# Can It See or Just Look? The Role of Eyes and Beliefs in Visual Perspective Taking with Social Robots

Fărăgău Maria[1], Boudhar Djamel[2], Junko Ueno Strowski[1], Markos Kontogiorgos[1], Yongtao Yu[1], Nikolaos Savvidis[1], Hongyu Zhou[1], Rieken Xenia[1], Mikeladze Lasha Giorgi[3], Yerim Lee[4], and Idil Izgi Akyurt[4]

[1]Technische Universität Berlin
[2]Berlin School of Public Health
[3]Charité – Universitätsmedizin Berlin
[4]Freie Universität Berlin

August 2, 2025

## Abstract

Visual perspective taking (VPT) is fundamental to social cognition and interaction, including in human-robot contexts. Recent work has shown that humans can spontaneously take the perspective of humanoid robots, supporting the mere-appearance hypothesis, the idea that human-like visual features alone can trigger perspective taking. However, it remains unclear which specific visual or cognitive cues are necessary to elicit this effect. We conducted three experiments to investigate the role of eyes and beliefs about visual capacity in eliciting VPT1 toward humanoid robots. In Experiment 1, we replicated the prior findings that used a humanoid robot with human-like eyes and confirmed the presence of altercentric intrusion effects (Wahn and Berio, 2023). In Experiment 2, we tested whether these effects persist when the robot lacks visible eyes. In Experiment 3, we assessed whether belief manipulation – informing participants that the robot without visible eyes can still see – would reduce or eliminate perspective taking. Across experiments, participants completed a dot-matching task in which reaction times and accuracy were used to measure altercentric intrusions. Results showed clear evidence of altercentric intrusion in Experiment 1: Participants were significantly slower on inconsistent compared to consistent trials when responding from their own perspective to a robot with visible eyes. In contrast, no such effect was observed in Experiment 2, where the robot lacked eyes. Accuracy was lower on inconsistent trials overall, but this effect did not vary by perspective or robot type, indicating a general conflict cost rather than altercentric intrusion. No conclusive results could be extracted from Experiment 3 due to an insufficient sample size.

## Introduction

### Theoretical background

In social environments, being aware of the spatial viewpoints of others can help facilitate communication (Wahn and Berio, 2023). Transfer of information often involves the cognitive ability referred to in developmental psychology as visual perspective taking (VPT) (Surtees et al., 2016). VPT allows to make two different qualitative predictions: (1) inferring whether an object is visible to another agent and (2) how an object is positioned in relation to another agent's egocentric reference frame – informtaion about where the individual is located in space (Brady et al., 2024). The former, also known as Level 1 of visual perspective taking (VPT1) emerges early in childhood and is based on subject-to-object relations that lead to body-centered (self-centered) representations, while the latter – Level 2 of visual perspective taking (VPT2) is a more complex skill acquired later in life and is based on identifying locations through object-to-object relationships, independent of the observer's view (Colombo et al., 2017).

One of the most frequently applied experimental designs for investigating VPT1 is the dot-matching task developed by Samson et al., 2010, in which participants were asked to indicate whether a predefined number matched the number of dots from their own perspective or from the perspective of a human avatar. As revealed in the experiment, the avatar's perspective had influenced the participants' judgments from their own perspective. The delays or errors in self-perspective judgments caused by interference from another's point of view are called "altercentric intrusions". Conversely, participants' own perspectives also disrupted the judgment they were instructed to make from the avatar's point of view ("egocentric intrusion"), indicating bidirectional interference between self and other perspectives.

With the growing relevance of robotic technologies in social contexts, human-robot interaction has become an important focus of scientific study. Research has extended traditional paradigms of VPT beyond human agents to include non-human entities such as avatars, animated characters, and robots. Understanding whether the cognitive mechanisms underlying the human-robot interaction resemble those used in human-human or human-avatar interactions is necessary to determine how robots are perceived socially.

Initial findings suggested that VPT1 occurs only for humans, not for objects or humanoid robots (Xiao et al., 2022). However, Wahn et al., 2025 showed that humanoid robots with clearly visible features can elicit VPT1, challenging earlier conclusions. These findings have sparked interest in identifying the specific factors that lead to VPT, giving rise to competing hypotheses.

