COGNITIVE SCIENCE
A Multidisciplinary Journal

Check for updates

Cognitive Science 47 (2023) e13385

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13385

Perspective Taking Reflects Beliefs About Partner Sophistication: Modern Computer Partners Versus Basic Computer and Human Partners

Jia E. Loy,[a] Vera Demberg[a,b]

[a]*Department of Language Science and Technology, Saarland University*
[b]*Department of Computer Science, Saarland University*

Received 18 November 2022; received in revised form 22 October 2023; accepted 3 November 2023

Abstract

We investigate partner effects on spatial perspective taking behavior in listeners, comparing behavior with a human versus a computer partner (Experiments 1 and 2), and with computer partners of different perceived capabilities (Experiment 3). Participants responded to spoken instructions from their partner which could be interpreted egocentrically (from their own perspective) or othercentrically (from their partner's perspective). In contrast to earlier work, we found that participants were more egocentric with a computer than a human partner. Participants were also more egocentric with a computer partner that appeared more modern and capable, compared to one that appeared outdated and limited in ability. Our results show that perspective taking behavior is sensitive to information about one's partner; in particular, listeners consider their partner's potential ability to collaborate, adjusting their egocentric tendencies accordingly. Moreover, we highlight what appears to be a shift in listeners' expectations regarding computers' collaborative capabilities, leading to greater willingness to push the burden of perspective taking onto a computer partner.

Keywords: Perspective taking; Partner effects; Human–computer interaction

Correspondence should be sent to Vera Demberg, Department of Language Science and Technology and Department of Computer Science, Saarland University, Campus C7.2, 66123, Saarbrücken, Germany. E-mail: vera@lst.uni-saarland.de

1. Introduction

Spatial perspective taking, the ability to mentally adopt another person's point-of-view, is a process crucial to everyday cognition and communication. For instance, people frequently find themselves in situations where they have to refer to objects, give directions, or describe locations, often from another person's perspective. Research has shown that speakers are flexible in their perspective taking tendencies, with the choice of whether to take their own or their addressee's perspective dependent on a range of cognitive, social, and representation cues (Galati & Avraamides, 2013). However, selecting and maintaining a perspective is cognitively effortful (Mainwaring, Tversky, Ohgishi, & Schiano, 2003), and speakers may at times mix perspectives (Taylor & Tversky, 1996), or produce spatially ambiguous descriptions (e.g., "the mug on the left"). Such descriptions may not pose a challenge when speaker and listeners' perspectives align; however, when perspectives differ, the task of disambiguation falls to the listener, who is forced to adopt a particular perspective—either their own (egocentric) or the speaker's (othercentric) perspective—to derive a meaningful interpretation (cf. Duran, Dale, & Kreuz, 2011). How do listeners decide which perspective to take? In the current study, we examine the role of partner identity on listeners' perspective taking behavior. In particular, we compare differences between interacting with a human partner and a computer partner, as well as with computer partners which appear to have different levels of capability.

1.1. Partner effects on spatial perspective taking

Studies on perspective choice and interaction highlight the influence of various aspects of one's partner on perspective taking. For instance, speakers tend to be more egocentric when their partner's perspective is more effortful to adopt, such as with cognitively complex misalignments in perspective (e.g., oblique 135 degree vs. orthogonal 90 or 180 degree offsets; Galati & Avraamides, 2015). This is in line with a view on perspective taking which emphasizes the primacy of egocentricism, and associated cognitive cost of othercentric perspective taking (Epley, Keysar, Van Boven, & Gilovich, 2004; Ferguson, Apperly, & Cane, 2017). Further evidence for this is seen from studies that show reduced perspective taking abilities when speakers face utterance planning difficulties such as time constraints or memory demands (Horton & Gerrig, 2005; Horton & Keysar, 1996). However, speakers also appear willing to invest in this cost if sufficiently motivated, based on inferences about their communicative partner. Schober (1993) showed that speakers were more likely to adopt their partner's perspective when interacting with a real addressee than with a hypothetical addressee in a noninteractive setting, highlighting the social motivation underlying othercentric perspective taking (cf. Tversky & Hard, 2009). When actual addressees are present, inferences about their communicative ability also influence perspective taking tendencies: speakers produce more othercentric descriptions when they perceive their partner to be constrained in some way, such as by being unable to provide feedback (Shelton & McNamara, 2004), or having poorer spatial abilities (Schober, 2009). These results are attributed to the principle of least collaborative effort (Clark & Wilkes-Gibbs, 1986), which posits that conversation partners are jointly responsible for ensuring mutual understanding through adapting their perspectives to minimize collective effort; accordingly, when one party is perceived to be hampered in their ability to contribute to the task, the other may expend greater effort to maintain shared understanding by taking an othercentric perspective (cf. Galati & Avraamides, 2015). Thus, conversation participants are prompted to tailor their utterances in a way such that they are easily understood by their addressee, a process known as audience design (Clark, 1996). Comparable adaptation strategies have also been observed in nonspatial communication tasks, whereby speakers appear sensitive to the types of knowledge their partners may have (Horton & Gerrig, 2002; Isaacs & Clark, 1987), and invest extra effort into making information more accessible for partners with potentially limited ability, such as addressees with less expertise (Fussell & Krauss, 1989, 1992), child addressees (Newman-Norlund et al., 2009), or "basic" computer systems (Branigan, Pickering, Pearson, McLean, & Brown, 2011).

Such flexible perspective taking is not limited to production but extends to comprehension as well. Duran et al. (2011) highlight the role of partner identity on listeners' interpretations of an ambiguous spatial description. Their study investigated spatial perspective taking using a computerized task in which participants responded to recorded instructions from a partner requesting for one of two identical objects on a tabletop. The participant and partner's locations around the table changed from trial to trial, and the partner's request made use of potentially ambiguous spatial terms *right, left, front*, or *back* (e.g., "give me the folder on the right"). Thus, on trials where perspectives differed (e.g., when the partner was seated 180 degrees across the table from the participant), participants could select the object based on an egocentric or othercentric perspective.

Duran et al.'s study revealed several findings of interest. First, response times on othercentric responses increased with increasing misalignment between the participant and partner's perspectives ($0<90<180$ degrees), highlighting the cognitive cost of perspective taking. This is in line with previous work on perspective rotation with orthogonal offsets (Michelon & Zacks, 2006; see Galati, Diavastou, & Avraamides, 2018 for evidence of additional costs associated with oblique offsets). Second, an additional analysis comparing the subset of right/left to front/back trials revealed that the processing cost effect lay largely in right/left trials. This suggests that the cognitive cost of perspective taking is not equal across directional axes: in particular, right/left appears to be more costly than front/back perspective taking. This finding is echoed by other studies on perspective taking (e.g., Mainwaring et al., 2003), and aligns with the broader literature on spatial encoding which shows that the right/left axis, which has greater body axis symmetry, is more difficult than front/back (Bryant, Tversky, & Franklin, 1992; Clark, 1973; Franklin & Tversky, 1990). Third, and most importantly, the amount of othercentricism exhibited by participants was sensitive to information about partner identity. By classifying participants as "egocentric" or "othercentric" based on their dominant mode of response, Duran et al. found that the proportion of othercentric responders was greater when participants were told their partner was a simulated computer than a real human. The authors interpreted this as a demonstration of the principle of least collaborative effort in that the motivation to take on a partner's perspective increases when the partner is perceived as being less capable of collaboration, such as with a computer system unable to share the burden of ensuring mutual understanding.

1.2. Social perception of computers and technology

While social effects on perspective taking have been tested extensively in the context of human–human communication, the topic has received less attention in the field of human–computer interaction. Human users are, however, known to apply social conventions and expectations to their interactions with computers much as they would with human interlocutors, in particular when computers exhibit human–like attributes, such as interactivity and spoken language capabilities (Nass, Steuer, & Tauber, 1994; Nass & Moon, 2000). Starting from the nineties, within the Computers-Are-Social-Actors (CASA) paradigm, Nass and colleagues have emphasized that the human–computer relationship is fundamentally social: for instance, people are known to exhibit social behavior such as politeness norms and personality endowment with computers in interactive scenarios (Moon & Nass, 1996; Nass, Moon, & Carney, 1999). Social cues in the form of speech, appearance, or behavior presented by computers also modulate human behavior, for instance, in the way gender stereotypes are triggered by computers with gendered-speech (Lee, Nass, & Brave, 2000). Crucially, people appear to attribute different degrees of competency to technology based on such social cues: much in the way people might consider the work of a specialist to be of better quality than a nonspecialist, viewers in an experiment-rated programs presented on "specialist" television sets more positively and higher in quality than the same programs presented on "generalist" television sets (Leshner, Reeves, & Nass, 1998).

The CASA framework demonstrates that human users extend social and interactional conventions to human–computer interaction across a wide spectrum of phenomena. These findings highlight a broader point on the perception of artificial intelligence (AI) technology: that of a degree of readiness to establish common ground and cooperation with computer partners and technology in general. Importantly, this notion has received growing attention in the last decade or two, in part from rapid advancements in the fields of AI and human–machine interaction (e.g., Ajoudani et al., 2018; Sheridan, 2016), leading to increasing integration of collaborative robot systems in everyday life. Farooq and Grudin's (2017) historical summary of the development of human–computer interaction describes a paradigm shift at the industrial level from positioning humans as operators and computers as tools, to a current landscape of shared-agency and cooperation. This change comes as a consequence of technological interfaces exhibiting increasing cognitive abilities, such as situational awareness, dynamic adaptation, and decision-making (Richert, 2018; Romero, Bernus, Noran, Stahre, & Fast-Berglund, 2016). Frameworks such as human–autonomy teams (Lyons, Sycara, Lewis, & Capiola, 2021) propose the role of machines as support within human teams, identifying factors such as the perception of machines as intelligent, autonomous teammates rather than automated tools as a rationale for the shift. At the societal level, we also see a movement toward increasingly positive attitudes toward robots and AI, even in areas that traditionally prioritize human–human interaction, such as the hospitality and service industries (Huang, Chen, Huang, Kong, & Li, 2021; Hou, Zhang, & Li, 2021; Kim, Kim, Badu-Baiden, Giroux, & Choi, 2021). In short, developments in human–computer interaction and collaboration have engendered changing attitudes toward AI, strengthening views on machines as increasingly viable partners in social situations and cognitive tasks.

1.3. Changes in the perception of computers as communicative partners

One domain in which the change in perception toward the collaborative abilities of computers may be observed is that of perspective taking. Duran et al.'s (2011) partner effects, in which participants were more likely to take the perspective of a supposed computer partner, are consistent with results from earlier production studies suggesting that speakers view robot addressees as having reduced perspective taking capabilities. For instance, when issuing directions to a robot to move goal objects about a floor space, participants were found to almost exclusively use othercentric instructions, occasionally even standing up from their chair to reorient themselves with the robot's spatial perspective (Fischer, 2006). In a similar vein, Moratz, Fischer, and Tenbrink (2001) observed that participants directing a robot frequently used explicit instructions that traced a path from the robot's perspective (e.g., "drive 1 meter ahead, roll a bit forward, drive north from your point of view") rather than straightforward, goal-naming strategies (e.g., "drive up to the right cube"). The perceived linguistic competence of the robot also appears to matter: with a nonverbal robot, participants almost always took the robot's perspective; a linguistically skilled robot capable of producing verbal output, however, elicited a wider range of spatial expressions, notably with the inclusion of egocentric descriptions as well as more varied and complex syntax (Fischer, 2005). These results suggest that speakers conceptualized robots as artificial communication partners with poorer ability to collaborate at the time of the studies, reflecting an underlying view that spatial cognition and perspective taking posed considerable challenges for AI (cf. Marin-Urias, Sisbot, & Alami, 2008).

