect in human-robot interaction. However, in a recent work, Koban, Haggadone, and Banks (2021) did not find the same results using a similar Stroop paradigm\textsuperscript{31}. The authors did not find a substantial difference in individuals’ performance when participants were playing alone or when they were observed by a robot or a human, suggesting the existence of additional contextual factors that might influence social facilitation processes in human-robot interaction (e.g., individual’s familiarity with the task or with the observer; task difficulty). Similarly, Belkaid et al.\textsuperscript{27} showed that social signals exhibited by a humanoid robot impair participants’ performance in a social decision-making task.

One potential reason for such conflicting findings might be that the paradigms that are usually adopted to study social facilitation in human-robot interaction rely on classical attentional tasks that are not interactive (for example, the Stroop task\textsuperscript{30}). In recent work, Irfan and colleagues (2018) highlighted the importance of adapting classical experimental paradigms to more natural human-robot interaction environments, intending to increase the ecological validity of the approach\textsuperscript{32}. Furthermore, performance measures (i.e. accuracy, answer times) are often the only indicators used to assess participants’ cognitive (and attentional) engagement during human-robot interaction experiments. Recent literature demonstrated the potential that other measures of attention, such as eye-tracking metrics, may have in exploring social cognition mechanisms that underpin human-robot interaction\textsuperscript{33–34}).

To explore the effect of the social presence in a human-robot interaction scenario, we designed a visual-search game, in which participants were asked to find a letter hidden within pictures of naturalistic photographs. To monitor subtle changes in the allocation of attentional resources, we asked participants to wear a mobile eye-tracker during the task.

The game was designed to be played both alone and with another player. Importantly, the design gave participants the time and the freedom to explore the environment (including the social environment) during the experiment. This was meant to mimic more natural contexts of being involved in an attentional task (e.g., monitoring bag scans in security control) while being embedded in a social (and noisy) environment that can also be distracting.
We decided to focus on two eye-tracking metrics, previously used in research on visual attention\(^35\text{--}36\), namely fixation duration and time to the first fixation. The importance of the use of eye-tracking metrics to better understand human cognition has been recognized in both human-computer interaction\(^37\) and human-robot interaction\(^38\) studies. Therefore, we decided to include such metrics along with participants’ response times, to examine the effects of the robot’s presence on performance and attention.

The first group of participants was asked to play the game alone. After two weeks, the same participants were asked to play the game again, but with the humanoid iCub robot\(^39\) observing them. The second group of participants was asked to play a turn-taking version of the game, where the robot was both observing and collaborating with them. The combination of these three versions of the game (i.e. “Solo”, “Observation”, and “Interaction”) allowed us to investigate further contextual aspects that might play a role in social facilitation/distraction, such as the difference between a “passive” robot observer and an “active” cooperative one. This difference might be of high relevance, as in everyday life interactions observers are not always “passive”, especially when it comes to scenarios where robots can be deployed as human assistants.

**Results**

**Search Times**

To assess the effect of the presence and the role played by the robot on task performance, we analyzed the differences, across conditions, within and between participants’ average Search Times (STs), using separate generalized mixed models (GLMM). Within-subjects comparisons revealed a main effect of the presence of the robot during the task \([\beta = -0.007, t(19) = -2.55, p = .011, 95\% \text{ CI} = (-0.013, -0.002); \text{Fig. 1A}]\). Namely, when the robot was present, participants performed slower than when they performed the task alone \((M_{\text{Solo}} = 11.08 \pm 1.94; M_{\text{Observation}} = 12.02 \pm 2.14)\). Between comparisons revealed a main effect of the role played by the robot during the task \([\beta = -0.011, t(34) = -2.54, p = .016, 95\% \text{ CI} = (-0.020, -0.003); \text{Fig. 1B}]\). Specifically, when the robot was playing with the participants, they performed slower than when the robot was just observing them \((M_{\text{Observation}} = 12.02 \pm 2.14; M_{\text{Interaction}} = 13.86 \pm 2.14)\). Although the comparisons between the Solo and the Interaction condition resulted significant \([t(36) = 773.2, p < .001])\, with the STs of the Interaction condition being larger than in the Solo condition, the eigenvalue of the Hessian (inverse curvature matrix) at the maximum likelihood lead the model to not converge (for details, see the description of Package ‘lme4’, 1.1–29; see https://cran.r-project.org/web/packages/lme4/lme4.pdf). As the reference model for this comparison was nearly unidentifiable, the difference between the Solo and Interaction conditions will not be further discussed.

