Co-Representation with Social Robots: Evidence from the Social Inhibition of Return effect

Orit Nafcha*[1,2] , Georgia E. Kapetaniou[1], & Emily S. Cross*[1]

(1) Social Brain Sciences Lab, Department of Humanities, Social and Political Sciences, ETH Zurich, Switzerland
(2) Brain Sciences, Weizmann Institute of Science, Rehovot, Israel

Keywords: Social robots, human—robot interaction, social attention, co-representation, Inhibition of return

Main Text

Summary

As artificial intelligence continues to advance, people increasingly expect humanoid robots to demonstrate sophisticated social behaviours. To meet these expectations, it is crucial to design robots that align with human models of social cognition. This study explores the extent to which humans co-represent actions performed by social robots and how prior attitudes influence this process. We use the social inhibition of return (SIOR) paradigm, where individuals are slower to respond to targets in locations previously acted upon by themselves or by a partner. In a within-subjects design, participants engaged first with a robot co-actor and then with another human co-actor. Overall, our findings suggest that interacting with a robot fosters positive perceptions of the robot and evokes co-representation of their actions. Additionally, we found a trade-off between self- and social-inhibitory processes when interacting with a robot, in contrast to interactions with humans. Furthermore, prior attitudes, particularly the tendency to attribute human-like qualities to robots, were paradoxically found to reduce the SIOR effect. We discuss these results in relation to the uncanny valley effect, as well as broader implications for how human and robot factors that drive people's co-representation of robot behavior may support more natural and effective human—robot cooperation.

*Authors for correspondence: ornafcha@gmail.com & ecross@ethz.ch
*Present address: Brain Sciences, Weizmann Institute of Science, 234 Herzl St. Rehovot 7610001, Israel

Introduction

Social robots are increasingly being deployed to assist people with daily life in contexts such as healthcare, education, and the home[1,2]. Robots designed to engage people on a social level (known as social robots) are currently serving as companions for elderly individuals[3], assisting with search and rescue tasks[4], and helping administer psychological and physical therapeutic interventions[5,6]. While robotics technology continues to advance from an engineering perspective, understanding of the the psychological mechanisms and consequences of human acceptance and collaboration with such robots remains in its infancy. One significant challenge in the field is to understand which factors might facilitate human—robot interaction (HRI), specifically to encourage people to trust robots. Here, we aim to explore trust objectively from a novel perspective by examining action co-representation as an implicit, indirect indicator of trust.

Co-representation. Co-representation, defined as the ability to represent another's mental content by observing both their behaviour and its outcome, is a crucial capacity that supports the understanding of others' goals and intentions and facilitates learning from their actions[7]. Co-representation can occur at different levels, such as action co-representation, which entails a representation of only another person's motor response[8–10], or task representation, which is when an observer co-represents a stimulus-response mapping[7]. Some authors suggest that individuals use the same mental model they use when preparing to act themselves[11]. One important prerequisite for co-representation is the ability to distinguish between the self and the other[12]. This distinction enables the formation of separate models for one's own actions and actions of others, which can be co-activated during social interactions. Impairment of co-representational abilities has been documented in some psychological disorders, such as autism spectrum disorders (ASD)[13]. While co-representation in joint action has been widely investigated in humans [14], it has received far less attention in the field of HRI [7], where such research could offer valuable insights into how robots are perceived along a spectrum ranging from social agents to mere objects. To date, co-representation of a robot as a co-agent has mainly been studied using the joint Simon task [15,16] where co-representation has been shown to be influenced by participants' beliefs about the robot as an intentional partner [17,18](but see [15]), as well as by people's prior sensorimotor experience with the robot[19]. In the current study, we focused on a task that is considered to be more reflexive and automatic, the social inhibition of return (SIOR)[20,21].

Inhibition of return (IOR) [22] refers to slower responses for targets that appear in previously attended, compared to unattended, locations. It is a reflexive and automatic effect that is considered to be associated with activation in the superior colliculus, a subcortical brain region[23]. The IOR effect has been attributed to an inhibitory mechanism that delays the ability to reorient attention to previously attended locations, having evolved to improve foraging abilities[23,24]. It has been suggested that, similarly to IOR, a social inhibitory effect exists regarding locations already investigated by another individual. This effect is termed the 'social inhibition of return' (SIOR)[20]. That is, SIOR refers to the phenomenon of slower reaction times toward locations already acted upon by the co-actor in a previous trial [20,25]. In the computerized version of this task [21] participants, each in turn, respond to a peripherally presented target in two successive trials. The first trial is performed after the other responds and is designed to examine the social inhibitory process. The second trial for each is designed to study self-induced inhibitory processes. Previously, it was established that sociality is an integral aspect of SIOR [21,26] see also [27]. Nafcha et al (2020) found that the social effect vanishes without a partner, but that merely believing a human partner exists can be a sufficient social cue to induce SIOR [21]. In the current study, we have explored whether the SIOR effect extends to artificial agents and assessed where robots are positioned on the spectrum of human perception, from viewing them merely as mechanical objects to adopting a more mentalistic perspective that regards them as autonomous agents. Overall, our primary question in this study was to explore whether and how the identity of the co-actor partner (human vs. robot) affects action co-representation. In addition, HRI is still relatively uncommon; therefore, pre-existing attitudes are likely to play a significant role in shaping perceptions of robots, either as a mechanistic object or a mentalistic social partner.

Prior attitudes and individual differences. Prior work clearly demonstrates that theory of mind (ToM), or the ability to attribute mental states to others (such as their beliefs or intentions), relates to the mastery of being trusting [28]. ToM enables individuals to forecast behavior and intentions, providing indications regarding someone's trustworthiness. Having the ability to represent others (and the self) as mental agents is critical for attributing internal states to nonhuman agents, that is, for anthropomorphism. Anthropomorphization is the assignment of human-like mental states and traits to entities that are not human, namely animals or objects [29,30]. For instance, autistic individuals who struggle with mentalizing tended to provide non-mental, mechanical descriptions of intention-implying animations [31]. Attributing human-like characteristics to artificial systems, especially robots, significantly influences how humans perceive, value, and engage with them[32]. Moreover, converging evidence suggests that both the trait and the temporary state of being inclined to anthropomorphize enhance people's enjoyment of interacting with robots and foster greater trust in them[33]. We predicted that participants' predisposition to anthropomorphism [34], prior attitudes towards robots [35], and the extent to which people adopt an intentional stance towards robotic actions[36] will modulate action representation. Furthermore, based on prior literature, individual differences in empathy as a trait may also influence mental state attribution[37]. Individual differences in empathy as a trait may also influence the development of trust in a robot [38] and people's propensity to attribute mental states to a robot. For instance, see [39] who found that robots' expression of fear about memory loss led participants to report higher empathy and to attribute more mental states. The term "empathy" is a broad concept that concerns our reaction to the observed experiences of another person, enabling individuals to act alongside others [40,41]. Empathy refers to both cognitive and emotional reactions[40]. The cognitive component of empathy refers to the ability to adopt another person's point of view, to evaluate and consider their intentions, will, and beliefs (i.e., theory of mind). On the other hand, affective empathy reflects emotional responsiveness to another person's feelings and distress.