However, recent studies investigating spatial perspective taking in human–robot interaction present a more varied picture with respect to speakers' perspective choice. Consistent with earlier results, some studies find that speakers produce more othercentric descriptions when interacting with robots compared to humans (e.g., Li, Scalise, Admoni, Rosenthal, & Srinivasa, 2016). On the other hand, others find no difference between human and robot addressees (e.g., Xiao, Xu, Sui, & Zhou, 2021), or even the opposite result with speakers more likely to take the perspective of humans compared to robots (e.g., Carlson, Skubic, Miller, Huo, & Alexenko, 2014; Li et al., 2016; Zhao, Cusimano, & Malle, 2016). For instance, Carlson et al. (2014) manipulated whether participants who had to describe the location of an object in a virtual environment addressed a human avatar or a robot avatar. They found that speakers produced more othercentric descriptions when addressing a human avatar, and conversely more egocentric descriptions with a robot avatar. Instructions in the two conditions were also qualitatively different: with the robot avatar, speakers used fewer words overall, and more static descriptions defining the object's location in relation to the environment (e.g., "the cell phone is on the table by the bed"); with the human avatar, speakers used more dynamic descriptions which resembled step-by-step directions guiding the listener through the environment (e.g., "go forward, turn right…"). The authors attribute these differences to the differential assumptions that speakers make about their addressee's capabilities, resulting in a preference for more streamlined communication, and consequently less accommodation with a robot addressee.

The apparent shift in perspective taking tendencies over time could reflect a more general change in the perception of and expectations toward robots as interactive agents. Recent, rapid progress in the field of AI has led to considerable developments in the design, functionality, and complexity of intelligent systems, which have gone beyond simply providing an automated interface to more advanced behavior, such as multimodal communication (Mazhar, Navarro, Ramdani, Passama, & Cherubini, 2019; Yohanan & MacLean, 2012) or incorporating virtual reality (Yu, Hsueh, Sun, & Liu, 2021). Human user behavior, in turn, is observed to be sensitive to the apparent capability of AI systems, such as the degree of autonomy they exhibit (Brügger, Richter, & Fabrikant, 2019; Taylor, Wang, & Jeon, 2023). Evidence also shows that social robots are now comparable to, and in some cases more effective than, humans in various learning and communicative situations (e.g., Beattie, Edwards, & Edwards, 2020; Edwards, Edwards, Spence, Harris, & Gambino, 2016; Kim et al., 2013). Crucially, these developments characterize the interactional aspect of AI capabilities, signaling a shift from earlier work which conceptualizes robots as socially limited communication partners (e.g., Tenbrink, Fischer, & Moratz, 2002). Recent research investigating the expectations of robots in daily life reveals that in addition to household chores and work-related tasks, some people expect to use robots for company or entertainment, or in other sociointeractional capacities (e.g., by reading a story to help them fall asleep; Horstmann & Krämer, 2019). In addition, participants' knowledge of and contact with robots in the past was found to be positively related to their expectations regarding robot abilities and their role in daily life. This signals how technological development in AI has come to shape peoples' expectations, either through their own experience with or via media exposure to information about robots.

Given this landscape of technological advancement, it is perhaps less surprising that robot partners now elicit less othercentric perspective taking in interactive tasks, as expectations regarding their spatial and communicative abilities have increased, in turn increasing people's willingness to shift the burden of perspective taking onto them. However, results thus far are largely limited to studies investigating the production of spatial descriptions directed at robots; less is known about the comprehension of such expressions produced by a robot. In other words, if people expect robots to have a better ability to understand spatial expressions from another perspective, would they have similar expectations regarding a robot partner's ability to produce such expressions? Duran et al. (2011) found that listeners had lower expectations regarding the perspective taking abilities of a partner they believed to be a computer than a human, as evidenced by higher rates of othercentric responding with the computer partner (i.e., participants interpreted utterances from the computer's spatial perspective). However, it is of interest whether this pattern of results would still hold in light of recent developments in AI and the corresponding shift in language production toward robots.

2. Current study

Across three experiments, we investigate the role of partner effects on spatial perspective taking in listeners. Experiments 1 and 2 provide a conceptual replication of Duran et al.'s study, using ambiguous spatial descriptions which could be interpreted egocentrically or othercentrically, and comparing listeners' perspective taking behavior with a human partner and a computer partner. In line with our hypothesis about a "change in technological expectations", we find in both experiments the opposite pattern of results to Duran et al. (2011) such that our listeners are more egocentric with a computer partner than a human partner. To further investigate this, in Experiment 3, we manipulate the perceived capability of the computer partner with a computer partner who appears to be modern and advanced and one who appears to be old-fashioned and basic. Here, we find that listeners are more egocentric with the "modern" compared to the "basic" computer partner. In addition, the first two experiments assess right/left (Experiment 1) and front/back (Experiment 2) descriptions separately. Thus, while not an original objective of the study, this allows us to make a secondary contribution on perspective taking behavior in each dimension without the influence of the other. A comparison of data from the two experiments shows that listeners are more egocentric in the right/left compared to the front/back dimension (Section 4.2.5). We discuss our results in the three experiments with respect to theories of perspective taking and the changing expectations in human–computer interaction.

3. Experiment 1

Experiment 1 investigated the role of partner identity (human vs. computer) in listeners' spatial perspective taking behavior. We employed a collaborative task modeled on Duran et al.'s (2011) study. Participants heard prerecorded instructions from their partner requesting for an object on a table top in a web-based virtual environment. We varied the location of the partner's avatar such that on some trials, the participant and partner were on the same side of the table (thus having the *same perspective*), and on other trials, they were on opposite sides (thus having *different perspectives*). In addition, we manipulated partner identity between-subjects: Participants were led to believe they were interacting with either another human or a simulated computer. On critical trials, the partner asked for one of two identical objects on the table using the spatial terms *left* or *right* (e.g., "Give me the stapler on the right"). We focused on right/left descriptions here since larger effects were reported for these over front/back descriptions in Duran et al. (2011). Thus, on different perspective trials, the partner's request was spatially ambiguous, and could be interpreted egocentrically (the participant's right in this example) or othercentrically (the partner's right).[1]

3.1. Method

3.1.1. Participants

Five hundred and twenty-four participants were recruited on Amazon Mechanical Turk (AMT).[2] Data collection for the study took place in June 2021. We used AMT filters to restrict recruitment to U.S.-based participants with a minimum of 1000 approved HITs and a 97% approval rating. For the analyses, we excluded data from participants who

(a) were non-native speakers of English (4; determined via self-report in the post-test questionnaire),

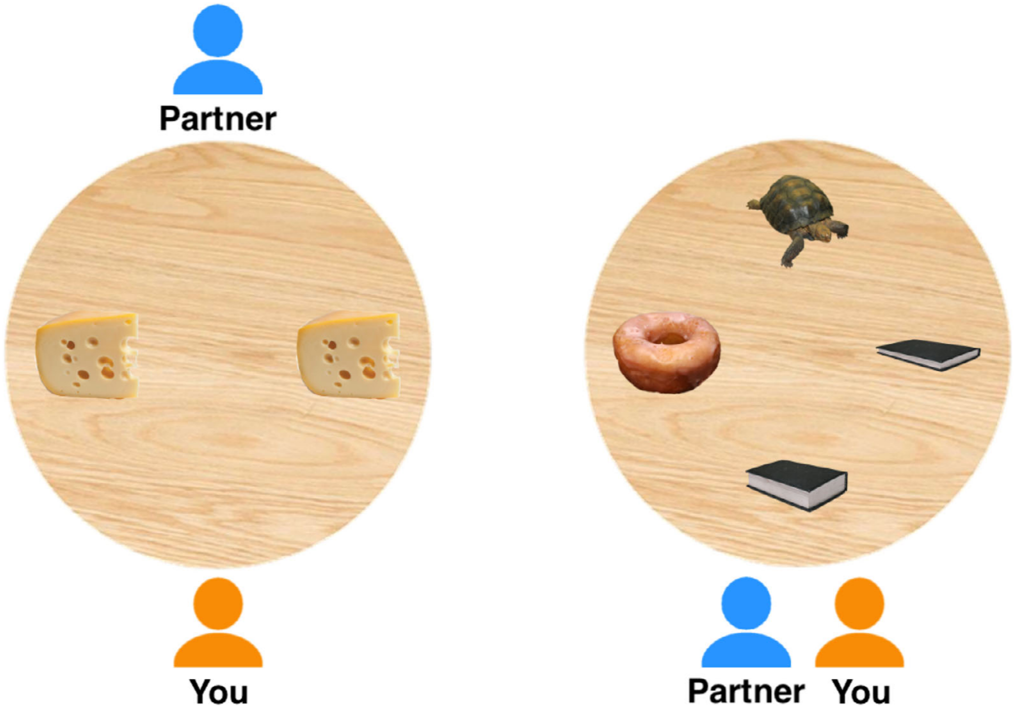


Fig. 1. Example of a critical trial in the different perspective condition (left) and a filler trial (right).

(b) failed to meet a minimum accuracy threshold of 80% on all trials except for those in the different perspective condition in which the partner's utterance was ambiguous (71), or
(c) indicated in the post-test questionnaire suspicion about the authenticity of their partner or the interaction (175).

Thus, the final dataset consisted of 274 participants (mean age = 36.1, range = 20–68), with 134 and 140 participants in the human and computer partner conditions respectively.

3.1.2. Materials and design

The experiment consisted of 12 critical and 36 filler trials. Each trial presented a number of objects (two, three, or four) arranged on a round table top viewed from above. Objects were located in one of four predetermined positions: top, left, bottom, or right of the center of the table. Each trial also featured two avatars representing the participant and their partner (orange and blue, respectively). The participant was always located at the bottom of the table; we manipulated whether the partner was located next to the participant (same perspective condition) or across the table from the participant (different perspective condition). Fig. 1 shows an example of a trial from the experiment.

Objects used in the experiment were images of everyday objects (e.g., clock, stapler, potato) taken from the Bank of Standardized Stimuli (BOSS; Brodeur, Dionne-Dostie, Montreuil, &

Table 1
Breakdown of display types in the experiment

Trial type	Display type	No. of objects	No. of trials
critical	same perspective	2	6
	different perspective	2	6
filler	color contrast	2/3/4	8
	size contrast	2/3/4	8
	location contrast	3/4	8
	no contrast	2/3/4	12

Lepage, 2010). In cases where no suitable images were available on BOSS, a free alternative was sourced from Google Image. A total of 64 images were used in the experiment: 32 unique objects and 16 contrast pairs (e.g., red/green apple). Contrast pairs were created by editing the original image to produce two versions of the object that differed only in the relevant contrast property, and were featured in filler trials (see below).