**Fixation Duration**

The effects of the presence and role of the robot on participants’ attentional engagement were analyzed by examining variations in Fixation Duration across conditions, using separate generalized mixed models
Within-participants comparisons showed no effect due to the presence of the robot \( [\beta = 0.019, \, t(19) = 0.37, \, p = .716, \, 95\% \, CI = (-0.086, \, 0.121)] \). Similarly, between-participants comparisons showed no effect due to the role of the robot \( [\beta = 0.084, \, t(34) = 1.00, \, p = .323, \, 95\% \, CI = (-0.087, \, 0.255)] \).

### Time to First Fixation

We also analyzed the effect of the presence/role of the robot on the initiation of engagement, by examining variations across conditions among Times to First Fixation (TTFF), using separate generalized mixed models (GLMM). Within-participants comparisons revealed a main effect of the presence of the robot during the task \( [\beta = 0.309, \, t(19) = 2.60, \, p = .013, \, 95\% \, CI = (0.008, \, 0.543); \, \text{Fig.} \, 2A] \). Specifically, the mere presence of the robot during the Observation condition negatively affected initiation of engagement in the task, delaying participants' TTFF compared to the Solo condition \( (M_{\text{solo}} = 0.08 \pm 0.16; \, M_{\text{Observation}} = 0.76 \pm 1.66) \). Between-participants comparisons revealed no effect due to the role played by the robot during the task \( [\beta = -0.238, \, t(34) = -1.75, \, p = .090, \, 95\% \, CI = (-0.514, \, 0.039); \, M_{\text{Observation}} = 0.76 \pm 1.66; \, M_{\text{Interaction}} = 0.19 \pm 0.33 \, \text{Fig.} \, 2B] \).

### Discussion

The present study aimed at investigating the impact of the social presence of a robot on human performance in an attentional task. To meet this aim, we designed a game that participants could play alone, under the gaze of the robot, or jointly with it, by taking turns.

Our first result on search times revealed that the presence of the robot negatively affected participants’ performance in the game. Specifically, participants became slower in providing their answers when the robot was observing them, compared to when they played the task alone. This is in line with the body of research highlighting the detrimental effect that social presence sometimes exerts on individuals' performance. In human-human interaction, choking under monitoring pressure may result from distraction, as well as from the interference of self-focused attention with the execution of automatic responses.

Past research also showed that completing a task in front of a robot observer elicits a higher perception of monitoring presence than completing the task alone or in front of another human. We can hypothesize that this is related to the unfamiliarity individuals might perceive when interacting with robots. Therefore, the mere presence of a robot observer might raise the feeling of monitoring pressure.

We can also speculate that the presence of the robot led participants to redefine the context of the task, distributing their attentional resources between the observer and the task itself, as if the observer needs to be monitored as an integral part of the game.

This explanation of distributed attentional resources is confirmed by the results we found on the time to first fixation. Indeed, participants took longer to make the first fixation on the task when the robot was present, relative to when it was absent. This confirms that the robot was acting as a distractor for the participants, delaying the allocation of attentional resources to the task, and potentially exerting...
monitoring pressure. We can speculate that this detrimental effect on performance is due to the unfamiliarity that participants perceived with the robot. Indeed, uncertainty about the unfamiliar has been vastly debated as one potential source of social distraction\(^42\). Robots, and humanoids, in particular, may constitute for most humans ambiguous and unknown entities, that cannot be subsumed under the category of “machines” but, at the same time, cannot be treated as natural agents either\(^43\). We speculate that such unfamiliarity contributed to the effect we found in this study. Indeed, performing a task in “co-presence” with such an unknown entity could require participants to “keep an eye” on it, drawing cognitive resources from the task they are performing.

This interpretation is also in line with the results we found when we compared the two roles played by the robot between the two groups of participants. Indeed, the moment the robot was playing an active role in the game, participants’ search time worsened even further. This suggests that while performing the task, participants were monitoring the robot even more extensively. We can speculate that being aware of the acting possibilities of the robot required the participants to distribute additional peripheral attentional resources between the Stage screen and the artificial co-agent while performing the task.