In the present study, we examined whether empathy modulates the SIOR effect. In particular, we predicted that the cognitive component would be related to the magnitude of the SIOR effect.

Overall, we aimed to achieve two main goals through this experiment goals. The first was to replicate previous findings [21], which showed that participants merely believing they were conducting the study with another person was sufficient to elicit the social effect. The second goal was to examine the extent to which acting with a robot as a co-actor evokes the SIOR effect. That is, we aimed to test whether participants would represent the robot's action to the location in the same way that people do when acting with a human co-actor and whether prior attitudes and experience with robots would modulate this effect.

Understanding whether and when people take into account the actions of artificial agents when acting together, and how prior attitudes shape this process, contributes to our understanding of when artificial agents are treated as social co-actors rather than tools, and may support the development of socially acceptable, trustworthy, and effective social robots with whom people can intuitively cooperate and whom they can trust.

Method

Participants.

We recruited twenty-five healthy participants (from the general population, pre-screened for Gender, handedness, and any history of major physiological or mental disorders) were recruited (age: M =26.24, SD=4.92). This number was based on a preregistered power calculation (G*Power 3.1.9.4) for a repeated measure ANOVA (within factors) with a significant level of 5% for the question of whether the identity of the co-actor influences the SIOR effect and assumed moderate effect size (0.25), a power of 80% is achieved with 24 participants. We recruited one more in case we would need to exclude someone. Moreover, the SIOR task is sensitive to social factors (e.g., gender, ingroup; see, e.g., [26,42]). Therefore, in the current study, we used a sample consisting only of women. We address this limitation in the Discussion. In addition, to maintain the same gender factor, we included a female human co-actor. Additionally, the robot introduced itself as a female robot. For a complete list of the inclusion and exclusion criteria, please refer to the supplementary material, Note 1. This study was approved by the ETH Zurich Ethics Commission (EK-2024-N-155).

Robotic Platform. In all the experiments, we used the Pepper robot (Softbank Robotics), a 1.2m tall, commercially available humanoid robot that functions in controlled and autonomous animation modes. Pepper features 20 degrees of freedom and runs a Linux operating system programmable using NAOqi libraries with Python or C++.

Transparency and Openness.

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the current study[43]. This information is detailed in a preregistration available on the Open Science Framework (OSF) as part of a project consisting of a series of studies that explore reciprocity and social attention: see https://doi.org/10.17605/OSF.IO/PAQ92. The SIOR task was programmed using E-Prime, and all questionnaires were administered via Qualtrics. For the interaction with Pepper, we developed dedicated applications that implement the scripted dialogue for the experiment; see the supplementary material, Note 8. Data analyses were performed with E-prime, R, Version 4.4.1(R Core Team, 2020), R Studio version 2024.9.0.375, and JASP (JASP Team, 2019, JASP Version 0.11.1). For all analyses, the alpha threshold was set to 5%. All data and analysis codes are available on OSF, see X.

Design and Procedure.

Before coming to the lab, participants first completed a series of questionnaires assessing attitudes toward robots: the *Robotic Social Attributes Scale* (RoSAS)[35], the Attitudes Towards Robots Measure (ARM)[44], the Intentional Stance Task (IST)[36], and individual differences related to theory of mind and empathy, the IRI [45]. This part conducted as part of an online survey before coming to the lab to assess prior attitude towards robots.

In the lab, the primary task involved participants performing the SIOR task in *a within-participant*, *counterbalanced* design. In one meeting, participants believed they were performing the task with another human, and in another meeting, they believed they were performing the task with a robot.

Upon arrival, and after signing an informed consent form in one room, participants were directed to meet the robot in a separate room. The introduction to the robot (Pepper robot, Softbank Robotics) in the experiment included Pepper joking and talking about its hobbies (see Video 1 in the supplementary material and Note 8 for the text). Next, the experimenter explained the SIOR task instructions to both the participant and the Pepper robot. The participant then moved to a separate room to begin the practice, after which the task was performed by the robot and participant in different rooms. Before leaving, participants observed Pepper turning to its computer, which displayed an image of the task instructions identical to the one shown to participants in their experiment's room. After completing a practice trial, participants proceeded to the SIOR task. The participants completed debriefing questions after each session and at the end of the study. In addition, after interacting with the Pepper robot, participants completed the ROSAS questionnaire again. At the end of the study, we provided all participants with a full disclosure document (see the supplementary material Note 2) and we informed them that they had the option to refuse the use of their data, now that they understood the robot had not actually performed the task. None of the participants declined to have their data used.

Tasks and measurements.

Co-representation task. To examine co-representation as an implicit measure, we used the computerized version of the social inhibition of return task (in contrast to the board display version). This was chosen to manipulate social perception of the co-actor while all other factors remain constant (e.g., no action observation). Participants were informed that they were completing the task alongside the Pepper robot (see Figure 1a,1b) or another human. In reality, all responses from the co-actor were predetermined and fixed. These responses followed a pseudo-random order to ensure that all conditions were met. To mimic human behavior, the reaction times of the co-actor (the robot/human) varied, distributed around the mean RT from a previous pilot study, see the supplementary material Note 3 for further information. This procedure is common in studies that examine co-representation from afar, see [21,46,47].

In the SIOR task, each participant has two successive trials in which they perform a localization task: responding with the right key when a target appears on the right, and with the left key when it appears on the left. Participants used the 'P' and 'Q' keys and were told and shown that Pepper would use the keys shown in Figure 1a, placed on the left and right in front of the robot. The first trial was designed to examine the social effect (SIOR), which refers to how the actions of the other player influence participants. The second trial was designed to examine the individual effect, the inhibition of return (IOR). Targets could appear either in the same location as in the previous trial or in a different location. Thus, there were two types of trials: *same* and *different*, and two types of effects to observe: self and social (see Figure 1d). As in previous studies [21], each participant received a color cue that indicated whether it was their turn or the other player's turn. After the color cue, a target appeared on the left or right. When it was the participant's turn, the other co-actor could not see the appearance or disappearance of the target (see [21], for a full explanation). However, after each response, **both** actors saw an arrow in the middle of the screen indicating the direction of the response. This corresponded to the location where the target had appeared in the previous trial and to which either the other player (the robot) or the participant had reacted. Each block consisted of 33 trials, and there were a total of 20 blocks. Reaction time (RT) in the task was calculated as the time from target onset until the participant pressed the response key. We emphasized to all participants that this was not a competition with the robot or the human co-actor. Still, we asked them to respond as quickly and accurately as possible, since the target would remain visible until they responded. We explained the instructions to the participant and the robot together in the same room. Ultimately, the robot inquired if it could be reminded of the color it had been assigned. Participants then left the robot's room to read the instructions again and perform a practice run in another room. Before leaving, they saw the robot turn toward the computer and position itself in front of the screen. Full instructions are available in the supplementary material Note 4 and in [21], which provides further information on the computerized version of the SIOR task.