Critical displays always featured two identical objects located in the left and right positions on the table. Half of the critical trials used the same perspective seating configuration and the other half used the different perspective configuration. The partner's request was always of the form "Give me the <object> on the right/left," with an equal number of left and right requests in each perspective condition.

To reduce the salience of critical displays, we included filler trials where the display varied the total number of objects presented (two, three, or four), and the type of contrast relevant for referent identification. Four types of contrast were used: color (e.g., red/green apple); size (e.g., long/short ruler); location (right/left ambiguity as in critical trials, but with the addition of one or two distractors); and no contrast (identifiable by the bare noun alone, e.g., "clock," with unrelated distractors). A higher proportion of three- and four-object filler displays were included to ensure participants saw roughly the same number of two-, three-, and four-object displays across the experiment. Table 1 provides a breakdown of the variation in display types used in the experiment.

Distractor objects on filler trials were chosen randomly from the full set of images with the constraints that

(a) any relevant contrast requirements were fulfilled, and
(b) no objects were repeated in the display.

In color contrast, size contrast, and no contrast displays, the position of objects was randomized and the partner's instruction unambiguously identified a single referent (e.g., "Give me the red apple," or "Give me the clock"). Half of the filler trials used the same perspective seating configuration and the other half the different perspective configuration.

Participants were randomly assigned to the human partner or the computer partner condition. In the human partner condition, participants were told they were paired with another worker on AMT; in the computer partner condition, they were told their partner was a simulated computer program. All participants were told the interaction was taking place in real-time. In fact, partner utterances were prerecorded and we simulated live interaction with waiting times (for a partner to sign up) and variable utterance onset latencies. The human partner instructions were recorded by a female native speaker of North American English. The computer partner instructions were synthesized with the Apple Macintosh built-in text-to-speech function ("Agnes" voice). All recordings were normalized to have the same mean intensity.

3.1.3. Procedure

Participants accessed the experiment online via the AMT website. The experiment was described as an activity in which two users carried out a joint task in a shared workspace; the participant's task was to move objects about the workspace in response to spoken instructions from their partner. Depending on their assigned condition, participants were told they would be partnered with another worker on AMT (human partner condition) or a simulated computer program (computer partner condition). The instructions emphasized that the online interface was still in its development stages, hence audio streaming only worked one-way (i.e., participants could hear but could not speak to their partner). Following this, participants in the computer partner condition were taken to the audio check phase. In the human partner condition, participants saw a page with a message informing them to wait for another worker to sign up to the task. The message, along with dynamic loading dots, remained on screen for 15 s. This was then replaced by a message informing them that they had been partnered with another worker, before the experiment progressed to the audio check phase.

The audio check phase served to ensure participants had audio turned on and volume adjusted to a suitable level. Participants followed their partner's instruction (ostensibly streamed in real-time) to click on a target image out of an array of four images. After selecting the correct image, the experiment began. Participants who selected the wrong image more than once were prevented from continuing with the experiment.

During the experiment, participants saw a display consisting of a table top viewed from above and two avatars representing the participant and their partner. On each trial, the display featured a number of objects arranged on the table top. Following a short delay simulating utterance planning time, participants heard an instruction from their partner requesting for one of the objects. The utterance initiation delay was fixed at 2500 ms for the first three trials (always filler trials), 1200 ms on critical trials, and variable between 800 and 2000 ms on all other trials (with lower probability assigned to larger values). In the human partner condition, the first two instructions included disfluencies and nonpropositional discourse markers to more closely mimic naturalistic spoken behavior (*"Um, ok, give me the…", "Ok, now give me the…"*). Participants manipulated objects by clicking on and dragging them over to their partner's avatar. Objects were not movable until the partner's utterance had finished playing. Once an object had been "given" to the partner, the objects disappeared and were replaced by the objects for the next trial. A progress bar at the top of the screen indicated how many trials the participant had completed. Trial order was randomized for each participant with the constraints that the experiment began with at least three filler trials, and critical trials were separated by at least one filler trial.

After the task, participants completed a post-test questionnaire collecting information about their age, gender, native language, and experience doing the task. This was followed by a series of questions aimed at verifying whether participants in the human partner condition suspected the authenticity of their partner or the interaction (as in Duran et al., 2011). These included an open-ended question about their impression of their partner ("How did you find your interaction with your partner?"), and two yes/no response questions ("Did you feel you were connected to an actual person?" and "Did you believe you were connected to an actual person?"). Each question appeared on a new page, with no option of returning to the previous page. Participants whose answer to the first question implied any suspicion about their partner or who indicated "no" to the second or third questions were excluded from the planned analyses.

3.2. Results

3.2.1. Analyses

The experiment yielded two main measures: the object selected by the participant and their response time on each trial. Object selection was coded as whether or not participants interpreted their partner's utterance egocentrically (i.e., "left" or "right" from the perspective of the participant's avatar). Response times were measured from the offset of the partner's utterance on each trial. In addition to participants who did not meet our inclusion criteria, we excluded individual data points that were over three standard deviations above the condition mean.

Following Duran et al. (2011), we categorized participants as "egocentric" or "othercentric" based on their dominant mode of response (defined as over 70% of trials) on different perspective trials—in our dataset, this corresponded to at least five out of six different perspective trials. Participants who were neither were categorized as "mixed." Statistical analyses were carried in R version 4.1.0 (R Core Team, 2021). We conducted two planned analyses. The first examined perspective taking tendencies based on the object selected by participants on each trial. We used logistic mixed effects regression to model the dependent variable of whether participants chose the object from their own avatar's perspective or not on each trial. The second analysis examined processing cost based on participants' trial response times. We used linear mixed effects regression to model the log-transformed response time (in ms) on each trial. Both models included perspective (sum-coded; –0.5 and +0.5 for "same" and "different" respectively) and partner (sum-coded; –0.5 and +0.5 for "human" and "computer" respectively) as fixed effects, and by-participant and by-item random intercepts and by-participant random slopes for perspective. *p*-values for linear mixed effects models were calculated using Satterthwaite approximations for degrees of freedom implemented in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). For all models, we report any significant effects in the text; full model outputs are provided in Appendix A in the Supplementary Material.

3.2.2. Distribution of responders

Table 2 shows the breakdown of egocentric, othercentric, and mixed responders in the human and computer partner conditions. Following Duran et al. (2011), we examined

Table 2
Breakdown of responder types in Experiments 1–3

Experiment	Partner	Egocentric	Othercentric	Mixed
1	Human	108 (81%)	17 (13%)	9 (6%)
	Computer	125 (89%)	13 (9%)	2 (1%)
2	Human	8 (8%)	85 (86%)	6 (6%)
	Computer	38 (31%)	77 (64%)	6 (5%)
3	"Basic" computer	53 (38%)	30 (22%)	55 (40%)
	"Modern" computer	88 (56%)	19 (12%)	39 (27%)

Note: Percentages are calculated out of the total number of participants in each partner condition.

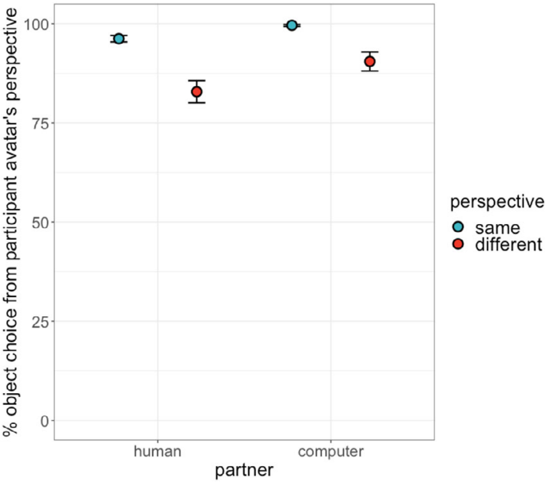
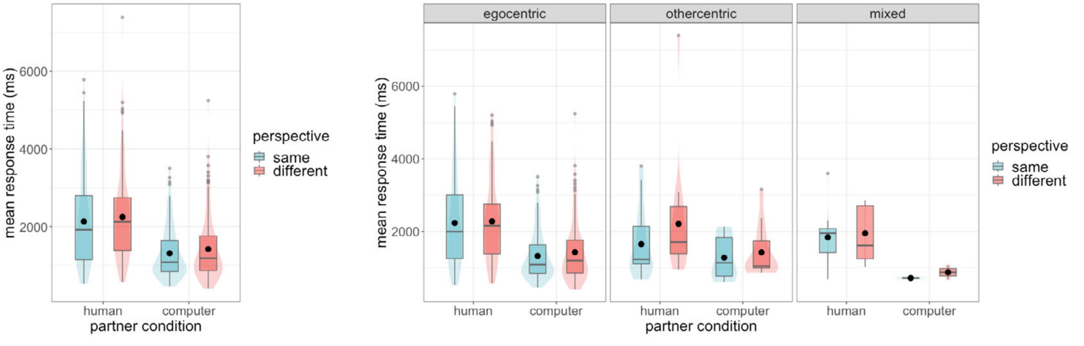


Fig. 2. Experiment 1: Percentage of trials on which participants chose the object from their own avatar's perspective (on same perspective trials, this was a shared perspective with the partner's avatar; on different perspective trials, this was across the table from the partner's avatar and, therefore, reflects egocentric perspective taking by the participant). Error bars represent ±1 of by-participant means.

the rate of egocentric responders and othercentric responders comparing across the two partner conditions. For egocentric responders, there was an 8% decrease in the human compared to the computer partner condition. This difference was marginally significant, $\chi^2(1) = 3.41$, $p = .06$. For othercentric responders, there was a nonsignificant 4% increase in the human compared to the computer partner condition, $\chi^2(1) = 0.50$, $p = .5$.

3.2.3. *Perspective taking tendencies*

Fig. 2 shows the percentage of trials on which participants chose the object from their own avatar's perspective. Of interest is participants' behavior on different perspective trials, where this response reflects egocentric perspective taking by the participant. The model showed a main effect of perspective, with participants less likely to choose the own-avatar-associated object on different perspective trials, $\beta = -2.43$, $z = -2.02$, $p = .04$; and of partner, with



(a) (b)

Fig. 3. Experiment 1 mean trial response times (in ms) (a) across all participants, and (b) split by responder group. Boxplots represent the median and interquartile range; black dots represent condition means.

participants more likely to choose the own-object-associated object in the computer partner condition, $\beta = 1.69$, $z = 2.43$, $p = .02$. There was no interaction between perspective and partner condition ($p > .5$).