Two main hypotheses explain what triggers visual perspective taking (VPT) with non-human agents: the mere-appearance hypothesis and the mind-perception hypothesis (Wahn and Berio, 2023). The mere-appearance hypothesis states that a human-like appearance of robots alone is enough to trigger VPT, and a mental capacity to perceive the environment is not required.

In a study by Zhao and Malle, 2022, participants were shown robots differing in human-likeness, a cat, and a doll, which was described either as a robot or a mannequin depending on the condition. The aim was to examine how perceived mental capacity and human-likeness influence the activation of VPT2. In the task, altercentric intrusions occurred with humanoid robots and the doll - regardless of whether it was framed as a mannequin or a robot - but were absent with the cat and the box-shaped robot. Given the results, Zhao and Malle, 2022 predict that human-likeness increases the likelihood of VPT2.

This evidence has been supplemented by a study by Wahn et al., 2025 which tested whether the mere-appearance hypothesis applies to VPT1 by varying a humanoid robot's head (human-like vs. camera-like) and its functioning status (on vs. off). Altercentric intrusions were found in all conditions, suggesting that a human-like body alone may be sufficient to trigger VPT1. The finding that even the robot with a camera-like head elicited intrusions raises questions about whether this is due to the body's human-likeness or the camera's association with human presence.

The mind-perception hypothesis introduced by Gray and Wegner, 2012, on the other hand, argues that VPT is triggered by the attribution of mental capacities to the agent. According to this view, it is not merely appearance that matters, but whether the observer believes the agent can perceive, think, or feel. Human-likeness might facilitate this attribution, but is not itself the decisive factor. Instead, when an agent is seen to be capable of mental activity, such as seeing or knowing, observers are more likely to adopt the agent's perspective.

Most prior studies confound these two dimensions, presenting robots with visual features while simultaneously implying visual functionality, which makes it difficult to isolate the contributions of each. This ambiguity is especially relevant in the context of social robots, where appearance and perceptual capability often diverge. The current study seeks to address this gap by disentangling physical appearance from perceived functionality. The present study poses the following research question: "To what extent do visible eyes and beliefs about visual capacity influence Level-1 visual perspective taking (VPT1) in human-robot interaction?". Building on these theoretical frameworks, the study tests three specific hypotheses to examine how visual cues and beliefs about perceptual abilities shape VPT1.

## The present study

This study aims to contribute to the ongoing debate surrounding visual perspective taking (VPT) in human-robot interaction. To elucidate the respective roles of visual cues and cognitive attributions in eliciting VPT1, three experimental scenarios were designed to test the corresponding hypotheses.

In our first experiment, which is a replication of the experiment by Wahn et al., 2025 and serves as our reference point, we predict that the results will be similar to those of the original study and that altercentric intrusions will be present, as the mere-appearance hypothesis suggests.

In the second variation, where a human-like robot is present but lacks eyes, we examine the role eyes play in this process and whether participants attend to the guideline that the robot cannot see. Eyes are a salient social cue in human cognition, often interpreted as indicators of visual attention, awareness, and agency (Senju and Johnson, 2009). Their presence may be sufficient to trigger implicit perspective-taking mechanisms.Our prediction is that altercentric intrusions will be much lower or even absent, due to the fact that eyes are a crucial human feature in perspective taking, and their absence will likely impact our results. Furthermore, this would support the idea that test subjects actively assess whether the robot possesses spatial awareness based on the information that it cannot see.

If that is not the case, and our findings still show altercentric intrusions, we would conclude that a general human-like appearance is sufficient to trigger VPT1, and that eyes do not play a significant role. It would also suggest that participants did not pay attention to the guideline indicating that the robot cannot see and instead based their judgments solely on appearance. Finally, this would support the theory that humans tend to attribute spatial and cognitive awareness to intelligent agents, similar to findings from studies on human perceptions of self-driving cars (e.g., "In the Eyes of the Beheld: Do People Think That Self-Driving Cars See What Human Drivers See?" by Thellman et al., 2023).

There are two factors influencing the outcome of Experiment 3: the absence of eyes and the general guideline that the robot can see. In Experiment 3, we are testing how much attention participants give to the guideline. The robot lacks eyes but is described as being able to see and perceive. Our prediction for this experiment is that altercentric intrusions will be present, which would support the mere-appearance hypothesis. However, the comparison of results with Experiment 2 will help determine to what extent participants pay attention to the guideline and whether the decision about VPT1 is made spontaneously or after deliberate cognitive processing.