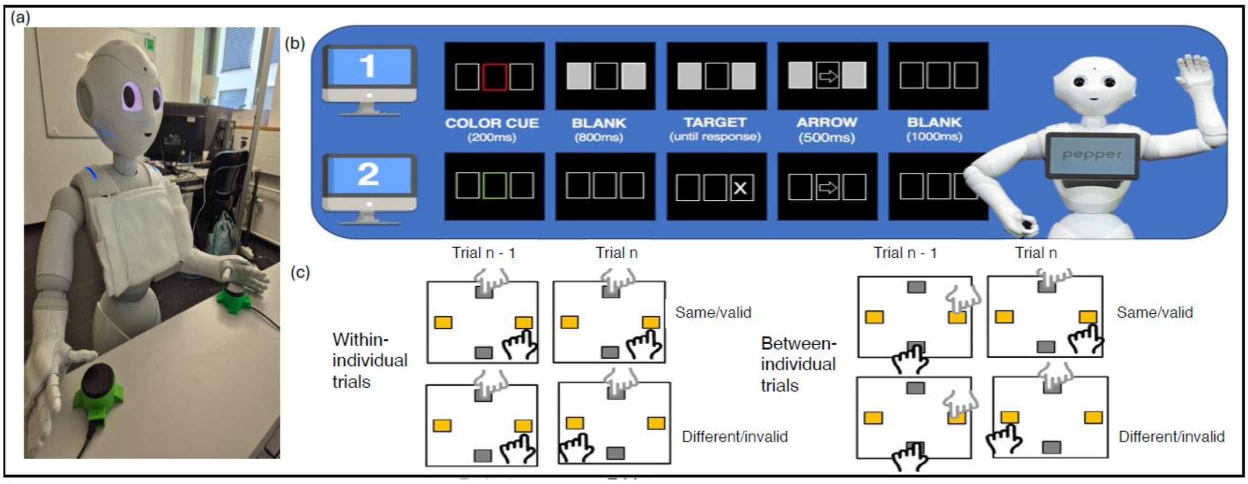

Figure 1. General Procedure and the Social Inhibition of Return task. (a). Pepper the robot in the lab, positioned in front of the screen and the response keys. (b) Trial sequence of the SIOR task. In this figure, it is the green actor's turn. (c) Task conditions: each participant has two successive trials. The first trial examines the social effect (between-individual trials), and the second examines the self-effect (within-individual trials). Each trial can be the same/valid (responding to the same location) or different/invalid (responding to a different location than in the previous trial.

Self-report questionnaires.

For examining attitudes toward social robots, we used the 'The Robotic Social Attributes Scale (RoSAS)'[35], the Attitudes Towards Robots Measure (ARM)[44] and the Intentional Stance task (IST)[36]. For identifying individual differences related to theory of mind and empathy, we used the Davis empathy questionnaire [45,48]. Participants completed these questionnaires at home prior to the main experimental session, except for the RoSAS, which they also filled out again after interacting with the robot in the lab. After the SIOR task, we asked participants several questions, including how much they enjoyed acting alongside their co-actor, their perception of time during the game, and how much they believed the robot or the human co-actor performed the actions.

'The Robotic Social Attributes Scale (RoSAS)'[35]. This scale is a validated tool that was used to get a better sense of the robot perception the participants hold (before and after an interaction with a social robot). The RoSAS is an 18-item scale with three subscales, each consisting of six items: Warmth (Organic, Sociable, Emotional, Compassionate, Happy, Feeling), Competence (Reliable, Competent, Knowledgeable, Interactive, Responsive, Capable), and **Discomfort** (Awkward, Scary, Strange, Awful, Dangerous, Aggressive). The higher the score in each subscale, the greater the user's perception of the robot possessing the characteristics described by the items in that subscale. Participants rated each subitem on a 9-point Likert scale, ranging from "definitely not associated " to " definitely associated." The rating refers to a general robot before the task and to Pepper after interacting with it, such that: Before: "Using the scale provided, how closely are the words below associated with social robots in general. And after:"Using the scale provided, how closely are the words below associated with the robot in the video/with the robot you performed the task with/Pepper?"

The Attitudes Towards Robots Measure (ARM)[44]. This questionnaire examines attitudes towards robots. It contains 15 items divided into three factors: prior anxiety (e.g., I am afraid that humanoid robots will make us forget what it is like to be human), prior acceptance (e.g., I would want to boast that I have a robot in my home), and prior anthropomorphism (e.g., I think robots will be able to perceive what I am going to do before I do it). These dimensions illustrate both negative (anxiety) and positive (acceptability) attitudes as well as expectations regarding robots' capabilities and their human-like characteristics. Participants rated each item on a 7-point Likert scale, ranging from "totally disagree" to "totally agree".

The Intentional Stance Test-2 (IST2)[36]. The Intentional Stance Task (IST) is a tool designed to measure the likelihood of people attributing mental states and intentions to robots, rather than viewing them purely as mechanical objects. This test measures the tendency to adopt an intentional stance towards robots by examining whether individuals interpret robot behavior through a mentalistic lens (attributing goals, desires, intentions) or a mechanistic lens (viewing actions as programmed responses). The task includes two factors: (1) a robot acting in isolation ("isolated robot") and (2) a robot interacting with human characters ("social robot"). There are six items per factor (12 in total). Each item presents a sequence of three photographs depicting a robot engaged in everyday activities. For each item, participants choose between two alternative descriptions of the scenario: one written in mentalistic terms (implying intentions, goals, or mental states) and the other in mechanistic terms (describing actions in physical or programmed language). Participants used a bipolar 100-point slider to indicate which description best matched their view. One sentence appeared at the left end of the scale and the other at the right. Higher scores indicate a more mentalistic view. Three items in each factor were reverse scored. They were explicitly instructed to move the slider toward the sentence they consider the more plausible account of the scenario. The order of the items was randomized.

The Davis empathy questionnaire [45,48]**.** The Interpersonal Reactivity Index (IRI) was utilized to evaluate multidimensional empathy. Participants rated how well each statement represented them on a scale from 1 to 5. They were instructed to read carefully and respond honestly. The IRI consists of four subscales, each with seven items: Perspective Taking (PT), Fantasy (FS), Empathic Concern (EC), and Personal Distress (PD). The items were presented in random order, and some were scored in reverse (See in the supplementary material, Note 5). Subscale scores were calculated by averaging the items. Higher scores indicate greater levels of Perspective Taking, Fantasy, Empathic Concern, or Personal Distress.

Results.

Exclusions. As stated in the preregistration, the following exclusions were applied:

Data exclusions. Only accurate trials were analyzed (87/16,000 were inaccurate; overall, 0.54% of all trials were excluded). Trials with durations shorter than 100ms, which are considered anticipatory errors, were removed from the dataset; this included 6 trials, accounting for 0.04%. Additionally, trials with reaction times (RTs) exceeding 2.5 standard deviations above or below the participant's mean RT for each specific condition (validity, effect, group, and participant) were also excluded. This accounted for a total of 418 trials out of 15,907, or 2.63%. The total number of valid trials included in subsequent data analyses is 15,489.

Participants exclusions. Participants whose overall mean RTs (across all variables) were more than 2.5 standard deviations above the average of their group (human or robot session) were excluded from the analysis. Overall, we did not identify any participants with a reaction time average that would exclude them based on this criterion.

Analyses.