3.2.4. *Processing cost*

Fig. 3 shows participants' mean trial response times. The model on log-transformed response times showed main effects of perspective, with longer response times observed in the different perspective condition, $\beta = 0.06$, $t = 2.81$, $p = .005$; and of partner, with shorter response times in the computer partner condition, $\beta = -0.46$, $t = -7.97$, $p < .001$. The latter result may reflect the larger proportion of egocentric responders we found in the computer partner condition, who (unlike othercentric responders) did not show the slow-down associated with perspective taking on different perspective trials. To further investigate this, we repeated the same analysis including responder group (egocentric vs. othercentric; sum-coded) as a predictor. We focused on the subset of egocentric and othercentric responders since mixed responders accounted for a very small percentage of participants (4%). In addition to effects of perspective, $\beta = 0.05$, $t = 2.60$, $p = .01$, and of partner, $\beta = -0.46$, $t = -7.87$, $p < .001$, this analysis revealed a perspective by responder group interaction, $\beta = 0.21$, $t = 3.17$, $p = .002$, reflecting faster responses by egocentric compared to othercentric responders on different perspective trials. This difference is visualized in the first two panels of Fig. 3b, which show an increase in response time on different perspective trials that is smaller for egocentric than othercentric responders, and was confirmed by separate models by responder group. For egocentric responders, there was no difference in response times on different and same perspective trials, $\beta = 0.03$, $t = 1.41$, $p = .15$. Othercentric responders, on the other hand, were significantly slower on different compared to same perspective trials $\beta = 0.26$, $t = 3.25$, $p = .003$. There was, however, no significant effect of partner ($p = .07$), nor did partner interact with perspective ($p > .2$).

3.3. Discussion

Experiment 1 investigated spatial perspective taking in listeners, comparing interaction with a human partner and a computer partner. We found an effect of partner identity, with higher rates of othercentric perspective taking when participants interacted with a human than a computer. This corroborates earlier results by Duran et al. (2011), which demonstrate that perspective taking behavior is sensitive to information about one's partner. Notably, our partner effects were in the opposite direction to Duran et al.'s, who observed more othercentric perspective taking when participants were told they were interacting with a computer partner. In contrast, we found higher rates of egocentricism with a computer partner. Our partner effects are, however, in line with more recent work on perspective taking in speakers in the context of human–robot interaction. These studies suggest that speakers are less likely to take their partner's perspective with a robot compared to a human partner (Carlson et al., 2014; Li et al., 2016), in contrast to earlier studies which found that speakers overwhelmingly used othercentric expressions when addressing a robot partner (Fischer, 2005; Moratz et al., 2001). This pattern of results may reflect a shift in peoples' perceptions of AI systems as communicative partners.

The distribution of responders in our experiment also reflected clear biases in individuals' perspective taking tendencies. As in Duran et al. (2011), we observed that mixed responders were a minority, with most participants committed toward egocentric or othercentric responding. This is in line with previous studies which demonstrate that perceivers tend to have a natural preference for adopting a perspective centered either on themselves or on others (Arnold, Spence, & Auvray, 2016), and implies a cognitive cost in the act of switching or mixing perspectives (Job, Kirsch, Inard, Arnold, & Auvray, 2021). However, unlike Duran et al., who found a roughly even split of egocentric and othercentric responders, we observed a much larger egocentric bias (>80% of responders) in our participants. On the surface, this appears to be in line with theories that emphasize an egocentric dominance and associated cost of othercentric perspective taking (e.g., Epley et al., 2004). Our response time results also demonstrate higher processing cost on different perspective trials, in particular for othercentric responders, who were more inclined to take their partner's perspective in the experiment. Why, then, were participants in Duran et al.'s study more willing to invest in the cost of othercentric perspective taking than our participants? One possibility could be differences in the processing cost associated with perspective taking on different directional axes. Duran et al. (2011) included ambiguous descriptions in the right/left as well as front/back dimensions, whereas our experiment solely focused on the former. These are likely harder for listeners to identify quickly on verbal command. Indeed, difficulty with discriminating "right" from "left," known as right–left confusion, is a well-established phenomenon in the spatial cognition literature (Vingerhoets & Sarrechia, 2009; Wolf, 1973), and has been observed even in neurologically healthy adults (van der Ham, Dijkerman, & van Stralen, 2021). In contrast, people do not appear to have the same extent of difficulty with conceptual representations in the front–back or up–down dimensions (Farrell, 1979; Vingerhoets & Sarrechia, 2009). Studies on the production of spatial descriptions have also observed that speakers often opt for alternative expressions such as cardinal directions to avoid using the terms "right" and "left," suggesting a difficulty with right and left that does not extend to other directional axes (Mainwaring et al., 2003). It is possible that this right–left confusability increased the difficulty of perspective taking in our experiment compared to Duran et al.'s, biasing our participants more strongly toward a cognitively simpler strategy of egocentricism. To further investigate this, we repeated the experiment using ambiguous front/back instead of right/left spatial descriptions. If the difference in othercentric tendencies between our participants and Duran et al.'s was indeed due to right–left difficulty, we would expect to see an increase in othercentricism in Experiment 2 compared to Experiment 1.

4. Experiment 2

We saw in Experiment 1 that spatial perspective taking behavior in listeners is sensitive to partner identity, with computer partners eliciting more othercentric perspective taking than human partners. Nevertheless, we observed a rather high rate of egocentricism across both partner conditions, which we speculate may be due to our exclusive use of right/left spatial descriptions, which are known to be more difficult for comprehenders compared to, for instance, front/back descriptions. To follow up on this, we designed Experiment 2 with the purpose of investigating ambiguous spatial descriptions in the front–back dimension. The experiment was a replication of Experiment 1 with the exception that we replaced right/left spatial descriptions with front/back ones, thus allowing us to perform a direct comparison of the two forms.

4.1. Method

4.1.1. Participants

Five hundred and eight participants were recruited on AMT. Data collection for the study took place in September 2021. We followed the same recruitment and exclusion criteria as in Experiment 1. Data were excluded from participants who were non-native speakers of English (7), who failed to meet the minimum accuracy criteria (146), or who indicated suspicion about the authenticity of their partner or the interaction (114). In addition, it emerged post-hoc that a subset of participants had confused the terms "front" and "back" by interpreting these from a top-down view rather than from the avatar's perspective (see Galati, Dale, & Duran, 2019 for a related discussion), thus mapping "front" to the top of the screen and "back" to the bottom (i.e., the opposite outcome to what was intended). Thus, in addition to our preregistered exclusion criteria, as a conservative measure we also excluded participants who selected the wrong object on more than one same perspective trial (21). The final dataset consisted of 220 participants (mean age = 38.7, range = 23–75), with 99 and 121 in the human and computer partner conditions respectively.

4.1.2. Materials, design, and procedure

These were identical to Experiment 1 other than the replacement of ambiguous right/left utterances with front/back ones (e.g.,"Give me the stapler in the front"). The corresponding objects appeared in the top and bottom positions on the table on these trials.

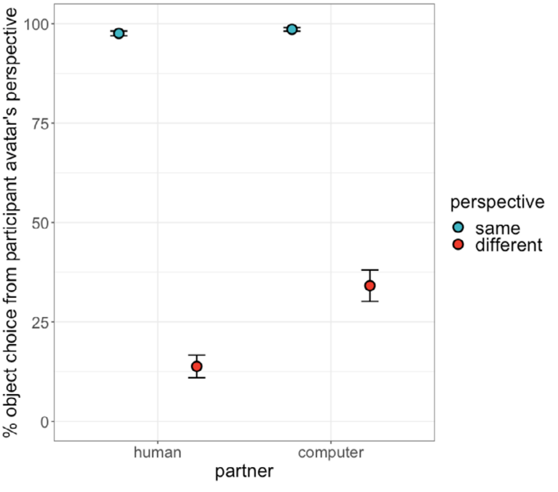


Fig. 4. Experiment 2: Percentage of trials on which participants chose the object from their own avatar's perspective (on same perspective trials, this was a shared perspective with the partner's avatar; on different perspective trials, this was across the table from the partner's avatar and, therefore, reflects egocentric perspective taking by the participant). Error bars represent ±1 standard error of by-participant means.

4.2. Results

4.2.1. Analyses

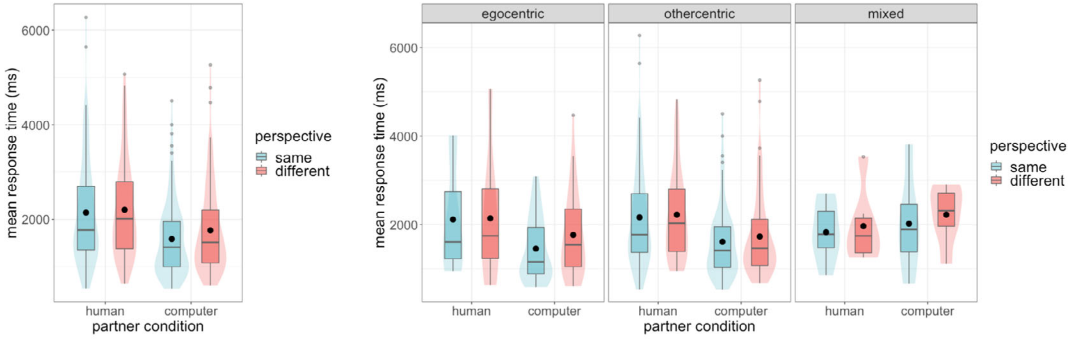
We followed the same analysis procedures as in Experiment 1.

4.2.2. Distribution of responders

Table 2 shows the breakdown of egocentric, othercentric, and mixed responders in the human and computer partner conditions. As in Experiment 1, mixed responders were in the minority, with most participants exhibiting a dominant egocentric or othercentric preference. Importantly, the distribution of responders shifted toward an othercentric bias here, suggesting that listeners' perspective taking tendencies were sensitive to the change from right/left to front/back ambiguity. As in Experiment 1, we examined the rate of egocentric and othercentric responding across the two partner conditions. For egocentric responders, there was a 23% decrease in the human compared to computer partner condition, $\chi^2(1) = 16.53$, $p < .001$. For othercentric responders, there was a 22% increase in the human compared to computer partner condition, $\chi^2(1) = 12.73$, $p = .001$.

4.2.3. Perspective taking tendencies

Fig. 4 shows the percentage of trials on which participants chose the object from their own avatar's perspective. Of interest is participants' behavior on different perspective trials, where this response reflects egocentric perspective taking by the participant. The model showed main effects of perspective, with participants less likely to choose the own-avatar-associated object on different perspective trials, $\beta = -11.66$, $z = -13.83$, $p < .001$; and of partner, with participants more likely to choose the own-avatar-associated object in the computer partner



(a) (b)

Fig. 5. Experiment 2 mean trial response times (in ms) (a) across all participants, and (b) split by responder group. Boxplots represent the median and interquartile range; black dots represent condition means.

condition, $\beta = 1.01$, $z = 2.06$, $p = .04$. There was no interaction between perspective and partner condition ($p > .8$).

4.2.4. *Processing cost*

Fig. 5 shows participants' mean trial response times. The model on log-transformed response times showed main effects of perspective, with longer response times observed in the different perspective condition, $\beta = 0.08$, $t = 3.29$, $p < .001$; and of partner, with shorter response times in the computer partner condition, $\beta = -0.24$, $t = -4.00$, $p < .001$. As in Experiment 1, we examined response time differences in egocentric and othercentric responders by including responder group as a predictor. This analysis showed no effect of responder group nor its interaction with any of the other predictors, suggesting that the processing cost of perspective taking was similar in the two groups.