The detrimental effect of robot presence is also supported by the eye movement data, time to first fixation (TTFF) specifically. TTFFs are thought to reflect pre-attentive processes\(^44\) or covert attention\(^45\), meaning that they provide a measure of how quickly participants can shift their attention to a new stimulus in their visual field. Longer TTFFs usually indicate that participants are taking longer to process or react to the visual stimulus, which can have negative implications for task performance. Indeed, in our task, the “monitoring” condition resulted in longer TTFFs. This indicates that participants took longer to orient their attentional focus on the *Stage* (the visual search display), spending a longer time monitoring the robot when the robot was monitoring them.

Interestingly, however, fixation duration data did not show entirely the same pattern. In the comparison between performance in the “solo” condition and the monitoring condition, fixation durations on the main screen with the task did not differ across these two experimental conditions. We hypothesize that this divergence in the data reflects two different sides of visual processing, that is, covert and overt attention\(^41\). Indeed, the duration of fixations is a direct consequence of physically directing the eyes to a stimulus (the *Stage* screen in our task) and reflects overt attention\(^46–47\) while the delay in search times and TTFFs in the monitoring condition (relative to solo condition) might be related to covert attentional processes. Therefore, the different pattern between fixation duration and TTFFs suggests that engagement of overt attention in the task does not change across the conditions, but covert attentional resources are being redistributed to monitor the robot when it becomes part of the environment.

Interestingly, in the second comparison (the condition of being observed vs. interacting) neither fixation duration nor the time to the first fixation differed based on the role played by the robot, even though search times differed across these two conditions. This might suggest that there was an additional process, not related to attentional mechanisms that delayed search times in the interaction condition. Perhaps the process was related to social cognition, for example, engaging theory of mind regarding the
interaction partner\textsuperscript{48}. In a recent study, Belkaid et al.\textsuperscript{27} explored the effects of a humanoid robot's gaze (mutual or averted) on how people strategically reason in social decision-making situations. In their experiment, participants' decision times became larger when the robot was gazing at them, compared to when the robot was averting their gaze. Such a “delay” was paralleled by a differential effect in synchronized alpha activity during the period of eye contact, with higher alpha synchronization compared to averted gaze. These findings suggest that the more social the robot appeared, the higher the need for suppression of distracting information for the participant. Such a suppression might affect performance, as it happens in our experiment. We speculate that in the \textit{Interacting} condition of our experiment (i.e. where participants were more “socially” engaged with the humanoid, still not knowing exactly what to expect from its behavior) the “social” presence of the robot influenced participants’ fluency in their performance in the task, while not affecting covert attention. This is an interpretation in line with another recent study by Roselli et al.\textsuperscript{49} on the vicarious sense of agency in human-robot interaction. In their experiment, Roselli et al. showed that participants' performance in a joint task with a humanoid robot worsen as a function of the level of intentionality attributed to the robot\textsuperscript{49}. The authors claimed that, when another social agent is present in a joint task, it activates spontaneous mentalizing processes, which are fundamental to predicting its actions. This can disrupt individuals’ ability to make decisions and take action smoothly, as mentalizing requires cognitive resources that can compete with action selection processes. Indeed, anecdotal information collected during the debriefing with the participants highlighted that most of them expected the robot to help them during their turns, as much as they were trying to help it during its turns, as if it was treated like a social partner in that condition. The fact that the robot was neither providing hints to the participants nor responding to their suggestions, might have elicited the need to monitor the agent's behavior and engage in theory of mind processes.

\textbf{Limitations and future directions}

It is important to highlight that the sample size for this study was limited. Although mixed effect models have been shown to be effective in keeping Type-I error down to the nominal $\alpha$\textsuperscript{50–51}, the power of the current study might be not sufficient to test the impact of all variables we considered. Indeed, an alternative explanation for the lack of significant differences in fixation duration across groups might be due to a lack of power. As also suggested by the relatively small $\beta$ values reported as a measure of effect size, future studies (especially those with between-subject designs) might require larger samples.

A further potential limitation is that our study focused exclusively on the effects of the social presence of a humanoid artificial agent, without considering how individuals would perform the same task in front of another human. This decision was dictated by our specific interest in human-robot interaction, as a follow-up of literature on social facilitation in human-human interaction. However, given the novelty of the experimental paradigm, future studies might explore differential effects between human and robotic co-agents.

The final aspect that might be worth further investigation is whether familiarity with the robot or with the task affects social facilitation. Based on our results, we speculated that the unfamiliarity with the
humanoid boosted its distracting effect. Indeed past research showed that increasing familiarity with robots positively affects individual performance in human-robot interaction [54]. Similarly, we can hypothesize that individuals who are frequently exposed to robots require less effort to monitor robot actions during the interaction. This might be further tested in a study in which the degree of exposure to robots is experimentally manipulated.