Main experimental question. As preregistered, our primary question in this study was to explore how the identity of the co-actor partner (human vs. robot) influences action co-representation. We used generalized linear mixed models (GLMMs) to test whether reaction times were predicted by fixed-effects predictors for validity (same or different location), Person (IOR or SIOR), Condition (robot or human co-actor), as well as the interaction between these variables using the lme4 package (Bates et al., 2015). We further modelled participant-specific random intercepts and random slopes for all within-subject predictors. Please note that, since we declared our intention to conduct ANOVA in the preregistration, we have also included these analyses in Supplementary Material Note 6.

A linear mixed-effects model was conducted to assess the effects of **condition** (human vs. robot), **validity, that is target location** (valid - same location vs. invalid- different location), and **type of trial** (self IOR vs. SIOR, in the model it's called WithinSubVal) on reaction times (RTs). The model included all two-way and three-way interactions among these predictors, as well as random intercepts and slopes by subject. final_model <- lmer(RT ~ condition*Validity*WithinSubVal+(1 + condition + Validity + WithinSubVal | Subject), data = cleaned_trials, REML = TRUE). Model comparison using AIC, BIC, and likelihood ratio tests indicated that this model provided the best fit among the candidate models (AIC = 175081, BIC = 175227, logLik = 87522). This model also converged without singularity and was therefore retained for all analyses. The model was then refitted with REML for the statistical analysis. See the supplementary material, Note 7 and Table S1, for all other models and AIC, BIC. Results revealed significant main effects of **condition** (robot or human co-actor) (b = 17.63, SE = 6.83, t(28.21) = 2.58, p = .015 η^2 = 0.18) **Specifically, responses were slower when acting with a robot compared to when acting with another human co-actor,** and **validity** (acting to a same or different location) (b = 20.65, SE = 2.90, t(74.15) = 7.13, p < .001 η^2= 0.61), while **type of trial** (self or social effect) was not significant (b = -3.78, SE = 3.97, t(40.34) = -.95, p = .35, η^2 = 0.17). There were also significant interactions between **condition and validity** (b= -7.69, SE = 3.1, $t(15,385)$ = -2.48, p = .013, η^2 = 0.16), and between **validity and type of trial** (b= -10.15, SE = 3.1, $t(15,385)$ = -3.27, p = .001, η^2 = 0.2), indicating that the effect of validity depended on both condition and the type of effect. The interaction between **condition × type of trial** was not significant (p = .82, η^2 = 0.01), as well as the **three-way interaction** (p = .14, η^2 = 0.06).

Planned contrasts to examined simple effects revealed robust inhibition-of-return (IOR) and social inhibition-of-return (SIOR) effects across both experimental conditions. We have found IOR (*WithinSubVal = 1*) and SIOR (*WithinSubVal = 2*) for both conditions; that is, the participants were consistently slower on valid (same) compared to invalid (different) trials, reflecting a robust inhibition of return effect. In **condition 1**, where participants acted with another human as a co-actor, they were slower on valid compared to invalid trials in both the IOR and SIOR effects. In the IOR (*WithinSubVal = 1*), valid/same location trials(M=319ms, SD = 43.27) were slower than invalid/different location trials (M=298 ms, SD = 39.45), b = 20.65, SE = 2.89, $t(74.34)$ = 7.13, p < .001, Cohen's d_z = 1.21). In the SIOR effect (*WithinSubVal = 2*), also the valid, same location trials (M=305ms, SD = 35.36) were slower than the invalid, different location trials (M=294.68ms, SD = 32.24), b = 10.5, SE = 2.89, $t(74.05)$ = 3.63, p < .001, Cohen's d_z = 0.99). In **condition 2**, where participants acted with the robot as a co-actor, the same pattern was present but with generally smaller effects. In the IOR effect, valid, same locations, trials (M=329 ms, SD = 43.88) were slower than invalid, different location, trials (M=316 ms, SD = 38.7), b = 12.96, SE = 2.89, $t(74.09)$ = 4.48, p < .0001, Cohen's d_z = 0.63). In the SIOR effect, valid, same locations trials (M=320.73 ms, SD = 44.75) were slower than invalid, different location, trials (M=311.6 ms, SD = 43.9), b = 9.13, SE = 2.88, $t(73.7)$ = 3.16, p = .002, Cohen's d_z = 0.81).

In the preregistration we predicted that the SIOR would be evoked with a human co-actor and would be absent or weaker with a robot. To test this, we compared the magnitude of the validity effect (valid − invalid) across conditions using planned contrasts. Contrary to the prediction, **the condition affected only the self-IOR, not the SIOR**. For IOR (WithinSubVal = 1), the effect was larger with the human than the robot (b = 7.69, SE = 3.1, $t(15,385)$ = 2.47, p = .013, Cohen's d_z = 0.421). For SIOR (WithinSubVal = 2), the between-condition difference was not significant (b = 1.37, SE = 3.09, $t(15,385)$ = 0.44, p = .65, Cohen's d_z = 0.095). **Thus, co-actor identity modulated the self-IOR effect but not the SIOR effect, which was observed similarly in both conditions** (see also Figure 2).