Taken together, these three experiments aim to clarify the respective roles of appearance-based and belief-based factors in triggering VPT1 in interactions with humanoid robots. By carefully manipulating the robot's visual features and the observer's beliefs about its perceptual capacities, the study seeks to provide a more nuanced understanding of the cognitive mechanisms underlying visual perspective taking with non-human agents.

## Methods

### Participants

A total of 36 participants completed the experiment (M = 28.44, SD = 8.66, 15 female, 20 male, 1 diverse). The recruitment was conducted via online convenience sampling using social media platforms, specifically WhatsApp and Facebook Messenger, as well as through word-of-mouth. Inclusion criteria included access to a computer with a stable internet connection and sufficient proficiency in the English language. One participant was excluded due to incomplete data. We obtained informed consent from all participants. The study did not require ethics approval as it followed standard procedures.

### Experimental Procedure

Participants performed the same dot-matching task as in Wahn and Berio, 2023 but with slightly different robot pictures. Participants were assigned one condition per experiment – either a left or right wall facing robot. In each trial, participants first viewed a fixation cross displayed for 500 milliseconds, followed by a blank screen for another 500 milliseconds. Then, a cue word (either *"You"* or *"Robot"*) was presented for 750 milliseconds, indicating the perspective from which the next task should be performed. After another 500-millisecond blank screen, a digit between 0 and 3 appeared for 750 milliseconds. This was followed by a picture of a room showing a robot avatar and a varying number (0–3) of red dots. The image remained on the screen until the participant responded. An example from one experimental trial can be seen in Figure 1.

The participants were instructed to judge whether the digit matched the number of visible red dots **from the specified perspective** (*"You"* or *"Robot"*). The red dots were arranged so that they either matched or mismatched the digit depending on the perspective. Importantly, the red dots seen by the participant and those visible to the robot could be the same (*consistent trials*) or different (*inconsistent trials*). Specifically, the robot could only see dots on the wall it was facing, while the participant could see dots on both walls. For example, if two red dots were placed on the wall in front of the robot, both the robot and the participant would see the same number of dots (consistent). In contrast, if one dot was on each wall, the participant would see two dots, while the robot would see only one (inconsistent).

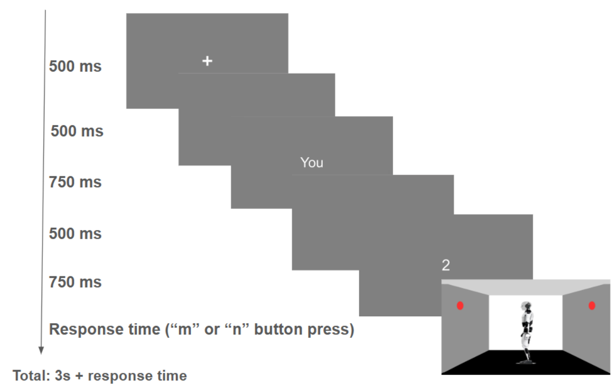


Figure 1: *One trial example from the 1st Experiment*

The participants responded by pressing the **'m'** key if the number of visible red dots corresponded to the digit, or the **'n'** key if it did not, according to the instruction perspective. They were asked to respond as quickly and accurately as possible. Each experiment consists of 160 trials. The trials were counterbalanced in both perspective and consistency. The whole experiment lasted around 20 minutes, with self-placed breaks in between trials.

### Experiment 1

18 participants (age range: 21–46 years, M = 26.24 years, SD= 5.11 years, 7 female, 10 male, 1 diverse) participated in Experiment 1. Eight in the condition with the robot facing the left wall, and ten in the condition with the robot facing the right wall. All participants were naive to the purpose of the study. The sample size was chosen based on previous studies, which reported large effect sizes for altercentric intrusion in humanoid VPT tasks. In this study, we chose the **robot condition** (Robot with or without visible eyes) as the between-subjects factor, while the within-subjects factors were **the perspective**