4.2.5. *Combined analysis of Experiments 1 and 2*

To evaluate how differences in right/left and front/back ambiguity influenced listeners' perspective taking tendencies, we conducted an additional analysis combining the data from Experiments 1 and 2. This dataset consisted of 515 participants: 274 from Experiment 1 and 220 from Experiment 2. We constructed the same model as our main analysis examining perspective taking tendencies, with the addition of experiment (1 vs. 2; sum-coded) as a predictor. The model showed a main effect of perspective, with participants less likely to choose the own-avatar-associated object on different perspective trials, $\beta = -8.67$, $z = -11.71$, $p < .001$; of partner, with participants more likely to choose the own-avatar-associated object in the computer partner condition, $\beta = 1.39$, $z = 2.17$, $p = .03$; and of experiment, with participants less likely to choose the own-avatar-associated object in Experiment 2, $\beta = -8.31$, $z = -10.79$, $p < .001$. There was also an experiment by perspective interaction, $\beta = -15.37$, $z = -10.03$, $p < .001$, reflecting the larger effect of perspective in Experiment 2 compared to Experiment 1. In other words, when the participant and partner's perspective differed, participants were less likely to choose the own-avatar-associated object following ambiguous front/back utterances (Experiment 2) than ambiguous right/left utterances (Experiment 1).

4.3. Discussion

The results of Experiment 2 corroborate those of Experiment 1 in that listeners were more likely to respond othercentrically with a human compared to a computer partner. The crucial difference lies in the distribution of egocentric and othercentric responders across participants. Unlike Experiment 1 which saw a large bias toward egocentricism, in Experiment 2, the pattern was reversed, with a large bias toward othercentricism. The combined analysis highlights that perspective taking behavior was sensitive to the dimension of spatial ambiguity, suggesting that the change in responder distribution can be attributed to the change from right/left to front/back spatial utterances. Furthermore, unlike Experiment 1, we found no difference in the response times of egocentric and othercentric responders in Experiment 2. This may reflect the fact that front-back discrimination is cognitively simpler than right-left discrimination (Farrell, 1979), thereby eliciting a smaller cost in othercentric perspective taking. However, a secondary implication of this is that perspective taking in the front-back dimension alone may be too trivial to evoke clear behavioral differences between egocentric and othercentric responders. This echoes Duran et al.'s (2011) observation that collapsing right/left and front/back trials weakened their effects, and stronger effects on processing cost were seen when right/left trials were considered alone.

In both Experiments 1 and 2, we found consistent results which indicate stronger othercentric tendencies with a human than a computer partner. These results stand in contrast to the higher rates of othercentricism with a computer partner observed by Duran et al. (2011). We speculated that our results might reflect a shift in people's perception of computers' capabilities arising from recent developments in AI technology. The increasing prevalence of computers as effective interactional partners in day-to-day life may have led to greater expectations regarding their collaborative abilities, prompting an increased willingness in people to shift the burden of perspective taking onto them. It follows, then, that people's expectations about the computer's collaborative abilities should be directly influenced by the apparent capabilities of the computer. Indeed, Branigan et al. (2011) found that participants were more likely to adopt the naming preferences of a computer that appeared to be limited in functionality compared to one that appeared to be highly capable—a finding attributed to people tailoring their level of communicative design based on their perception of their computer partner's communicative capacity. Similarly, we might expect perspective taking tendencies to vary based on people's perception of the computer's communicative capacity—a computer perceived as more capable should elicit higher expectations of its perspective taking ability, leading to more egocentric behavior in listeners; conversely, a less capable computer should elicit lower expectations and, therefore, more othercentric behavior in listeners. To provide a direct test of this, in Experiment 3, we focused on the computer partner condition and manipulated its outward presentation. This allowed us to induce different expectations in listeners with respect to its communicative capabilities.

5. Experiment 3

Experiment 3 investigated spatial perspective taking in listeners with computer partners of different perceived capabilities. We hypothesized that people should be egocentric to different degrees based on their expectations about their partner's communicative capacities and, therefore, potential perspective taking ability. Specifically, we expected more egocentric behavior with a computer that appeared to be more capable, because people are more likely to shift the burden of perspective taking onto such a partner.

We induced different expectations by manipulating the computer's appearance and linguistic behavior: in the "basic" partner condition, participants interacted with a software system that had an old-fashioned appearance and limited linguistic capabilities; in the "modern" partner condition, the software had a more advanced appearance and better linguistic capabilities. In addition, we included both ambiguous right/left and front/back utterances here in attempt to elicit a more even distribution of egocentric and othercentric responders. All other aspects of the task remained the same as Experiments 1 and 2.

5.1. Method

5.1.1. Participants

Three hundred and two participants were recruited on Prolific.[3] Data collection for the study took place in May 2022. We used Prolific filters to restrict recruitment to native English speakers over the age of 35 based in the United States. We restricted our sample to this age group based on the rationale that younger adults would presumably not be familiar with an "outdated" form of AI, and therefore unlikely to show differences based on our computer manipulation. Participants were randomly assigned to the "basic" computer partner (146) or the "modern" computer partner (156) condition. No participants were excluded on the basis of English non-nativeness or failing to meet the accuracy threshold. As in Experiment 2, we excluded the subset of participants who appeared to mis-interpret "front" and "back," defined as those who selected the wrong object on more than one same perspective trial (18). Thus, the final dataset consisted of 284 participants (mean age = 44.9, range = 35–74), with 138 and 146 in the "basic" and "modern" computer partner conditions respectively.

5.1.2. Materials, design, and procedure

The experiment was the same as Experiment 1 except for the following changes:

1. The number of critical trials increased to 16—8 same and 8 different perspective, with 4 right/left and 4 front/back utterances in each perspective condition.
2. Location contrast fillers were omitted, leaving color contrast, size contrast, and no contrast fillers (12 each). Target objects that originally appeared in location contrast fillers were used as targets in the additional critical trials. The total number of filler trials remained the same (36).
3. All participants interacted with a computer partner. Participants were randomly assigned to the "basic" or the "modern" computer partner condition. Prior to beginning, participants saw an application start-up window (see Fig. 6) and heard a record-

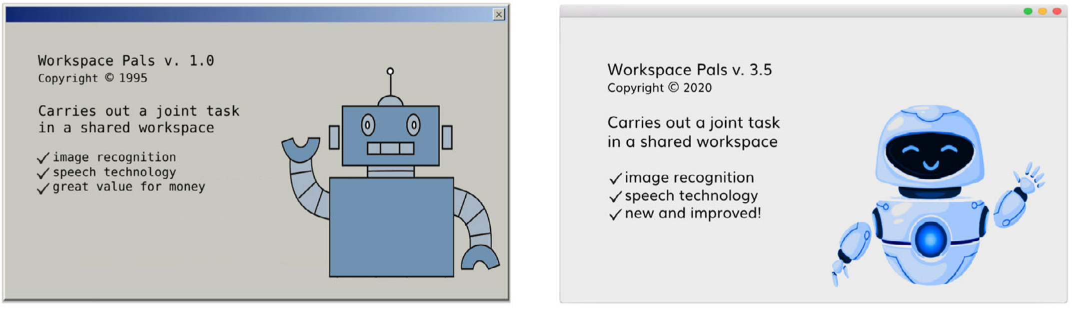


Fig. 6. Start-up windows seen by participants in the basic computer partner (left) and modern computer partner (right) conditions.

ing of the software introducing itself as their partner. The basic computer partner window suggested an outdated and simplistic software system. Recordings were synthesized using the eSpeak text-to-speech software, an open-source synthesizer that uses the "formant synthesis" method, with pitch, speed, and gap parameters set to 95, 200, and 7, respectively.[4] The modern computer partner window suggested an up-to-date and sophisticated software system. The same set of utterance recordings from the computer partner condition in Experiments 1 and 2 were used.

4. During the task, a "?" button appeared at the bottom-right of the screen on each trial, which participants were instructed to click in the event that they did not understand their partner's request.
5. The post-test questionnaire at the end included a question asking participants to rate on a scale of 1 (very easy) to 5 (very difficult) how easy they found their partner to understand. This was used to verify that the partner manipulation evoked perceptual differences between the two partner conditions as intended.

5.2. *Results*

5.2.1. *Analyses*

We first compared participants' understanding ratings of their partner in the two partner conditions, using linear regression with rating as the dependent variable and partner condition (basic vs. modern; sum-coded) as the predictor. We then analyzed the distribution of responders, listeners' perspective taking tendencies, and their processing cost, following the same analysis procedures as in Experiment 1. We also conducted exploratory analyses examining differences by age on perspective taking. These results are reported in Appendix B in the Supplementary Material.

5.2.2. *Understanding ratings*

Table 3 shows the mean understanding ratings provided by participants in the two partner conditions. Participants in the modern computer partner condition rated their partner as less difficult to understand, $\beta = -0.26$, $p =< .001$. This confirms that our partner manipu-

Table 3
Mean understanding ratings of their partner (standard deviation in parentheses) provided by participants in the two partner conditions

Computer partner	Mean rating
Basic	2.33 (0.97)
Modern	2.07 (0.91)

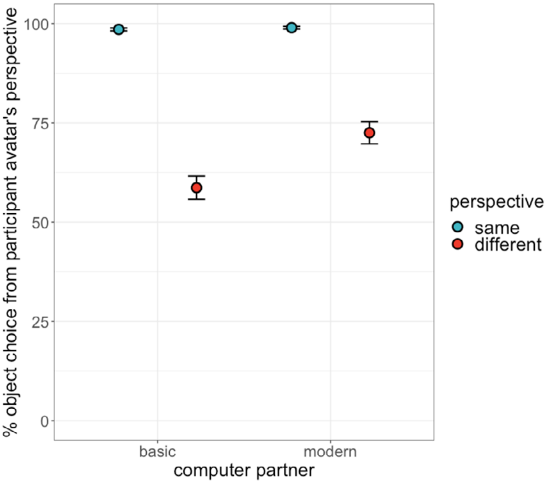


Fig. 7. Experiment 3: Percentage of trials on which participants chose the object from their own avatar's perspective (on same perspective trials, this was a shared perspective with the partner's avatar; on different perspective trials, this was across the table from the partner's avatar and, therefore, reflects egocentric perspective taking by the participant). Error bars represent ±1 standard error of by-participant means.

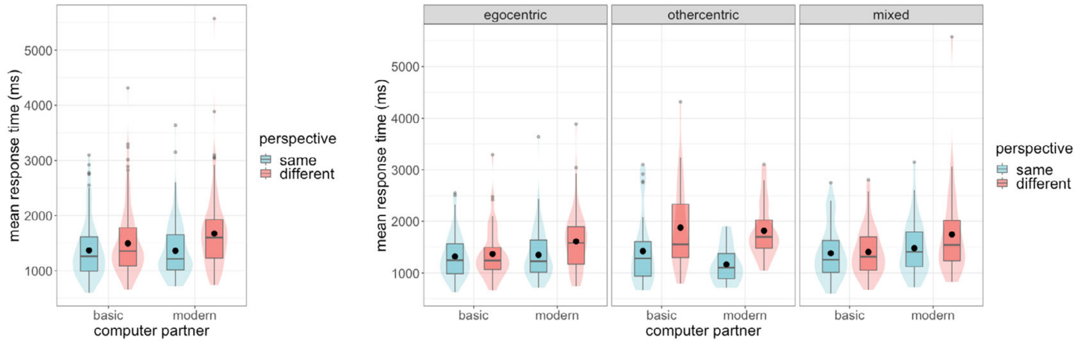
lation was successful at inducing perceptual differences in the two computers' communicative capacities.