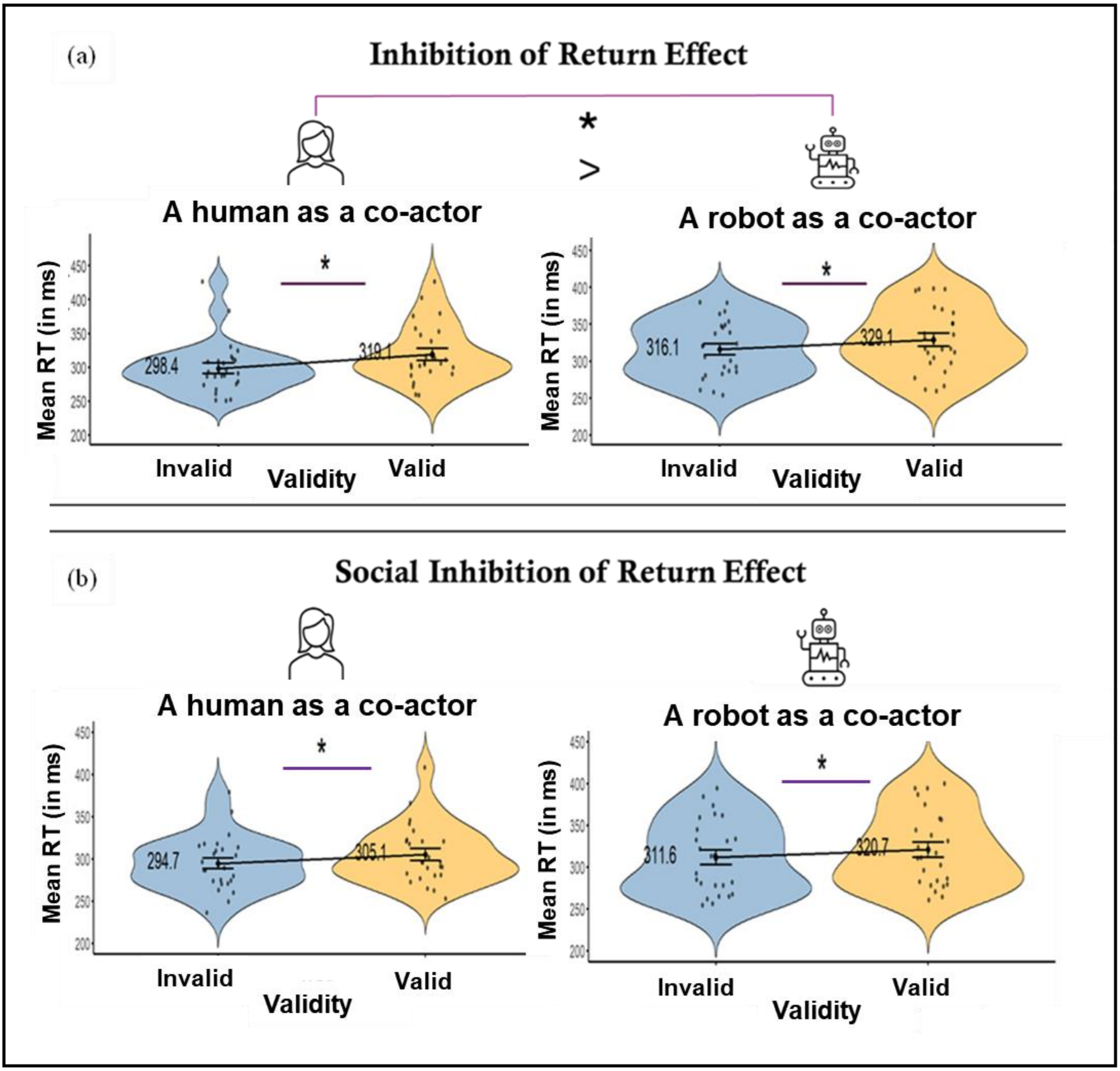

Figure 2. SIOR and IOR results. (a) Self-Inhibition of Return (Self-IOR): This refers to the effect of a participant's own previous response on their reaction time in subsequent trials. The Y-axis represents the mean reaction time (in ms), while the X-axis indicates the validity variable (where valid means the target is in the same location as the previous response, and invalid means it is in a different location). The points on the graph represent mean Rt per participant. The right panel shows data for a robot co-actor, while the left panel displays data for a human co-actor. (b) *Social-Inhibition of Return (SIOR): This measures the effect of the co-actor's previous response toward a location on the participant's reaction times. The axes and panel order remain the same as in (a).

Prior attitudes.

In our second preregistered experimental question we aimed to test whether prior attitudes toward robots modulate the SIOR effect. We predicted that higher scores on the RoSAS, IST, and ARM would each be positively correlated with SIOR in the robot condition, such that participants who attribute more social qualities, intentionality, or positive attitudes toward robots would show stronger co-representation of the robot's actions. In contrast to our predictions, we found that **perceiving the robot as more human-like attenuated the automatic inhibitory effect**. This suggests that greater attribution of human-like qualities to the robot **reduced**, rather than enhanced, the strength of SIOR.

ROSAS. To assess participants' attitudes toward robots, we administered the **Robotic Social Attributes Scale (RoSAS)** both at home (before the experiment) and again in the laboratory following the SIOR task. The RoSAS consists of three subscales: **Competence**, **Warmth**, and **Discomfort**. At home, participants rated how closely these attributes described **social robots in general**. In the lab, they rated how closely the same attributes described the robot with which they performed the task, namely, Pepper. To examine whether attitudes toward robots changed following the SIOR task, we first compared participants' RoSAS ratings before (general robots at home) and after (specific robot in the laboratory) the task. A paired-samples *t*-test revealed a significant **decrease in perceived competence** after the task, $t(24) = 3.98$, $p < .001$, Cohen's $d = 0.795$ (before: $M = 6.95$, $SD = 1.22$; after: $M = 5.87$, $SD = 1.14$). Perceptions of **warmth** did not significantly change, $t(24) = –1.03$, $p = .314$, Cohen's $d = –0.21$ (before: $M = 3.70$, $SD = 1.63$; after: $M = 4.08$, $SD = 1.67$). In contrast, ratings of **discomfort** significantly decreased following the task, $t(24) = 5.93$, $p < .001$, Cohen's $d = 1.185$ (before: $M = 3.29$, $SD = 1.57$; after: $M = 1.97$, $SD = 0.89$). Additionally, we investigated whether prior attitudes toward robots (RoSAS scores) were related to the SIOR effect. No significant correlations were observed between any of the RoSAS subscales taken before and the SIOR (Competence r=-0.107, p = .612, Warmth r=-0.105, p=0.616, Discomfort r=0.049, p=0.816). When examining the ROSAS subscales from after the SIOR task, we found a trend such that the greater the **Warmth** perception, the less SIOR effect (r=-0.38, p = .05); no other effect was found (Competence r=0.038, p = .85, Discomfort r=-0.12, p=0.544).

Intentional Stance. To further examine individual differences, we included a measure of intentional stance, in which participants rated a series of pictures on a scale ranging from a mechanistic interpretation (evaluating the robot as machine-like) to a mentalistic interpretation (evaluating the robot as more human-like). **Pearson correlational analyses revealed that in the robot condition, higher social scores were associated with a reduced SIOR effect** (r= –.43, p = .032). **Likewise, higher solo scores were also negatively correlated with SIOR** ($r(23) = –.465$ $p = .019$). These findings may suggest that participants who tended to adopt a more **mentalistic stance** toward the robot, whether in social or solo contexts, showed a **reduced SIOR effect**, indicating that anthropomorphizing the robot may diminish automatic processes, such as the inhibition of return.

ARM Questionnaire. We further examined whether prior attitudes toward robots, as measured by the Attitudes toward Robots Questionnaire (ARM), were associated with the SIOR effect in the Pepper condition. No significant associations were observed for **prior anxiety** ($r(23) = –.16$, $p = .45$) and only a trend for the **prior acceptance** ($r(23) = –.37$, $p = .07$). In contrast, **prior anthropomorphism** was significantly correlated with SIOR ($r(23) = –.53$, $p = .006$), indicating that participants who tended to anthropomorphize robots more strongly before the task exhibited a reduced SIOR effect.

Empathy. In the third experimental question we asked whether empathy, and specifically cognitive empathy as measured by the perspective-taking scale, modulates the SIOR effect and interacts with the identity of the co-actor. In order to examined this we used the IRI questionnaire scores [48] We predicted that individuals high in perspective-taking would generally show stronger SIOR, and that co-actor identity would matter more for those low in perspective-taking, with larger SIOR when interacting with a human than with a robot. In contrast to our predictions in the preregistration, no effect was found (all $p > 0.05$). See the discussion about the limitations of that tool.