(self vs. robot) and t**he consistency (consistent vs. inconsistent).**

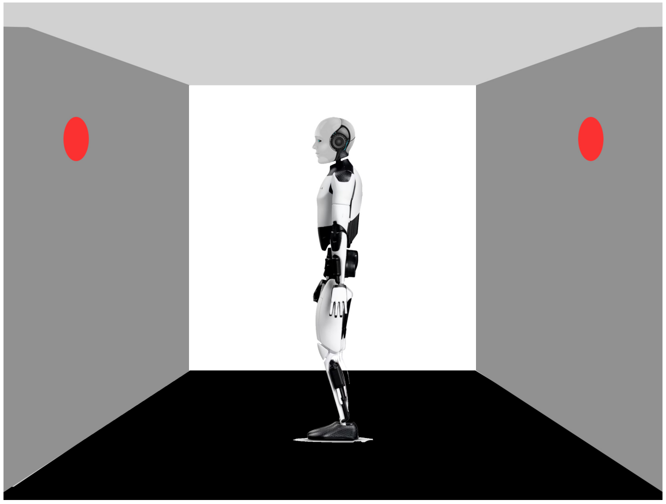

Figure 2: *Robot with visible eyes*

This experiment closely replicated the original study by Wahn and Berio, 2023, using a dot-matching paradigm designed to induce Level 1 VPT. Participants were shown images of a humanoid robot sitting at a table, facing either the same or a different direction than the participant. Crucially, the robot in this condition had clearly visible, human-like eyes, as seen in Figure 2. Dots appeared on either side of the screen and participants were instructed to indicate how many dots they could see (”Self” trials) or how many the robot could see (”Other” trials). During the experiment, participants are asked how many dots can be seen from their own perspective or from the robot's perspective, and their response time and accuracy were collected.

**Experiment 2**

A total of 18 participants (age range: 18–66 years, M = 30.65 years, SD = 11.13 years, 8 female, 10 male, 0 diverse) participated in the second experiment. Eight in the condition with the robot facing the left wall and ten in the condition with the robot facing the right wall. This second experiment differs from the first as follows: instead of using a robot with eyes, the paradigm displays a robot without eyes and with a red lamp attached on the chest. This lamp symbolizes the incapacity of the robot to see; the meaning was explained to the participants in the beginning of the experiment. Figure 3 shows the screen containing the robot included in this experiment. Otherwise, the procedure and task remained identical. The robot maintained the same posture and body orientation, just the facial area was designed without anthropomorphic eye features. This allows to isolate the role of eyes in eliciting VPT.

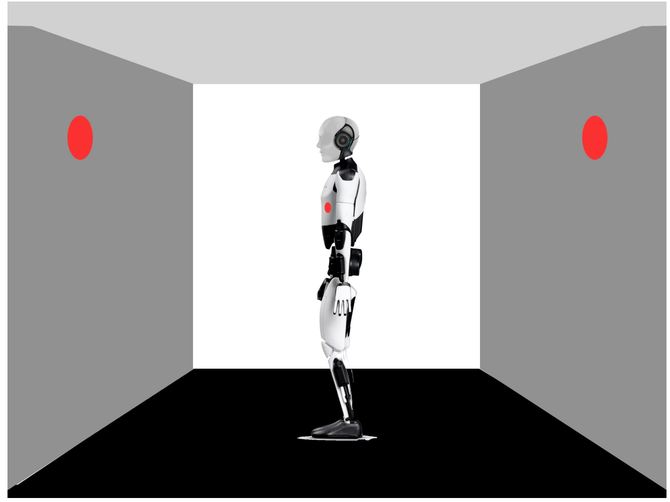

Figure 3: *Robot without eyes and a red lamp, indicating the robot can not see*

**Experiment 3**

Three participants participated in Experiment 3. One in the condition with the robot facing to the left wall and the other two, in the condition where robots were facing the right wall. The third experiment differs from the second experiment as follows: It uses a picture of a robot without eyes and a green lamp and symbolizes the fact that the robot can see, as shown in Figure 4. The robot maintained the same posture and body orientation, just the color of the light indicating on/off functionality status and ability to see changed. This allows to isolate the role of eyes in eliciting VPT. The procedure for the third experiment was the same as in previous experiments.