5.2.3. Distribution of responders

Table 2 shows the breakdown of egocentric, othercentric, and mixed responders in the basic and modern computer partner conditions. As before, we used Duran et al.'s (2011) threshold to determine group membership (at least 70% of trials showing a given response type), which in this case corresponded to six out of eight different perspective critical trials. We compared the rate of egocentric and othercentric responding across the two partner conditions. For egocentric responders, there was a 20% increase in the modern compared to basic computer partner condition, $\chi^2(1) = 11.46$, $p < .001$. For othercentric responders, there was an 8% decrease in the modern compared to basic computer partner condition, $\chi^2(1) = 3.29$, $p = .07$.

5.2.4. Perspective taking tendencies

Fig. 7 shows the percentage of trials on which participants chose the object from their own avatar's perspective. Of interest is participants' behavior on different perspective trials, where



(a) (b)

Fig. 8. Experiment 3 mean trial response times (in ms) (a) across all participants, and (b) split by responder group. Boxplots represent the median and interquartile range; black dots represent condition means.

this response reflects egocentric perspective taking by the participant. The model showed main effects of perspective, with participants less likely to choose the own-avatar-associated object on different perspective trials, $\beta = -5.04$, $z = -13.98$, $p < .001$; and of partner, with participants more likely to choose the own-avatar-associated object with the modern computer partner, $\beta = 0.89$, $z = 2.52$, $p = .01$. There was no interaction between perspective and partner.

5.2.5. *Processing cost*

Fig. 8 shows participants' mean trial response times. The model on log-transformed response times showed a main effect of perspective, with longer response times observed in the different perspective condition, $\beta = 0.13$, $t = 7.43$, $p < .001$. There was also a main effect of partner, $\beta = 0.08$, $t = 2.44$, $p = .02$, superseded by a partner by perspective interaction, $\beta = 0.09$, $t = 2.61$, $p = .01$, reflecting a greater slowdown on different perspective trials with the modern compared to the basic computer partner. We did not have any predictions regarding processing cost differences between the two computer partners; however, to evaluate whether this result was due to different proportions of egocentric/othercentric responders in the two conditions, we repeated the same analysis including responder group (egocentric vs. othercentric) as a predictor.

This model showed a main effect of perspective, $\beta = 0.16$, $t = 7.52$, $p < .001$, and a perspective by partner interaction $\beta = 0.13$, $t = 3.08$, $p = .002$. In addition, there was a perspective by responder group interaction, $\beta = 0.26$, $t = 5.23$, $p < .001$, suggesting differences in processing cost of taking a partner's perspective in egocentric versus othercentric responders. We conducted separate analyses by responder group to explore the interaction. For othercentric responders, the model revealed an effect of perspective, $\beta = 0.33$, $t = 7.03$, $p < .001$, with longer response times on different perspective trials. This reflects the expected cost of taking a partner's perspective, and is visualized in the second panel of Fig. 8. There was no effect of partner nor an interaction between partner and perspective (all $p > .2$), suggesting that this cost was similar with both computer partners. For egocentric responders, there were main effects of perspective, $\beta = 0.10$, $t = 4.20$, $p < .001$, and of partner, $\beta = 0.12$, $t = 2.53$, $p = .01$. There was also a perspective by partner interaction, $\beta = 0.13$, $t = 2.75$, $p = .006$, reflecting longer response times on different perspective trials, in particular with the modern computer partner. This result is visualized in the first panel of Fig. 8. This is again the expected effect due to taking a partner's perspective on different perspective trials; however, we do not have a clear explanation for why it is only present with the modern computer partner. It is likely that egocentric responders in the basic computer partner condition did not take their partner's perspective enough for a processing cost effect to emerge. Notably, this is consistent with our processing cost results in Experiment 1, where we saw no difference in processing cost between same and different perspective trials in the egocentric responder group.

5.3. Discussion

Experiment 3 manipulated listeners' perception of their computer partner's communicative capacity. We found that listeners were more egocentric with a computer that appeared to be modern and advanced, than one that appeared to be basic and limited in functionality. This suggests that listeners' perspective taking behavior was tied to their expectations about the computer's potential ability to collaborate. A computer perceived as more capable likely also induced higher expectations of its ability to take perspective; conversely, one perceived as less capable likely induced lower expectations, prompting listeners to take on the burden of perspective taking themselves. This is in line with Clark and Wilkes-Gibbs's (1986) principle of least collaborative effort, in which interlocutors adapt their behavior in order to maintain mutual collaborative understanding with their partner. Experiment 3's results further highlight the role of technological expectations on listeners' adaptive behavior: listeners appear to go beyond making a distinction between human and computer partners, to tailoring their perspective taking tendencies to computer partners of different capacities. Importantly, the directionality of effect reveals stronger egocentric tendencies with a computer perceived as having higher communicative capabilities. This is consistent with our hypothesis that the stronger egocentric tendencies we observed with computers in Experiments 1 and 2 are likely due to higher expectations regarding computer partners' collaborative abilities today.

6. General discussion

In this paper, we report three experiments investigating partner effects on spatial perspective taking behavior in listeners. In a task of simulated interaction, participants heard instructions from their partner such as "Give me the stapler on the left," and were asked to move the corresponding object over to their partner's avatar. On critical trials, the referent was spatially ambiguous between two objects, licensing an egocentric or an othercentric interpretation.

Experiments 1 and 2 provide a conceptual replication of Duran et al.'s (2011) study (Experiments 1 and 3) comparing behavior with a human and a computer partner. We found that participants were more egocentric with a computer than a human partner. While these results are at odds with Duran et al.'s, they could be explained by a change in listeners' expectations regarding computer partners' collaborative abilities. To further investigate this, Experiment 3 compared listeners' behavior with computer partners of different perceived communicative capacities. We found that participants were more egocentric with a computer that appeared to be more capable than one that appeared less capable, highlighting how technological expectations come to influence perspective taking in human–computer interaction.

6.1. *Partner effects on spatial perspective taking*

Overall, our results are in line with accounts of perspective taking which emphasize the cognitive cost of taking another's perspective (Epley et al., 2004; Horton & Keysar, 1996; Horton & Gerrig, 2005). Different perspective trials consistently elicited a higher processing cost, in particular in participants who were classified as "othercentric" based on their predominant mode of response. Importantly, however, our results also demonstrate the flexible nature of perspective taking: the fact that we still observed othercentric behavior suggests that listeners were willing to invest in this cost, the extent to which depending on who their communicative partner was. Thus, in accordance with audience design theories, our experiments show the partner-specificity of perspective taking behavior. Specifically, we found higher rates of egocentricism when listeners believed they were interacting with a computer than a human, and with a computer that they perceived as having higher communicative competence.

Studies on audience design typically focus on the speaker domain, showing how speakers adapt their utterances with respect to a specific audience (e.g., Fussell & Krauss, 1989). Here, we find evidence for the relevance of similar mechanisms in the listener domain—listeners appear to construct a mental model of the speaker based on their perception of the speaker's collaborative ability, and tailor their interpretation of the message accordingly. Thus, when confronted with an ambiguous spatial description, listeners adjust their perspective taking tendencies to be more or less egocentric depending on the perspective they expect the speaker to take. A question that follows then is *how* listeners engage in such adaptation. One possibility is that listeners have an existing mental model of a "generic" category of speaker, based on, for example, prior knowledge or experience (e.g., a general impression of how good computers may be at perspective taking), which they draw upon. Another possibility, of course, is that listeners are additionally able to adapt (perhaps dynamically) to specific speakers based on for instance an inference about the speaker that arises from some aspect of their behavior (e.g., tailoring their impression of how good a computer is at perspective taking based on its appearance or behavior). While the exact nature of the mechanism is not something we set out to address, our results from Experiment 3—that listeners distinguish between computers that present as more or less capable —suggest some degree of an inference-based adaptation. This is reinforced by the difference in understanding ratings we observed in the two partner conditions, which indicate perceptual differences that listeners detected in the two computers' communicative capacities. Thus, at least in the context of our third experiment, listeners appear to tailor their perspective taking expectations in line with how their partner presents themselves.

6.2. *Partner effects in the context of human–computer interaction*

Within the domain of human–computer interaction, the behavior of our listeners is in accordance with findings by Nass and colleagues in the CASA framework. The fact that listeners showed othercentric tendencies even with computer partners (albeit to a smaller degree than with human partners) embodies the sort of "social" response to and cooperation with computers that Nass and colleagues allude to. Moreover, listeners are sensitive to the degree of competency displayed by the computer (Experiment 3; cf. Leshner et al., 1998), much like they would be to the degree of knowledgeability shown by a human interlocutor (Ferguson & Breheny, 2012; Isaacs & Clark, 1987). In our study, we manipulated both the visual appearance (avatar) and the linguistic ability (speech/voice quality) of the computer partner to create distinct impressions of the computer partner's competence. However, studies on human–computer interaction have shown that either visual (e.g., De Visser et al., 2016; Leshner et al., 1998) or voice-based (e.g., Nass & Lee, 2000) cues alone are sufficient at influencing social responses in the trust and attitude domains. A question that remains for future research then is whether these cues on their own similarly influence perspective taking behavior in listeners.

Notably, our partner effects in Experiments 1 and 2 were in the opposite direction to Duran et al.'s (2011), who observed a higher rate of egocentricism when participants believed they were interacting with a human rather than a computer. There are several potential explanations we can consider for this difference. First, methodological differences between their study and ours may have contributed to the disparity in results. The design of Duran et al.'s experiment utilized only critical utterances with a perspective ambiguity and unambiguous controls; in contrast, our study included a variety of filler utterances to distract participants from the perspective manipulation. In part due to this design choice, our study also featured fewer different-perspective critical trials compared to Duran et al.'s. It is possible that these differences in trial numbers may have influenced participants' perspective taking behavior, for instance, by diverting their attention from the spatial ambiguity in different perspective trials, thereby reducing their likelihood of developing othercentric tendencies over time. However, while this may have led to an overall increase in egocentric tendencies in our study, we see no reason for such an effect to differ across the two partner conditions. Another methodological difference lies in our additional use of a voice-based cue, in the form of different audio recordings for the two speakers, whereas Duran et al. (2011) relied solely on explicit instruction. This may have strengthened our partner effects by reinforcing listeners' conceptualizations of their partner's identity (see, e.g., Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008 for a demonstration of voice-based cues on listener inferences); however, it also does not predict a reverse in the directionality of the partner effects.