Belief analysis. Finally, in our fourth experimental question, we aimed to test whether the strength of participants' belief that the co-actor was actively performing the task modulated SIOR. Participants rated this belief on a scale ranging from 1(not at all) to 10 (very much). We predicted that participants who report not believing the robot or the human as a co-actor actually performing the task would not show SIOR. To examine this question, we fitted a linear mixed-effects model to reaction times from social trials (WithinSubVal = 2) with RT ~ condition × validity × belief_z + (1 + condition | Subject), where belief was z-standardized and the model included random intercepts and condition random slopes by subject. This model yielded the second-best AIC (86415) but the best BIC (86499) while also converging without singularity. Please note that we had one missing data point in the belief report in the human condition. In contrast to our prediction, the main effect of belief on RT was not statistically significant ($b = 5.19$, $SE = 6.08$, $t(41.27) = 0.85$, $p = .398$, $\eta^2_p < .0001$). In addition, no significant interactions involving belief were observed: belief × condition (($b = 0.41$, $SE = 3.66$, $t(23.44) = 0.11$, $p = .913$, $\eta^2_p < .0001$), belief × validity ($b = 0.69$, $SE = 0.81$, $t(7550.08) = 0.85$, $p = .394$, $\eta^2_p < .0001$), and the three-way interaction ($b = -0.24$, $SE = 0.81$, $t(7550.11) = -0.30$, $p = .766$, $\eta^2_p < .0001$). These findings indicate that individual differences in belief regarding whether the co-actor was actually performing the task did not modulate the SIOR effect.

Discussion and conclusions

The primary question we addressed here was how the identity of the co-actor partner (human vs. robot) affects action co-representation. Three main findings were revealed. First, we found the social inhibition of return (SIOR) across both conditions, which not only replicates previous findings with human co-actors, but also extends prior work by demonstrating people also co-represent the action of a social robot. Second, the IOR differed between conditions as a function of co-actor identity, indicating a trade-off between the two effects. Third, the questionnaire responses converge on the idea that the extent to which participants attribute human-like qualities to robots, whether through warmth ratings, mentalistic interpretations, or anthropomorphic tendencies, results in reduced SIOR effects. That is, perceiving the robot as more human-like paradoxically appears to attenuate the automatic social inhibitory effect.

Inhibition of return is an automatic reflexive effect [24]. Finding that the present manipulation modulates the effect such that a larger IOR is observed when acting with humans compared to a robot as a co-actor can be understood as a cognitive resource allocation process, whereby HRI requires a resource-demanding control process to represent the action of an artificial co-actor. This cognitive demand may result in less automatic processing, leading to a smaller IOR in the robot condition. Interacting with a robot likely imposes greater cognitive demands on participants, thus diverting control resources away from automatic, self-inhibitory orienting and thereby attenuating IOR. This interpretation fits with evidence that social cognitive interpretations are the default and less cognitively costly, especially under high load, whereas more mechanistic/physical processing requires extra cognitive resources and can interfere at early representational stages[49]. Further support for this interpretation comes from a key observation in the current study. Participants consistently demonstrated faster response times in the human condition compared to the robot condition, suggesting that interaction with a robot may impose a greater cognitive effort. Overall, these findings suggest a trade-off between automatic processes tied to one's own action (self-IOR) and those influenced by observing another's action (social processes). Greater IOR with a human co-actor, or reduced IOR with a robot, may reflect increased top-down monitoring that draws resources away from automatic attentional mechanisms. When social processing proceeds more automatically, self-related inhibition is likewise more automatic and pronounced. This interpretation needs to be followed with further replication and exploration into the exact mechanisms at play, for instance, future studies should explore this effect using cognitive load and examine whether the trade-off between the social and self-effect will be pronounced as well.

While both the human and social robots as co-actors elicited significant SIOR, the strength of this effect was modulated by participants' beliefs about the robot's intentionality and human likeness. One possible explanation involves the uncanny valley framework [50,51]. When a non-human agent behaves in highly human-like ways, especially in social contexts, it can evoke feeling of discomfort and arousal, increasing cognitive control demands and disrupting the formation of a coherent social representation of the agent. As a result, this can hinder the automatic co-representation necessary for SIOR, leading to a weaker effect even when the perceived social presence is high and positive. Thus, the absence of effect might not reflect a lack of social meaning but rather too much of it, which may be experienced as unsettling and arousing. Consistent with this view, prior neuroimaging research has reported that perception of human-like robots is associated with heightened amygdala–prefrontal responses and arousal [52]. Furthermore, in another study it was found that anthropomorphic design might not only fail to reveal positive effects on trust but also actively distract participants [53]. Future studies should incorporate physiological indices of arousal (e.g., skin conductance, heart-rate variability, pupillometry etc.) to test whether elevated arousal accompanies attenuated SIOR and, critically, whether high arousal can eliminate the effect.

Limitations. In this study, we employed a female-only sample to replicate previous findings while keeping the gender variable constant. While this may limit the generalizability of our findings, from an evolutionary perspective, it is reasonable to suggest that women, who historically tended to search for food collaboratively, will demonstrate a larger SIOR, especially since this effect is believed to enhance foraging abilities [23]. In contrast, men, as hunters in a group, may have prioritized goal pursuit and representation over inhibiting locations acted upon by another member of the group. It would therefore be crucial to investigate the same effects in a male-only population and explore potential gender differences. Additionally, another direction for future research could explore whether excessive exposure to robots impacts performance. Specifically, it would be valuable to determine if interacting with robots becomes less effortful over time, causing both IOR and SIOR when acting with a robot to resemble more closely the pattern of effects observed when humans are co-actors. Regarding the lack of findings related to the Empathy questionnaire, we suggest that future studies use more nuanced cognitive measures that can reliably quantify participants' perspective-taking skills[54,55], counteracting the possible confounding effects of self-report tools [56].

To conclude, this study replicates previous findings demonstrating that merely believing one is acting with another agent is sufficient to elicit the SIOR effect [21], and that acting with social robots can also elicit this effect. Our findings raise several significant insights into our understanding of attention processes within social contexts and HRI. First, they demonstrate that the cognitive mechanisms underlying IOR are not purely automatic but are instead influenced by the social and cognitive demands of the interaction context. Second, our results challenge the assumptions about anthropomorphism in the robotics field, revealing that greater human-like qualities do not necessarily facilitate social cognitive processes. The paradoxical decrease in SIOR associated with stronger beliefs in a robot's human likeness suggests that a more complex relationship is at play, where, for instance, uncertainty about just how humanlike a robot actually is may interfere with automatic social representation processes. From a practical standpoint, the findings of this study emphasize the importance of recognizing that distinct (social) cognitive strategies may be required when working and collaborating with robots, as opposed to those employed when interacting with humans. Moreover, the influence of people's beliefs and expectations underscores the importance of managing user perceptions through interaction with and working alongside robots, rather than relying solely on the robot's physical appearance.

Acknowledgments

We used OpenAI's ChatGPT and Grammarly to improve grammar and writing clarity in portions of the manuscript. All authors reviewed and take full responsibility for all content. We thank Fanny Mougel-Alléoud and Greta

Bramow for helping as co-actors, Manuel Widmer for helping with programming Pepper, ETHZ's Decision Sciences Laboratory for helping with participant recruitment, and Pepper for behaving so nicely.

Author contributions

Conceptualization: O.N., and E.S.C, Methodology: O.N., and E.S.C. Software: O.N. Investigation: O.N. Formal Analysis: O.N, Code review G.E.K. Writing–Original Draft, O.N. Writing–Review & Editing, O.N., E.S.C., and G.E.K. Resources: E.S.C. Supervision: E.S.C.

Competing Interests

We have no competing interests.