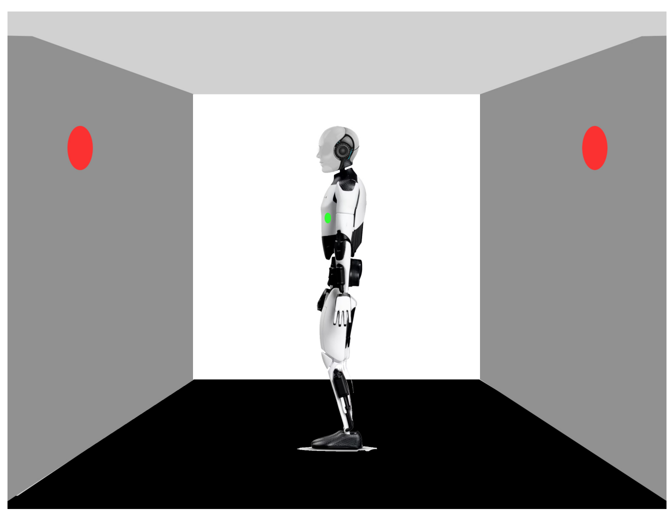

Figure 4: *Robot without eyes and a green lamp, indicating the robot can see*

## Results

The primary goal of the analysis was to test for altercentric intrusion effects – defined as the difference in response time (RT) between inconsistent and consistent trials during self-perspective judgments – under two experimental conditions: Experiment 1, where Robots have eyes (Figure 2) and Experiment 2, where Robots do not have eyes, and are described as unable to see (Figure 3). Experiment 3 was excluded from the result section due to insufficient data.

### Altercentric Intrusion

Figure 5 depicts the response time in milliseconds per experiment and perspective and shows a longer response time for inconsistent trials than consistent trial. This suggests altercentric-intrusion and is further investigated in the following: A mixed-design ANOVA (see Table 1)

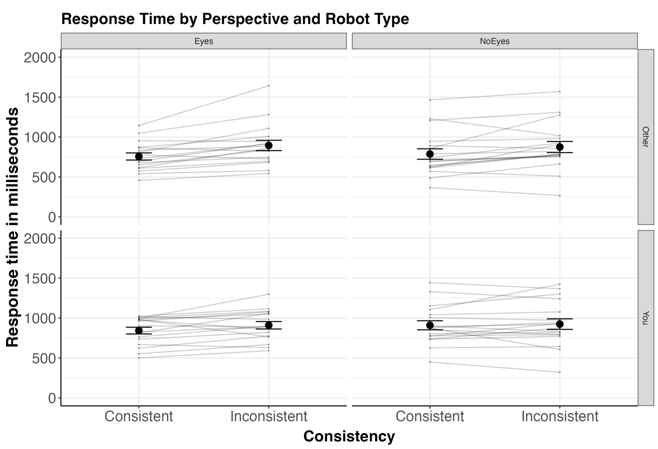


Figure 5: *Response Time in Milliseconds per Experiment and Perspective.*

with robot type (Eyes vs. No Eyes) as a between-subjects factor and consistency (consistent vs. inconsistent) and perspective (self vs. robot) as within-subjects factors revealed a significant main effect of consistency, F(1, 33) = 15.99, p < .001, and perspective, F(1, 33) = 10.22, p = .003. Importantly, there was a significant consistency × perspective interaction, F(1, 33) = 8.36, p = .007, indicating that the impact of consistency differed by perspective. No significant interactions involving robot type were found, including the three-way interaction (p = .974).

To assess altercentric intrusion directly, we have to look at self-perspective trials (see Table 2). In Experiment 1, participants were significantly slower on inconsistent trials compared to consistent trials, t(16) = –2.22, p = .041, d = –0.36, indicating clear altercentric intrusion. In contrast, in Experiment 2, there was no significant difference between inconsistent and consistent trials, t(17) = –0.47, p = .645, d = –0.05. This suggests that the removal of visible eyes from the robot eliminated the altercentric intrusion effect.