As for the issue of novelty of the task, the within-group comparison showed that search times increased the second time participants were exposed to the task. This might be due to a decline in engagement, as participants were already familiarized with the task. However, the second group of participants, who were naïve to the task, showed longer search times than the first group who were familiar with the task. Therefore we hypothesize that the un/familiarity with the task did not bias our results, which were rather driven by the presence of the robot and not by a novelty effect. However, it would be interesting to explore with future studies whether repeated exposures in a full within-participants design affects search times in similar paradigms.

The current study reported that the presence of a robot negatively affects participants' performance in an attentional task, potentially due to the perceived unfamiliarity with the robot. The detrimental effect of the robot's presence was also supported by eye movement data. In particular, time to first fixation (TTFF) showed that participants spent a long time monitoring the robot when the robot was monitoring them. These findings have practical implications for the design and implementation of robotic technology, particularly in situations where robots are used to accompany human operators. It suggests that the design of robots should be carefully considered to ensure that they do not impact human performance negatively, particularly in attentional tasks. Furthermore, the study highlights the importance of familiarity with robotic technology, as lack of familiarity may contribute to the negative effects observed. Additionally, the study suggests that the engagement of theory of mind processes regarding the interaction partner may also affect performance fluency. This implies that the design of robotic technology should consider the impact of the presence of a robot for human performance in tasks that require attentional resources.

**Methods**

**Participants**

Thirty-six participants completed this experiment (age = 26.44 ± 6.35, 10 males). All participants reported normal or corrected-to-normal vision and no history of psychiatric or neurological diagnosis, substance abuse, or psychiatric medication. The first group of participants (n = 19, age = 24.11 ± 4.11, 4 males) played the visual search game twice. Participants within this group played the task alone during the first game (Solo condition), after three weeks, they played an identical game with the same task under the surveillance of an iCub robot (Observation condition). This within-subjects design allowed for identifying the effects of the mere presence of the robot on participants’ performance. The second group of participants (n = 17, age = 28.89 ± 7.45, 6 males) played the same visual search game only once, taking
tells with the same robot, and cooperating with it to accomplish the task (*Interaction* condition). We decided to not expose this group to the solo condition to minimize the effect of learning/transfer across conditions. In a between-subjects design, we aimed at comparing this group with the first group to understand whether active social presence (interaction) has a different impact on performance than mere passive social presence. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. Our experimental protocols followed the ethical standards laid down in the Declaration of Helsinki and were approved by the local Ethics Committee (Comitato Etico Regione Liguria). All participants provided written informed consent to participate in the experiment.

**Experimental Design**

Main task and Stimuli. The main task consisted of a visual search game. Participants were instructed to search for one letter (either a “Z” or an “M”) that was hidden within high-resolution images (3008 x 2000 pixels) of real-life environments (Royalty-free images were retrieved from Unsplash.com). By pressing the left button of a Logitech Adaptive Gaming Kit participants communicated that they detected the “Z” letter while by pressing the right button, they communicated the detection of the “M” letter. For each trial, participants were given 30 seconds to find the hidden letter and were instructed to be as fast as possible. Such a relatively long timeout was provided to increase the chance that participants distribute their attention to the entire environment, instead of focusing solely on the visual search task. For each correct answer, participants accumulated +1000 points, while for each incorrect answer they lost −500 points. Timeouts were associated with the highest penalty, corresponding to a subtraction of -1000 points. This choice was made to increase the chance that participants’ (overt) attention was focused mainly on the task, rather than solely on the exploration of the environment. Indeed, despite the high penalty in case of timeout, it is important to note that the average response time of participants at this task was 12.26 ± 2.33, suggesting that, in most of the trials, participants were able to detect the letter well before the timeout occurrence. Also, the high accuracy of participants (i.e. correct response provided in 87% of the trials on average) showed that participant did not commit too many mistakes even if it “costs” less than exceeding search time beyond timeout. The experimental setup comprised three screens (23.8” LCD screens, resolution: 1920 × 1080, see Fig. 3): (1) one horizontal, central screen, located in front of the participant, at the center of the table; (2) one vertical lateral screen, located on the participant’s right; (3) a second vertical lateral screen, located on the participant’s left.