This brings us to a third explanation that participants in Duran et al.'s study and our study may have had different expectations with respect to their partners' perspective taking capabilities. Duran et al. interpret their results as listeners' use of attributional cues to motivate perspective taking—specifically, computers were perceived as having a lower potential to collaborate, prompting listeners to take on the burden of perspective taking themselves (cf. Clark & Wilkes-Gibbs, 1986). Assuming listeners in our study made similar collaborative attributions to their partners, it would appear that our participants favored computers as more proficient collaborators capable of adopting the burden of perspective taking. This is in accordance with recent studies on perspective choice in speakers, which demonstrate what appears to be a shift toward less othercentric perspective taking with robot compared to human partners (Carlson et al., 2014; Li et al., 2016), in contrast to a strong bias toward othercentricism with robot partners observed in earlier work (Moratz et al., 2001; Tenbrink et al., 2002). Such an explanation would be in line with Clark and Wilkes-Gibbs's (1986) principle of least collaborative effort, which posits that interlocutors make behavioral adjustments (e.g., adapting their perspective) taking into account both parties' relative abilities to contribute to mutual understanding, while allowing us to reconcile the different pattern of partner effects in the two studies. Recent work has shown that children as young as 7 apply the principle of least collaborative effort in their interaction with a NAO robot by adjusting their initially egocentric perspective choice to accommodate the robot after it fails to complete the task (Yadollahi, Couto, Dillenbourg, & Paiva, 2022). These results align within the landscape of technological evolvement as even a younger cohort of users demonstrates adaptability to the collaborative abilities of AI partners.

Our results from Experiment 3 further strengthen this notion of a shift in technological expectations toward computers as communicative partners. Here, we saw that listeners were more egocentric with a computer partner perceived as being more capable of collaborating. These results are significant on three fronts. First, they highlight the adaptive nature of listeners' perspective taking behavior —in particular, this adaptation appears to go beyond simple discrimination between humans and computers to a more nuanced attribution of differences between computers of different specifications. Second, they show that audience design mechanisms, which have been shown to influence interaction with human partners based on partner knowledge and expertise (e.g., Fussell & Krauss, 1989; Isaacs & Clark, 1987), also apply to computer partners. Third, they reveal *how* listeners' expectations about the computer's expertise matter: a seemingly outdated computer that listeners had more trouble understanding elicited a stronger tendency toward othercentricism in our study. Analogously, we might expect that a computer partner from the time of Duran et al.'s (2011) study could have elicited similar (low) expectations about its collaborative abilities, and therefore a stronger othercentric bias, compared to a computer partner today.

We note that the rate of othercentricism observed in Duran et al.'s computer partner condition is in fact higher than our basic computer condition (52% vs. 20%); in contrast, they found a lower rate of mixed responding than we did (9% vs. 43%). This is somewhat surprising, particularly when considering that our proportion of mixed responders in Experiments 1 and 2 was similar to theirs. While we do not have a definitive explanation here, we speculate that variations in design across the two studies could have contributed to the difference. For instance, Duran et al. included a higher number of different perspective trials than we did (20 vs. 8)—it is possible that participants became increasingly othercentric over time, leading to overall more othercentricism in their study. In designing our study, we prioritized having a more naturalistic interaction context by including a variety of fillers, thus limiting the number of critical trials we had; however, future studies could consider opting for a higher number of critical trials to allow for a time course analysis on perspective taking behavior. Duran et al. also included more degrees of partner rotation—notably, the addition of a 90° offset in addition to 0° and 180°. Although they do not provide a breakdown of othercentric tendencies by angle of rotation, their response time measures suggest that participants largely found 90° to be easier than 180°; it is thus conceivable that the inclusion of a 90° offset increased othercentric behavior in their study. Finally, Duran et al. had a higher proportion of different perspective trials in their experiment (50%), whereas our inclusion of filler trials lowered this percentage (15%)—participants in Duran et al.'s study therefore had more opportunities to take their partner's perspective, potentially boosting their othercentric tendencies.

Notably, the increase in mixed responding in our study was only observed in Experiment 3, where all participants interacted with a computer partner. Thus, another possibility of course is that at the time of our study, even with a basic and seemingly limited computer, participants still had higher expectations of its collaborative ability compared to participants at the time of Duran et al. (2011). This would be consistent with the idea of an update in technological expectations from the time of their study to today. We also note that in a separate study utilizing the same design with a computer partner but employing a data-driven method of classifying participants rather than predefined manual thresholds, we observed a similar split of three responder groups across participants (40/30/30% egocentric/mixed/othercentric; Loy & Demberg 2023). The threshold for mixed responders estimated by that model (37.5–75%) is also similar to Duran et al.'s cutoff of 30–70%. Thus, the distribution of responders we observed, at least in the context of the current design and at the time of the current study, appears to be robust.

Taken together, the current study alongside Duran et al. (2011) demonstrate how the evolving landscape of human–computer interaction has come to influence human behavior. Such a shift in behavior likely reflects a larger change in attitudes and expectations toward computers as competent communicative partners, prompted by advancements in AI and the growing presence of interactive AI systems in everyday life (Horstmann & Krämer, 2019). In particular, in the last decade or so, we have seen a proliferation of research on human–robot interaction and collaboration (see Ajoudani et al., 2018; Vianello et al., 2021 for reviews). These studies highlight methodological and technological developments in human–robot interaction, the application of AI in various everyday contexts, and notably, robot as well as human adaptability in collaborative contexts (e.g., Nikolaidis, Hsu, & Srinivasa, 2017). Our study adds to this growing literature on user expectations in the face of AI progress, particularly within the domain of communication and collaboration. The current results show that people's expectations in a partner's perspective taking abilities extend from the computer's comprehension of spatial descriptions (in earlier studies) to their production of such utterances here. In other words, listeners perceive computers as capable of upholding the burden of perspective taking by constructing utterances from an othercentric perspective. Since the time of our experiments, the technological capabilities of AI have continued to grow, in both the social and language domains (Katz, Bommarito, Gao, & Arredondo, 2023; Kortemeyer, 2023; Wilcock & Jokinen, 2022). Public awareness of this has also increased, fuelled notably by the recent advent of ChatGPT, the publicly accessible AI-powered chatbot developed by OpenAI[5]

(Hill-Yardin, Hutchinson, Laycock, & Spencer, 2023; Kasneci et al., 2023; Lo, 2023). Given these developments, we could expect that the tendency toward more egocentricism with computer partners that we observe would also continue to increase. In addition, it would be interesting to examine how these tendencies might be modulated by listener-specific characteristics, such as participants' attitudes toward technology, or their personal experiences with AI.

6.3. Differences in right/left versus front/back perspective taking

An additional finding that emerged in our study is that of perspective taking differences with right/left and front/back utterances. While we did not set out with the intention to examine differences in this domain, by including only one form each in Experiments 1 and 2, we were able to capture perspective taking behavior of each response type in isolation, without expectations due to inclusion of the other. Here, we saw clear differences between the two forms of description. In Experiment 1, we found that ambiguous right/left utterances elicited high rates of egocentricism, whereas in Experiment 2, front/back utterances elicited high rates of othercentricism. Our combined analysis confirmed that listeners were more egocentric in Experiment 1 than in Experiment 2. When both forms were included in Experiment 3, listeners' tendencies still appeared to be influenced on an utterance-level basis, as reflected by the higher rates of "mixed" responders we saw in Experiment 3 compared to Experiments 1 and 2. This suggests that at least some of the trade-off can be attributed to listeners who opted for both perspectives, rather than the strong preference for either egocentricism or othercentricism that has been noted in previous studies (e.g., Arnold et al., 2016; Duran et al., 2011; Wilke, Bender, & Beller, 2019). It is also notable that Duran et al. found a significant increase in othercentric responders when comparing across partner conditions, whereas the analogous increase in our Experiment 1 was not statistically borne out. While this difference may simply be due to partner effects in the opposite direction, the inclusion of (or lack of) front/back utterances in the stimuli may have also played a role: intermixing both utterance forms in their study likely boosted participants' othercentric tendencies, whereas our inclusion of only right/left utterances in Experiment 1 may have weakened the tendency to be more othercentric with the human partner. This suggests that the decision to present the two forms of utterance in conjunction or in insolation may have consequences on perspective taking behavior —intermixing the two may lead to one response type influencing the other, such as by increasing the degree of othercentricism that might otherwise occur with right/left utterances alone, or decreasing that with front/back utterances alone.

Indeed, across our three experiments, we see a clear trend toward more othercentricism as the proportion of front/back to right/left utterances increases. Why, then, are othercentric tendencies so much higher with front/back utterances? Researchers who have observed similar effects typically cite cognitive reasons, such as difficulty accessing right/left representations or discriminating right/left verbal commands (Franklin & Tversky, 1990; Newcombe & Huttenlocher, 1992). These explanations echo findings on right–left confusion reported in the spatial cognition literature (Wolf, 1973). Self-reported right–left confusion has also been shown to correlate with performance in the Money Road Map Task, with high-confusability participants being slower and less accurate, suggesting links between right/left discrimination ability and spatial perspective taking (Yamashita, 2013). Our processing cost results support this notion of right–left difficulty. Although we observed slower response times overall on different perspective trials, breaking down the difference by responder group reveals a striking pattern. In particular, othercentric responders showed a greater processing cost with right/left descriptions than their egocentric counterparts (Experiment 1), whereas this othercentric disadvantage was not observed with front/back descriptions (Experiment 2). Notably, Duran et al. (2011) report a similar effect, with right/left trials characterized by greater processing costs in othercentric responders, whereas processing costs for front/back trials were virtually identical in the two groups.

Our results emphasize that the difficulty associated with spatial perspective taking is not equal across dimensions; rather, a right/left disadvantage is apparent, in both listeners' willingness to adopt an othercentric perspective, as well as the processing costs involved when they do so. These results have implications for future work addressing spatial perspective taking in comprehension. Here, we see that a simple difference in the form of utterances can result in opposite biases toward egocentric or othercentric tendencies; in addition, the cognitive cost associated with othercentric perspective taking is notably greater with right/left than with front/back utterances. This demonstrates that spatial perspective taking is not operationally equivalent across linguistic contexts; in particular, spatial relations along the right/left and front/back dimensions are conceptually distinct, to the extent that they entail different processing costs associated with othercentric perspective taking.

7. Conclusion

In three experiments, we investigated the role of partner effects on spatial perspective taking behavior in listeners. We found that listeners were consistently more egocentric with a computer compared to a human partner, and with a computer partner perceived to have higher communicative competence. Our findings stand in contrast to earlier results which demonstrate stronger egocentric tendencies with humans compared to computers, and point toward an update in expectations regarding the collaborative abilities of computer partners—a shift characterized by transition from viewing computers as socially limited communicative partners to being capable of advanced cognitive behavior, such as perspective taking.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 programme (ERC Starting Grant "Individualized Interaction in Discourse", grant agreement No. 948878).

Open access funding enabled and organized by Projekt DEAL.

Open Research Badges

This article has earned Open Data. Data is available at https://osf.io/83brp/, https://osf.io/gmcfa/ and https://osf.io/aqpcs/.

Notes

1 Preregistration details and data for the experiments can be found at: https://osf.io/xvfr5/(Experiments 1 and 2) and https://osf.io/3xzmy/(Experiment 3).
2 www.mturk.com
3 https://prolific.co
4 http://espeak.sourceforge.net
5 https://chat.openai.com

References

Ajoudani, A., Zanchettin, A. M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., & Khatib, O. (2018). Progress and prospects of the human–robot collaboration. *Autonomous Robots*, *42*, 957–975.

Arnold, G., Spence, C., & Auvray, M. (2016). Taking someone else's spatial perspective: Natural stance or effortful decentring? *Cognition*, *148*, 27–33.