Table 1: *Mixed ANOVA Results for Reaction Time*

| **Effect** | **DFn** | **DFd** | **F** | **p** |
|---|---|---|---|---|
| R | 1 | 33 | 0.09 | .765 |
| C | 1 | 33 | 15.99 | <.001* |
| P | 1 | 33 | 10.22 | .003* |
| C × P | 1 | 33 | 8.36 | .007* |
| R × C | 1 | 33 | 1.84 | .184 |
| R × P | 1 | 33 | 0.68 | .415 |
| R × C × P | 1 | 33 | 0.00 | .974 |

R = Robot, C = Consistency, P = Perspective, * indicates $p < .05$

Table 2: *Paired t-Tests and Effect Sizes for Response Time by Experiment and Perspective*

| **Exp.** | **Persp.** | ***t*(df)** | *p* | *d* | **CI** |
|---|---|---|---|---|---|
| 1 | Self | –2.22 (16) | .041* | –0.36 | [–0.70, –0.02] |
| | Robot | –4.10 (16) | .001* | –0.50 | [–0.77, –0.24] |
| 2 | Self | –0.47 (17) | .645 | –0.05 | [–0.28, 0.18] |
| | Robot | –2.57 (17) | .020* | –0.30 | [–0.55, –0.06] |

*d* = Cohen's d, CI = 95% Confidence Interval. * indicates $p < .05$.

These findings were supported by a within-subjects ANOVA in each experiment, which can be found in Table 3. In Experiment 1, there was a significant main effect of consistency, F(1, 16) = 15.75, p = .001, and a marginal consistency × perspective interaction, F(1, 16) = 3.62, p = .075 . In Experiment 2, the consistency × perspective interaction reached significance, F(1, 17) = 4.83, p = .042, although follow-up analyses showed this effect was driven by the robot perspective, not the self-perspective.

Table 3: *Within-Subjects ANOVA Results by Experiment for Reaction Time*

| **Exp.** | **Effect** | **DFn** | **DFd** | **F** | **p** |
|---|---|---|---|---|---|
| 1 | C | 1 | 16 | 15.75 | .001* |
| | P | 1 | 16 | 1.69 | .212 |
| | C × P | 1 | 16 | 3.62 | .075 |
| 2 | C | 1 | 17 | 3.37 | .084 |
| | P | 1 | 17 | 18.33 | <.001* |
| | C × P | 1 | 17 | 4.83 | .042* |

C = Consistency, P = Perspective, * indicates $p < .05$

### Egocentric Intrusion

While altercentric intrusion was our primary focus, we also examined whether egocentric interference (i.e., self-perspective interfering with robot judgments) was present (see Table 2). In Experiment 1, participants were slower on inconsistent than consistent trials in the other-perspective condition, t(16) = –4.10, p = .001, d = –0.50. In experiment 2, a similar but weaker pattern was observed, t(17) = –2.57, p = .020, d = –0.30.

### Accuracy

Figure 6 shows a generally high accuracy, while being a bit lower for inconsistent trials. To assess whether trial consistency also affected accuracy, a mixed-design ANOVA was conducted with robot type (Eyes vs. No Eyes) as a between-subjects factor and consistency (consistent vs. inconsistent) and perspective (self vs. robot) as within-subjects factors (see Table 4).

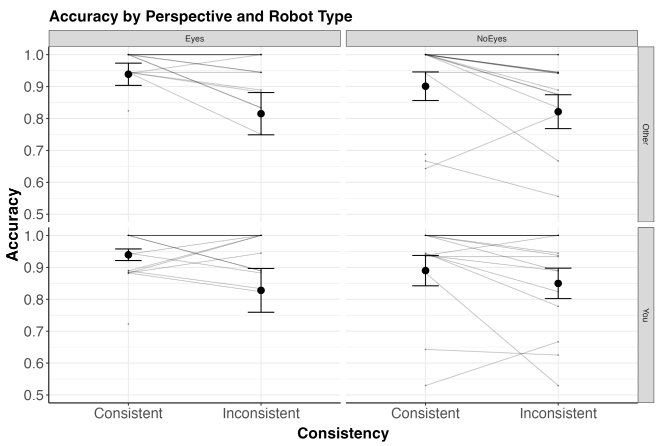


Figure 6: *Accuracy per Experiment and Perspective.*

The analysis revealed a significant main effect of consistency, F(1, 33) = 9.00, p = .005, generalized eta squared = .045, indicating that participants made more errors on inconsistent trials, regardless of perspective or robot type. There were no main effects of Robot, F(1, 33) = 0.06, p = .815, or Perspective, F(1, 33) = 0.69, p = .411, and no significant two-way or three-way interactions (ps >.22), suggesting that the accuracy cost of inconsistency was stable across conditions.