Visual-search stimuli were presented on the horizontal screen, named the “Stage”, while the lateral screens were providing additional task-related information. Namely, the right screen, named the “Timer”, was presenting the time left for the participant to provide the answer, turning on at the beginning of each trial, and turning off after the response was provided, or the timeout occurred. Additionally, after the first 10 seconds from the presentation of the stimulus on the Stage, if participants hesitated in providing the answer, a hint would appear on the Timer screen. The hint consisted of highlighting, on a black square, the quadrant of the stimulus in which the letter was hidden. At the end of each trial, based on the correctness of the participant’s response, the left screen named the “Scorer” turned on, presenting feedback and an update of the score (see Fig. 4). The Timer and the Scorer were added to the Stage to
make the task more immersive and engaging for the participant, to simulate a gaming-like scenario. Specifically, we designed the setup to stimulate spontaneous shifts of attention toward lateral screens, so that the robot was not the only “distracting” element in the environment. However, it is important to point out that attending to the lateral screens was not needed for completing the task, as much as attending to the robot (when present) was not needed. This choice also allowed for isolating the impact of social presence per se in the within-group comparison between the Solo and Observation conditions, as both conditions involved other distractors (i.e. lateral screens), and they differed only concerning the social presence of the iCub robot.

Experimental conditions. The experiment comprised three conditions: (1) Solo condition, in which participants were playing the game alone; (2) Observation condition, in which participants were playing the game while an iCub robot was “observing” them; (3) Interaction condition, in which participants were taking turns with the iCub robot in playing the game. The Solo and Observation conditions comprised 30 trials each, while the interacting turn-taking condition comprised 60 trials (30 per player). During the Solo and Observation conditions, participants were told their scores would be compared with other players’ scores. During the Interaction condition, participants were asked to cooperate with the iCub to achieve the highest team score, which would be eventually compared with other teams’ scores. At the end of each session, a fake top-10 ranking of the other players was presented, so that the actual player/team always was the winning one. Eventually, nineteen participants underwent conditions (1) and (2), while seventeen participants underwent only condition (3).

Collected data. The sample size was estimated a-priori based on previous similar experiments\textsuperscript{52–53}. During each trial, we collected participants’ search times (calculated as the time between the stimulus onset and the participant’s response). Participants performed the task wearing an eye-tracker (Tobii Pro Glasses 2), which enabled collecting also gaze-related data (i.e. fixation durations, and time to the first fixation), and monitoring participant’s gaze location during each phase of the experimental sessions. Before and after the Observation and the Interaction conditions, participants were also asked to complete the Intentional Stance Test\textsuperscript{54} to investigate participants’ tendency to adopt the intentional stance towards the robot and how it changes throughout the experiment. However, analyses and discussions of the InStance Test data go beyond the scope of the current paper and are omitted from the manuscript\textsuperscript{55}.

The behavior of the robot. The behavior of the robot was different between the Observation and the Interaction conditions. In the former, the robot was “monitoring” the participant and the task. Specifically, the robot was mainly observing the Stage screen and turned its head only if and when participants were looking at one of the lateral screens. Such a turn of the head was introduced to make the behavior of the robot contingent on the participants’ behavior, potentially increasing the naturalness of the interaction. During the Interaction condition, the robot was not only observing participants but actively playing the game during its turns. To maximize the human-likeness of such a condition, the robot’s behavior was derived from the participants’ recordings collected during the Observation condition. Specifically, the robot was programmed to perform as an “average” participant in terms of search times and accuracy.
Additionally, the gaze of the robot was programmed to switch from the *Stage* to the other screens with the same timing and frequency as an average participant.

**Data Analysis**

To explore the effects of the presence and the role played by the robot on participants’ performance during the task, we adopted various mixed models on behavioral and eye-tracking data, using R Studio\textsuperscript{56}. For all the models we used the same approach, considering metrics acquired during the experiment as separate dependent variables, and the subjects’ intercept as a random factor. We used within-group comparisons to assess the effect related to the mere presence of the robot, by contrasting the *Solo* condition with the *Observation* condition. Then, we used between-groups comparisons to assess the effect related to the role of the robot, by contrasting the *Observation* condition with the *Interaction* condition. Thus, experimental conditions were included as the fixed factor within each model.

Regarding the metrics, we analyzed search times (STs) as our main indicator of task performance. Correctness was not considered a reliable indicator, given the high accuracy participants displayed across conditions (87.27\% on average). Given the positively skewed distribution of search times, we adopted two separate generalized mixed models (GLMM) to analyze the data, in which we considered the inverse Gamma distribution as a reference for the model.