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, *71*(3), 409–427.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41–57.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One*, *5*(5), e106953.

Brügger, A., Richter, K.-F., & Fabrikant, S. I. (2019). How does navigation system behavior influence human behavior? *Cognitive Research: Principles and Implications*, *4*, 1–22.

Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, *31*(1), 74–98.

Carlson, L., Skubic, M., Miller, J., Huo, Z., & Alexenko, T. (2014). Strategies for human-driven robot comprehension of spatial descriptions by older adults in a robot fetch task. *Topics in Cognitive Science*, *6*(3), 513–533.

Clark, H. H. (1973). Space, time, semantics, and the child. In J. T. Moore (Ed.), *Cognitive development. The acquisition of language* New York: Academic Press.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *22*(3), 331.

Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, *121*(1), 22–40.

Edwards, A., Edwards, C., Spence, P. R., Harris, C., & Gambino, A. (2016). Robots in the classroom: Differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher". *Computers in Human Behavior*, *65*, 627–634.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327.

Farooq, U., & Grudin, J. T. (2017). Paradigm shift from human computer interaction to integration. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1360–1363).

Farrell, W. S. (1979). Coding left and right. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 42.

Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology*, *70*(8), 1646–1660.

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, *48*(1), 257–263.

Fischer, K. (2005). Discourse conditions for spatial perspective taking. In *Proceedings of WoSLaD Workshop on Spatial Language and Dialogue, Delmenhorst*.

Fischer, K. (2006). The role of users' concepts of the robot in human–robot spatial instruction. In *International Conference on Spatial Cognition* (pp. 76–89). Springer.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, *119*(1), 63.

Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, *19*(6), 509–525.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*(3), 378.

Galati, A., & Avraamides, M. N. (2013). Flexible spatial perspective-taking: Conversational partners weigh multiple cues in collaborative tasks. *Frontiers in Human Neuroscience*, *7*, 618.

Galati, A., & Avraamides, M. N. (2015). Social and representational cues jointly influence spatial perspective-taking. *Cognitive Science*, *39*(4), 739–765.

Galati, A., Dale, R., & Duran, N. D. (2019). Social and configural effects on the cognitive dynamics of perspective-taking. *Journal of Memory and Language*, *104*, 1–24.

Galati, A., Diavastou, A., & Avraamides, M. N. (2018). Signatures of cognitive difficulty in perspective-taking: Is the egocentric perspective always the easiest to adopt? *Language, Cognition and Neuroscience*, *33*(4), 467–493.

Hill-Yardin, E. L., Hutchinson, M. R., Laycock, R., & Spencer, S. J. (2023). A Chat (GPT) about the future of scientific publishing. *Brain, Behavior, and Immunity*, *110*, 152–154.

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*, 939.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589–606.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127–142.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117.

Hou, Y., Zhang, K., & Li, G. (2021). Service robots or human staff: How social crowding shapes tourist preferences. *Tourism Management*, *83*, 104242.

Huang, D., Chen, Q., Huang, J., Kong, S., & Li, Z. (2021). Customer–robot interactions: Understanding customer experience with service robots. *International Journal of Hospitality Management*, *99*, 103078.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26.

Job, X., Kirsch, L., Inard, S., Arnold, G., & Auvray, M. (2021). Spatial perspective taking is related to social intelligence and attachment style. *Personality and Individual Differences*, *168*, 109726.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274.

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). Gpt-4 passes the bar exam. *Available at SSRN 4389233*.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, *43*(5), 1038–1049.

Kim, S. S., Kim, J., Badu-Baiden, F., Giroux, M., & Choi, Y. (2021). Preference for robot service or human service in hotels? Impacts of the COVID-19 pandemic. *International Journal of Hospitality Management*, *93*, 102795.

Kortemeyer, G. (2023). Using artificial-intelligence tools to make LaTeX content accessible to blind readers. *arXiv preprint arXiv:2306.02480*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotype. In *CHI'00 Extended Abstracts on Human Factors in Computing Systems* (pp. 289–290).

Leshner, G., Reeves, B., & Nass, C. (1998). Switching channels: The effects of television channels on the mental representations of television news. *Journal of Broadcasting & Electronic Media*, *42*(1), 21–33.

Li, S., Scalise, R., Admoni, H., Rosenthal, S., & Srinivasa, S. S. (2016). Spatial references and perspective in natural language instructions for collaborative manipulation. In *25th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 44–51). IEEE.

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, *13*(4), 410.

Loy, J. E., & Demberg, V. (2023). Individual differences in spatial orientation modulate perspective taking in listeners. *Journal of Cognition*, *6*(1), 52.

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology*, *12*, 589585.

Mainwaring, S. D., Tversky, B., Ohgishi, M., & Schiano, D. J. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial Cognition and Computation*, *3*(1), 3–42.

Marin-Urias, L. F., Sisbot, E. A., & Alami, R. (2008). Geometric tools for perspective taking for human–robot interaction. In *2008 Seventh Mexican International Conference on Artificial Intelligence* (pp. 243–249). IEEE.

Mazhar, O., Navarro, B., Ramdani, S., Passama, R., & Cherubini, A. (2019). A real-time human–robot interaction framework with robust background invariant hand gesture detection. *Robotics and Computer-Integrated Manufacturing*, *60*, 34–48.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, *68*, 327–337.

Moon, Y., & Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human–computer interaction. *Communication Research*, *23*(6), 651–674.

Moratz, R., Fischer, K., & Tenbrink, T. (2001). Cognitive modeling of spatial reference for human–robot interaction. *International Journal on Artificial Intelligence Tools*, *10*(04), 589–611.

Nass, C., & Lee, K. M. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 329–336).

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology*, *29*(5), 1093–1110.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78).

Newcombe, N., & Huttenlocher, J. (1992). Children's early ability to solve perspective-taking problems. *Developmental Psychology*, *28*(4), 635–643.

Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A., De Ruiter, J. P., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition*, *111*(1), 46–54.

Nikolaidis, S., Hsu, D., & Srinivasa, S. (2017). Human–robot mutual adaptation in collaborative tasks: Models and experiments. *International Journal of Robotics Research*, *36*(5–7), 618–634.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richert, A. (2018). Socializing with robots. In R. M. K. North & O. Haas (Eds.), *Knowledge management in digital change: New findings and practical cases* (pp. 97–110). Springer.

Romero, D., Bernus, P., Noran, O., Stahre, J., & Fast-Berglund, Å. (2016). The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In *Advances in Production Management Systems. Initiatives for a Sustainable World: IFIP WG 5.7 International Conference, APMS 2016, Iguassu Falls, Brazil, September 3–7, 2016, Revised Selected Papers* (pp. 677–686). Springer.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*(1), 1–24.

Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. *Spatial Language and Dialogue*, *1*, 23–39.

Shelton, A. L., & McNamara, T. P. (2004). Spatial memory and perspective taking. *Memory & Cognition*, *32*(3), 416–426.

Sheridan, T. B. (2016). Human–robot interaction: Status and challenges. *Human Factors*, *58*(4), 525–532.

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, *35*(3), 371–391.

Taylor, S., Wang, M., & Jeon, M. (2023). Reliable and transparent in-vehicle agents lead to higher behavioral trust in conditionally automated driving systems. *Frontiers in Psychology*, *14*, 1121622.

Tenbrink, T., Fischer, K., & Moratz, R. (2002). Spatial strategies in human–robot communication. *KI*, *16*(4), 19–23.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124–129.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.

van der Ham, I. J., Dijkerman, H. C., & van Stralen, H. E. (2021). Distinguishing left from right: A large-scale investigation of left–right confusion in healthy individuals. *Quarterly Journal of Experimental Psychology*, *74*(3), 497–509.

Vianello, L., Penco, L., Gomes, W., You, Y., Anzalone, S. M., Maurice, P., Thomas, V., & Ivaldi, S. (2021). Human–humanoid interaction and cooperation: A review. *Current Robotics Reports*, *2*(4), 441–454.

Vingerhoets, G., & Sarrechia, I. (2009). Individual differences in degree of handedness and somesthetic asymmetry predict individual differences in left-right confusion. *Behavioural Brain Research*, *204*(1), 212–216.

Wilcock, G., & Jokinen, K. (2022). Conversational AI and knowledge graphs for social robot interaction. In *2022 17th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 1090–1094). IEEE.

Wilke, F., Bender, A., & Beller, S. (2019). Flexibility in adopting relative frames of reference in dorsal and lateral settings. *Quarterly Journal of Experimental Psychology*, *72*(10), 2393–2407.

Wolf, S. M. (1973). Difficulties in right-left discrimination in a normal population. *Archives of Neurology*, *29*(2), 128–129.

Xiao, C., Xu, L., Sui, Y., & Zhou, R. (2021). Do people regard robots as human-like social partners? Evidence from perspective-taking in spatial descriptions. *Frontiers in Psychology*, *11*, 578244.

Yadollahi, E., Couto, M., Dillenbourg, P., & Paiva, A. (2022). Do children adapt their perspective to a robot when they fail to complete a task? In *Interaction Design and Children* (IDC'22), June 27–30, 2022.

Yamashita, H. (2013). Self-rated right–left confusability and performance on the money road-map test. *Psychological Research*, *77*(5), 575–582.

Yohanan, S., & MacLean, K. E. (2012). The role of affective touch in human–robot interaction: Human intent and expectations in touching the haptic creature. *International Journal of Social Robotics*, *4*(2), 163–180.

Yu, S.-J., Hsueh, Y.-L., Sun, J. C.-Y., & Liu, H.-Z. (2021). Developing an intelligent virtual reality interactive system based on the ADDIE model for learning pour-over coffee brewing. *Computers and Education: Artificial Intelligence*, *2*, 100030.

Zhao, X., Cusimano, C., & Malle, B. F. (2016). Do people spontaneously take a robot's visual perspective? In *11th ACM/IEEE International Conference on Human–Robot Interaction* (pp. 335–342). IEEE.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Experiment 1: objectChoice perspective * partner
Table S2 Experiment 1: logRT perspective * partner
Table S3 Experiment 1: logRT perspective * partner * responderGroup
Table S4 Experiment 2: objectChoice perspective * partner
Table S5 Experiment 2: logRT perspective * partner
Table S6 Experiment 2: logRT perspective * partner * responderGroup
Table S7 Experiments 1 and 2 combined: objectChoice perspective * partner
Table S8 Experiment 3: objectChoice perspective * partner
Table S9 Experiment 3: logRT perspective * partner
Table S10 Experiment 3: logRT perspective * partner * responderGroup
Figure S1 Experiment 1 different perspective trials: Relationship between participants' age and their percentage of object choice from their own avatar's perspective (i.e., egocentric perspective taking).
Table S11 Experiment 1: objectChoice perspective * partner * age
Figure S2 Experiment 2 different perspective trials: Relationship between participants' age and their percentage of object choice from their own avatar's perspective (i.e., egocentric perspective taking).
Table S12 Experiment 2: objectChoice perspective * partner * age

Figure S3 Experiment 3 different perspective trials: Relationship between participants' age and their percentage of object choice from their own avatar's perspective (i.e., egocentric perspective taking).

Table S13 Experiment 3: objectChoice perspective * partner * age

Data S1