Table 4: *Mixed ANOVA Results for Accuracy*

| **Effect** | **DFn** | **DFd** | **F** | **p** |
|---|---|---|---|---|
| R | 1 | 33 | 0.06 | .815 |
| C | 1 | 33 | 9.00 | .005* |
| P | 1 | 33 | 0.69 | .411 |
| C × P | 1 | 33 | 1.51 | .228 |
| R × C | 1 | 33 | 0.96 | .334 |
| R × P | 1 | 33 | 0.01 | .920 |
| R × C × P | 1 | 33 | 0.41 | .527 |

R = Robot, C = Consistency, P = Perspective, * indicates $p < .05$

To further explore this consistency effect, separate within-subjects ANOVAs were conducted for each robot type: In Experiment 1, there was a marginal effect of consistency, F(1, 16) = 4.36, p = .053, indicating a trend toward reduced accuracy on inconsistent trials compared to consistent trials. There was no effect of perspective, F(1, 16) = 0.25, p = .626, and no significant consistency × perspective interaction, F(1, 16) = 0.12, p = .734. This suggests that while there may be a general performance cost in inconsistent trials, accuracy was not strongly modulated by whether participants were taking their own perspective or the robot's.

In Experiment 2, the effect of consistency on accuracy was statistically significant, F(1, 17) = 8.40, p = .010, showing lower accuracy on inconsistent trials. Similar to Experiment 1, neither perspective (p = .506) nor the consistency × perspective interaction (p = .131) reached significance. These results indicate that while inconsistency impaired accuracy, this impairment was not selective to the self-perspective and therefore does not provide direct evidence of altercentric intrusion in terms of accuracy.

Table 5: *Within-Subjects ANOVA Results by Experiments for Accuracy*

| **Exp.** | **Effect** | **DFn** | **DFd** | **F** | **p** |
|---|---|---|---|---|---|
| 1 | C | 1 | 16 | 4.36 | .053 |
| | P | 1 | 16 | 0.25 | .626 |
| | C × P | 1 | 16 | 0.12 | .734 |
| 2 | C | 1 | 17 | 8.40 | .010* |
| | P | 1 | 17 | 0.46 | .506 |
| | C × P | 1 | 17 | 2.52 | .131 |

C = Consistency, P = Perspective, * indicates $p < .05$

## Discussion

The present study aimed to investigate the role of physical appearance – specifically visible eyes – and belief in visual capacity in eliciting Level 1 of visual perspective taking (VPT1) toward humanoid robots. Across three experiments using a dot-matching paradigm, we tested the predictions of two competing frameworks: the mere-appearance hypothesis and the mind perception hypothesis. Our findings offer valuable insights into how humans engage in spontaneous perspective taking with robotic agents and shed light on the minimal cues required to trigger such processes.

In Experiment 1 ($N$ = 18), participants saw a robot with clearly visible eyes. We found a significant altercentric intrusion effect, with slower response times in inconsistent trials compared to consistent ones, although participants were instructed to respond from their own perspective. This replicates previous findings and suggests that people automatically compute the visual perspective of another agent when that agent looks human (Samson et al., 2010; Wahn et al., 2025). The eyes probably acted as a strong social cue, triggering implicit perspective-taking processes, which are usually fast and hard to suppress.

In Experiment 2 ($N = 18$), the robot's eyes were removed and the participants were told that it could not see – this was also signaled by a red lamp. Under these conditions, the altercentric intrusion effect disappeared. This suggests that eyes may not just support perspective taking, they might actually be necessary for it to occur in this context. The result supports previous work showing that visible eyes play a unique role in how people attribute awareness or agency to non-human entities (Wahn and Berio, 2023; Zhao and Malle, 2022). It is also possible that the combination of visual absence and explicit instruction (that is, the robot 'can't see') made it easier for participants to suppress any default assumptions about the robot having a perspective.

In Experiment 3, a new belief was introduced: the robot still had no eyes, but we told participants that it could see (signaled with a green lamp). Only three people completed this condition, so the data was not usable. Thus it remains uncertain whether belief alone, with disregard to visual cues, is sufficient to elicit perspective taking. Future studies could examine this question with a complete sample, to clarify the extent to which belief-based cues must be strengthened to override conflicting visual information.

In Experiments 1 and 2, the accuracy was generally high but slightly lower in inconsistent trials. Since this effect did not vary by condition, it probably reflects a general conflict or interference cost rather than perspective-taking specifically. Reaction time turned out to be a more sensitive and useful measure for capturing VPT1-related interference in this task, consistent with earlier research.