For eye-tracking data, we defined three main areas of interest (AOI) a-priori: (1) The *Stage*, (2) the *Timer*, and (3) the *Scorer*. 88.12\% of total xations were recorded within the *Stage* AOI, 4.56\% within the *Timer*, and 7.33\% within the *Scorer*. Considering the insufficient amount of data points within the non-*Stage* AOIs, we focused our analyses only on the *Stage* area. We considered two main parameters for data analyses: (1) Fixation Duration, which was used as an indicator of the attentional engagement displayed by participants during the task\textsuperscript{57}, and (2) Time to First Fixation, which was used as an indicator of participants’ attentional focus on the task\textsuperscript{58}. Both metrics were log-transformed before data analysis, and separate linear mixed models (LMM) were applied.

Analyses were conducted using the lme4 package\textsuperscript{59} in R. We reported t-statistics along with p-values and parameter estimates ($\beta$), corrected using the Satterthwaite approximation for degrees of freedom\textsuperscript{60}, to show the magnitude of single effects, with bootstrapped 95\% confidence intervals\textsuperscript{61}. Due to the way mixed models partition variance, and the lack of consensus on the calculation of standardized effect sizes for individual model terms\textsuperscript{62}, only t- and p-values are associated with the main effects in the “Results” section. Nevertheless, it is important to point out that mixed models allow for superior control for Type I errors than alternative approaches\textsuperscript{51}. Additionally, all comparisons report parameter estimates ($\beta$) and relative confidence intervals, providing a measure of effect size\textsuperscript{63–64}. We reported the mean values of each dependent variable divided by condition to ease the interpretation of the results.

**Declarations**
Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant awarded to A. Wykowska, titled “InStance: Intentional Stance for Social Attunement. Grant agreement No: 715058). The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

All authors contributed to the design of the present study. AWL and DDT provided technical support. DG performed data collection. DG and AW analyzed the data. DG and AW wrote the manuscript. All authors reviewed the manuscript and approved it.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing Interests Statement

The authors declare that they have no known competing interests or personal relationships that could have appeared to influence the work reported in the paper.

References


Conference on Intelligent Robots and Systems (IROS) (pp. 6979-6985). IEEE.


https://doi.org/10.1037/met0000184


**Figures**

**Figure 1**

Rain-cloud plot summarizing the fixed effects of the robot on participants’ Search Times (STs). On the left: main effect due to the presence of the robot during the task; on the right: main effect due to the role of the robot in the task. Dots indicate data points. Horizontal lines inside the boxplots indicate the median value of each distribution. Horizontal lines outside the box plot indicate significant comparisons. Asterisks denotes a significant difference (**p < .001; **p < .01; *p < .05)
Figure 2

Rain-cloud plot summarizing the fixed effects of the robot on participants’ Time to First Fixation (TTFF). On the left: main effect due to the presence of the robot during the task; on the right: main effect due to the role of the robot in the task. Dots indicate data points. Horizontal lines inside the boxplots indicate the median value of each distribution. Horizontal lines outside the box plot indicate significant comparisons. Asterisks denotes a significant difference (***p < .001; **p < .01; *p < .05)
Figure 3

Experimental setup. The stimuli appeared on the central screen (*Stage*). The time left to provide the answer in each trial was presented on the screen located on the left of the participant (on the right of the robot). After each trial, the score was updated on the screen located on the right of the participant (on the left of the robot). Markers located on the corner of each screen allowed for the identification of the regions of interest (ROIs), facilitating subsequent eye-tracking analyses.
Figure 4

Example of trial sequence. The experiment comprised three main screens: a central screen presenting the stimuli, named “Stage”; a right screen presenting the time left to provide the answers during the task, named “Timer”; a left screen presenting the updates of participants’ points after each trial, named “Scorer”. Each trial started with (1) a fixation cross appearing on the Stage for 500 ms; (2) then, a stimulus was presented on the same screen, activating a countdown of 30 sec on the Timer; (3) after the first 10 sec, a visual cue could appear on the Timer screen, highlighting the quadrant of the Scorer in which the letter was hidden; (4) the trial continued until the timeout or participant response; (5) in either case, the Timer, and the Stage turned off, while the Scorer was activated, providing visual feedback related to the correctness of the answer.