References

1. Henschel A, Hortensius R, Cross ES. 2020 Social Cognition in the Age of Human–Robot Interaction. *Trends Neurosci.* **43**, 373–384. (doi:10.1016/j.tins.2020.03.013)

2. Heyselaar E, Caruana N, Shin M, Schilbach L, Cross ES. 2023 Editorial: Do we really interact with artificial agents as if they are human? *Front. Virtual Real.* **4**, 1201385. (doi:10.3389/frvir.2023.1201385)

3. Birks M, Bodak M, Barlas J, Harwood J, Pether M. 2016 Robotic Seals as Therapeutic Tools in an Aged Care Facility: A Qualitative Study. *J. Aging Res.* **2016**. (doi:10.1155/2016/8569602)

4. Delmerico J *et al.* 2019 The current state and future outlook of rescue robotics. *J. Field Robot.* **36**, 1171–1191. (doi:10.1002/ROB.21887)

5. Bekele E, Crittendon JA, Swanson A, Sarkar N, Warren ZE. 2013 Pilot clinical application of an adaptive robotic system for young children with autism. *http://dx.doi.org/10.1177/1362361313479454* **18**, 598–608. (doi:10.1177/1362361313479454)

6. Basteris A, Nijenhuis SM, Stienen AHA, Buurke JH, Prange GB, Amirabdollahian F. 2014 Training modalities in robot-mediated upper limb rehabilitation in stroke: a framework for classification based on a systematic review. *J. NeuroEngineering Rehabil. 2014 111* **11**, 1–15. (doi:10.1186/1743-0003-11-111)

7. Kirtay M, Wudarczyk OA, Pischedda D, Kuhlen AK, Abdel Rahman R, Haynes JD, Hafner VV. 2020 Modeling robot co-representation: State-of-the-art, open issues, and predictive learning as a possible framework. *ICDL-EpiRob 2020 - 10th IEEE Int. Conf. Dev. Learn. Epigenetic Robot.* (doi:10.1109/ICDL-EPIROB48136.2020.9278031)

8. Sebanz N, Knoblich G, Prinz W. 2005 How Two Share a Task: Corepresenting Stimulus-Response Mappings. *J. Exp. Psychol. Hum. Percept. Perform.* **31**, 1234–1246. (doi:10.1037/0096-1523.31.6.1234)

9. Sebanz N, Knoblich G. 2009 Prediction in Joint Action: What, When, and Where. *Top. Cogn. Sci.* **1**, 353–367. (doi:10.1111/j.1756-8765.2009.01024.x)

10. Sebanz N, Knoblich G, Stumpf L, Prinz W. 2005 Far from action-blind: Representation of others' actions in individuals with autism. *Cogn. Neuropsychol.* **22**, 433–454. (doi:10.1080/02643290442000121)

11. Hommel B, Müsseler J, Aschersleben G, Prinz W. 2001 The Theory of Event Coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.* **24**, 849–878. (doi:10.1017/S0140525X01000103)

12. Novembre G, Ticini LF, Schütz-Bosbach S, Keller PE. 2012 Distinguishing Self and Other in Joint Action. Evidence from a Musical Paradigm. *Cereb. Cortex* **22**, 2894–2903. (doi:10.1093/CERCOR/BHR364)

13. Baron-Cohen S, Leslie AM, Frith U. 1985 Does the autistic child have a "theory of mind" ? *Cognition* **21**, 37–46. (doi:10.1016/0010-0277(85)90022-8)

14. Sebanz N, Knoblich G, Prinz W. 2003 Representing others' actions: just like one's own? *Cognition* **88**, B11–B21. (doi:10.1016/S0010-0277(03)00043-X)

15. Stenzel A, Dolk T, Colzato LS, Sellaro R, Hommel B, Liepelt R. 2014 The joint Simon effect depends on perceived agency, but not intentionality, of the alternative action. *Front. Hum. Neurosci.* **8**, 595. (doi:10.3389/fnhum.2014.00595)

16. Strait M, Lier F, Bernotat J, Wachsmuth S, Eyssel F, Goldstone R, Sabanovic S. 2020 A three-site reproduction of the joint simon effect with the NAO robot. *ACMIEEE Int. Conf. Hum.-Robot Interact.* , 103–110. (doi:10.1145/3319502.3374783)

17. Wiese E, Metta G, Wykowska A. 2017 Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Front. Psychol.* **8**, 281017. (doi:10.3389/FPSYG.2017.01663/BIBTEX)

18. Stenzel A, Chinellato E, Bou MAT, del Pobil ÁP, Lappe M, Liepelt R. 2012 When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 1073–1077. (doi:10.1037/a0029493)

19. Abubshait A, McDonald CG, Wiese E. 2021 Examining Social Cognition with Embodied Robots: Does Prior Experience with a Robot Impact Feedback-associated Learning in a Gambling Task? *J. Cogn.* **4**, 1–16. (doi:10.5334/JOC.167)

20. Welsh TN, Elliott D, Anson JG, Dhillon V, Weeks DJ, Lyons JL, Chua R. 2005 Does Joe influence Fred's action? Inhibition of return across different nervous systems. *Neurosci. Lett.* **385**, 99–104. (doi:10.1016/j.neulet.2005.05.013)

21. Nafcha O, Shamay-Tsoory S, Gabay S. 2020 The sociality of social inhibition of return. *Cognition* **195**, 104108. (doi:10.1016/j.cognition.2019.104108)

22. Posner MI, Cohen Y. 1984 Components of visual orienting. *Atten. Perform. Control Lang. Process.* , 531–556. (doi:10.1162/jocn.1991.3.4.335)

23. Klein RM, MacInnes WJ. 1999 Inhibition of Return is a Foraging Facilitator in Visual Search. *Psychol. Sci.* **10**, 346–352. (doi:10.1111/1467-9280.00166)

24. Klein RM. 2000 Inhibition of return. *Trends Cogn. Sci.* **4**, 138–147. (doi:10.1016/S1364-6613(00)01452-2)

25. Cole GG, Skarratt PA, Billing RC. 2012 Do action goals mediate social inhibition of return? *Psychol. Res.* **76**, 736–746. (doi:10.1007/s00426-011-0395-7)

26. Nafcha O, Morshed-Sakran A, Shamay-Tsoory S, Gabay S. 2020 The effect of co-actor group membership on the social inhibition of return effect. *Acta Psychol. (Amst.)* **208**, 103119. (doi:10.1016/j.actpsy.2020.103119)

27. Gobel MS, Tufft MRA, Richardson DC. 2018 Social Beliefs and Visual Attention: How the Social Relevance of a Cue Influences Spatial Orienting. *Cogn. Sci.* **42**, 161–185. (doi:10.1111/cogs.12529)

28. Vinanzi S, Patacchiola M, Chella A, Cangelosi A. 2019 Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philos. Trans. R. Soc. B* **374**. (doi:10.1098/RSTB.2018.0032)

29. Heijnen S, de Kleijn R, Hommel B. 2019 The impact of human-robot synchronization on anthropomorphization. *Front. Psychol.* **9**, 408999. (doi:10.3389/FPSYG.2018.02607/BIBTEX)

30. Sanders T, Oleson KE, Billings DR, Chen JYC, Hancock PA. 2011 A model of human-robot trust: Theoretical model development. In *Proceedings of the Human Factors and Ergonomics Society*, pp. 1432–1436. SAGE PublicationsSage CA: Los Angeles, CA. (doi:10.1177/1071181311551298)