Although the results mostly support the mere-appearance hypothesis, there are some important limitations. The biggest one is that Experiment 3 was underpowered, so no claims can be made about how belief affects VPT1 without eyes. Another limitation of the study is the reliance on a single task paradigm and a specific humanoid robot model. Although the dot-matching task has strong internal validity, its ecological validity may be limited. Moreover, the belief manipulation was delivered through instruction and was not reinforced dynamically during the interaction, which may have reduced its impact. Future work should consider using interactive scenarios or VR-based manipulations to more fully test the influence of beliefs and engagement.

Despite these limits, the study contributes to the understanding of how people mentally engage with robots and gives some insight into human-robot interactions. The findings support the idea that visual features like eyes are not just decorative, but actually affect how humans process a robot as a social agent. This has practical implications: If designers want robots to be intuitive partners in educational or caregiving contexts, then giving them eyes or at least features that clearly signal visual access might improve interaction quality. In short, making a robot "look like it can see" might help humans automatically consider what the robot is seeing, which could support better communication and coordination.

## Conclusion

Taken together, the results of this study provide compelling evidence that visual morphology, specifically the presence of visible, human-like eyes, is a critical factor in eliciting spontaneous visual perspective taking (VPT1) toward humanoid robots. Across two adequately powered experiments using a dot-matching paradigm, we found that altercentric intrusion effects emerged only when the robot possessed visible human-like eyes, suggesting that such features serve as strong perceptual cues for social agency. When the robot lacked the eyes, the intrusion effect was absent, indicating that just anthopomorphic bodily appearance may not be sufficient to trigger VPT1 in this context. However, since this conclusion is based on a null effect, it should be interpreted with caution and tested further in future studies. These findings support the mere-appearance hypothesis (Wahn and Berio, 2023; Wahn et al., 2025) and align with theories emphasizing the importance of low-level visual features in social cognitive processes (Samson et al., 2010; Wahn and Berio, 2023). While Experiment 3 was inconclusive due to low sample size, future work should aim to disentangle the relative contributions of appearance and belief using stronger manipulations and larger samples. These insights have important implications for both theories of social cognition and the design of socially interactive robots.

## References

Brady, N., Leonard, S., & Choisdealbha, Á. N. (2024). Visual perspective taking and action understanding. *Acta Psychologica*, *249*, 104467. https://doi.org/10.1016/j.actpsy.2024.104467

Colombo, D., Serino, S., Tuena, C., Pedroli, E., Dakanalis, A., Cipresso, P., & Riva, G. (2017). Egocentric and allocentric spatial reference frames in aging: A systematic review. *Neuroscience & Biobehavioral Reviews*, *80*, 605–621. https://doi.org/10.1016/j.neubiorev.2017.07.012

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Scott, S. E. B. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134. https://doi.org/10.1016/j.tics.2008.11.009

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52. https://doi.org/10.1016/j.cognition.2016.01.014

Thellman, S., Pettersson, M., Holmgren, A., & Ziemke, T. (2023). In the eyes of the beheld: Do people think that self-driving cars see what human drivers see? *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 612–616. https://doi.org/10.1145/3568294.3580158

Wahn, B., & Berio, L. (2023). The influence of robot appearance on visual perspective taking: Testing the boundaries of the mere-appearance hypothesis. *Consciousness and Cognition*, *116*, 103588. https://doi.org/10.1016/j.concog.2023.103588

Wahn, B., Berio, L., Weiss, M., & Newen, A. (2025). Try to See it My Way: Humans Take the Level-1 Visual Perspective of Humanoid Robot Avatars. *International Journal of Social Robotics*, *17*(3), 523–534. https://doi.org/10.1007/s12369-023-01036-7

Xiao, C., Fan, Y., Zhang, J., & Zhou, R. (2022). People do not automatically take the level-1 visual perspective of humanoid robot avatars. *International Journal of Social Robotics*, *14*(1), 165–176. https://doi.org/10.1007/s12369-021-00773-x

Zhao, X., & Malle, B. F. (2022). Spontaneous perspective taking toward robots: The unique impact of humanlike appearance. *Cognition*, *224*, 105076. https://doi.org/10.1016/j.cognition.2022.105076