31. Castelli F, Frith C, Happé F, Frith U. 2002 Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* **125**, 1839–1849. (doi:10.1093/brain/awf189)

32. Hortensius R, Cross ES. 2018 From automata to animate beings: The scope and limits of attributing socialness to artificial agents. *Ann. N. Y. Acad. Sci.* **1426**, 93–110. (doi:10.1111/nyas.13727)

33. Zlotowski JA. 2015 Understanding Anthropomorphism in the Interaction Between Users and Robots. (doi:10.26021/2451)

34. Riek LD, Adams A, Robinson P. 2011 Exposure to Cinematic Depictions of Robots and Attitudes Towards Them. (doi:10.1007/s00146-006-0052-7)

35. Carpinella CM, Wyman AB, Perez MA, Stroessner SJ. 2017 The Robotic Social Attributes Scale (RoSAS): Development and Validation. *ACMIEEE Int. Conf. Hum.-Robot Interact.* **Part F127194**, 254–262. (doi:10.1145/2909824.3020208)

36. Spatola N, Marchesi S, Wykowska A. 2021 The Intentional Stance Test-2: How to Measure the Tendency to Adopt Intentional Stance Towards Robots. *Front. Robot. AI* **8**. (doi:10.3389/FROBT.2021.666586)

37. Lawrence EJ, Shaw P, Baker D, Baron-Cohen S, David AS. 2004 Measuring empathy: reliability and validity of the Empathy Quotient. *Psychol. Med.* **34**, 911–920. (doi:10.1017/S0033291703001624)

38. Cramer H, Goddijn J, Wielinga B, Evers V. 2010 Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 141–142. (doi:10.1109/HRI.2010.5453224)

39. Gena C, Manini F, Lieto A, Lillo A, Vernero F. 2023 Can empathy affect the attribution of mental states to robots? In *Proceedings of the 25th International Conference on Multimodal Interaction*, pp. 94–103. New York, NY, USA: Association for Computing Machinery. (doi:10.1145/3577190.3614167)

40. Shamay-Tsoory SG. 2011 The neural bases for empathy. *Neuroscientist* **17**, 18–24. (doi:10.1177/1073858410379268)

41. Decety J, Bartal IBA, Uzefovsky F, Knafo-Noam A. 2016 Empathy as a driver of prosocial behaviour: Highly conserved neurobehavioural mechanisms across species. *Philos. Trans. R. Soc. B Biol. Sci.* **371**. (doi:10.1098/rstb.2015.0077)

42. Tufft MRA, Gobel MS. 2022 Gender and perceived cooperation modulate visual attention in a joint spatial cueing task. *Vis. Cogn.* **30**, 6–27. (doi:10.1080/13506285.2021.1976892)

43. Simmons JP, Nelson LD, Simonsohn U. 2012 A 21 Word Solution. (doi:10.2139/ssrn.2160588)

44. Spatola N, Wudarczyk OA, Nomura T, Cherif E. 2023 Attitudes Towards Robots Measure (ARM): A New Measurement Tool Aggregating Previous Scales Assessing Attitudes Toward Robots. *Int. J. Soc. Robot.* **15**, 1683–1701. (doi:10.1007/S12369-023-01056-3/TABLES/12)

45. Davis MH. 1983 Measuring individual differences in empathy: Evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* **44**, 113–126. (doi:10.1037/0022-3514.44.1.113)

46. Atkinson MA, Millett AC, Doneva SP, Simpson A, Cole GG. 2018 How social is social inhibition of return? *Atten. Percept. Psychophys.* **80**, 1892–1903. (doi:10.3758/s13414-018-1546-3)

47. Schmitz L, Wahn B, Krüger M. 2023 Attention allocation in complementary joint action: How joint goals affect spatial orienting. *Atten. Percept. Psychophys.* , 1–20. (doi:10.3758/s13414-023-02779-1)

48. Davis MH. 1980 A multidimensional approach to individual differences in empathy. *JSAS Cat. Sel. Doc. Psychol.* **10**, 85. (doi:10.11249/JSBPJJPP.21.2_137)

49. Spatola N, Marchesi S, Wykowska A. 2022 Cognitive load affects early processes involved in mentalizing robot behaviour. *Sci. Rep.* **12**, 14924. (doi:10.1038/s41598-022-19213-5)

50. Mori M, MacDorman KF, Kageki N. 2012 The Uncanny Valley [From the Field]. *IEEE Robot. Autom. Mag.* **19**, 98–100. (doi:10.1109/MRA.2012.2192811)

51. Zhang J, Li S, Zhang J-Y, Du F, Qi Y, Liu X. 2020 A Literature Review of the Research on the Uncanny Valley. In *Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments* (ed P-LP Rau), pp. 255–268. Cham: Springer International Publishing. (doi:10.1007/978-3-030-49788-0_19)

52. Vaitonytė J, Alimardani M, Louwerse MM. 2023 Scoping review of the neural evidence on the uncanny valley. *Comput. Hum. Behav. Rep.* **9**, 100263. (doi:10.1016/J.CHBR.2022.100263)

53. Onnasch L, Hildebrandt CL. 2021 Impact of Anthropomorphic Robot Design on Trust and Attention in Industrial Human-Robot Interaction. *J Hum-Robot Interact* **11**, 2:1-2:24. (doi:10.1145/3472224)

54. Holland C, Shin SM, Phillips J. 2021 Do you see what I see? A meta-analysis of the Dot Perspective Task. *Proc. Annu. Meet. Cogn. Sci. Soc.* **43**.

55. Samson D, Apperly IA, Braithwaite JJ, Andrews BJ, Bodley Scott SE. 2010 Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 1255–1266. (doi:10.1037/a0018729)

56. Joel S, Sakaluk JK, Kim JJ, Khera D, Qin HY, Stanton SCE. 2025 Pseudo Effects: How Method Biases Can Produce Spurious Findings About Close Relationships. *Psychol. Sci.* **0**, 09567976251370262. (doi:10.1177/09567976251370262)

Figure and table captions

Figure 1. General Procedure and the Social Inhibition of Return task. (a). Pepper the robot in the lab, positioned in front of the screen and the response keys. (b) Trial sequence of the SIOR task. In this figure, it is the green actor's turn. (c) Task conditions: each participant has two successive trials. The first trial examines the social effect (between-individual trials), and the second examines the self-effect (within-individual trials). Each trial can be the same/valid (responding to the same location) or different/invalid (responding to a different location than in the previous trial.

Figure 2. SIOR and IOR results. (a) Self-Inhibition of Return (Self-IOR): This refers to the effect of a participant's own previous response on their reaction time in subsequent trials. The Y-axis represents the mean reaction time (in ms), while the X-axis indicates the validity variable (where valid means the target is in the same location as the previous response, and invalid means it is in a different location). The points on the graph represent mean Rt per participant. The right panel shows data for a robot co-actor, while the left panel displays data for a human co-actor. (b) *Social-Inhibition of Return (SIOR): This measures the effect of the co-actor's previous response toward a location on the participant's reaction times. The axes and panel order remain the same as in (a).