


## Signatures of cognitive difficulty in perspective-taking: is the egocentric perspective always the easiest to adopt?

Alexia Galati, Anthi Diavastou & Marios N. Avraamides


To cite this article: Alexia Galati, Anthi Diavastou & Marios N. Avraamides (2017): Signatures of cognitive difficulty in perspective-taking: is the egocentric perspective always the easiest to adopt?, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2017.1384029](https://doi.org/10.1080/23273798.2017.1384029)

To link to this article: <http://dx.doi.org/10.1080/23273798.2017.1384029>

 View supplementary material 

 Published online: 06 Oct 2017.

 Submit your article to this journal 

 Article views: 17

 View related articles 

 View Crossmark data 

REGULAR ARTICLE



## Signatures of cognitive difficulty in perspective-taking: is the egocentric perspective always the easiest to adopt?

Alexia Galati <sup>a,b</sup>, Anthi Diavastou<sup>a</sup> and Marios N. Avraamides<sup>a,c</sup>

<sup>a</sup>Department of Psychology, University of Cyprus, Nicosia, Cyprus; <sup>b</sup>Cognitive and Information Sciences, University of California Merced, Merced, CA, USA; <sup>c</sup>Centre for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus

### ABSTRACT

In a series of experiments we examined factors that contribute to the difficulty of spatial perspective-taking and influence perspective selection. Listeners received instructions to select an object from a speaker whose depicted position varied (Experiments 1, 2, 2B). Responding from the speaker's perspective was slower than responding egocentrically, and was slower at large oblique offsets (135°, 225°) than at the maximum offset (180°). Experiment 2B confirmed that this was not due to the number of objects in configurations. Experiment 3 suggested that the ease of adopting the imagined egocentric perspective depended on its alignment with the sensorimotor perspective. Still, perspective preference was not influenced by the documented cost of adopting perspectives, but rather by social attributions (e.g. believing that the partner was the experimenter, Exp 1, vs. another participant, Exp 2, 2B, 3). These findings have implications for understanding behaviour in contexts where interlocutors interact remotely while adopting disembodied perspectives.

### ARTICLE HISTORY

Received 21 May 2017  
Accepted 14 September 2017

### KEYWORDS

Perspective-taking; audience design; mental rotation; spatial cognition; eye-tracking

### Introduction

In a variety of daily tasks that involve coordinating with others, we make choices – whether explicitly or implicitly – about the perspective from which to produce or interpret utterances. This is perhaps most obvious when we convey spatial information to one another, as for example when providing route directions to a prospective guest to our house, when jointly setting the table for dinner, or when figuring out which direction to move our body in yoga class. In many of these scenarios there are alternative options for conceptualising and, thus, describing or interpreting spatial relationships (e.g. Levinson, 2003). For instance, when providing route directions, we may adopt the perspective of the navigator (a so-called *route* perspective), or we may use absolute spatial terms (e.g. *north* and *south*) adopting an allocentric *survey* perspective (e.g. Taylor & Tversky, 1996). In a yoga class, when following the instructor's verbal cue to “rotate the feet to the right”, we have to decide whether this instruction is intended from our perspective, or from the instructor's counteraligned perspective at the head of the class. As this example suggests, the multiplicity of options for perspective can occasionally lead to ambiguity for how language users interpret spatial terms and map them onto space. The

competition of perspectives may sometimes lead to temporary misalignment in how people interpret the language or actions of others, and such misalignment can be corrected by monitoring others' behaviour (e.g. Keysar, Barr, & Horton, 1998). For example, upon realising that our “triangle pose” is misaligned with everyone else's in yoga class, we may summarily rotate our feet in the other direction to correct our orientation.

Given the multiplicity of perspective options, which often require explicit negotiation among interlocutors (e.g. Galati, Panagiotou, Tenbrink, & Avraamides, 2017; Garrod & Anderson, 1987; Schober, 2009), it's important to investigate the factors that motivate the selection of a particular perspective in a given context. One might predict that language users would generally opt for the perspective that is easiest to adopt. Still, what makes a perspective relatively “easy” or relatively “difficult” to adopt remains underexplored (Avraamides, Hatzipanayioti, & Galati, 2015). In the present study, we tease apart some of the factors that could contribute to the difficulty of perspective-taking and could thus influence perspective selection; we use spatial reasoning as our domain of inquiry.

An array of evidence about how people coordinate in both spatial and non-spatial tasks suggests that the *egocentric* perspective is often the easiest to adopt. In

**CONTACT** Alexia Galati  agalati@ucmerced.edu,  alexia.galati@gmail.com

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/23273798.2017.1384029>

© 2017 Informa UK Limited, trading as Taylor & Francis Group

visuospatial tasks, one's "egocentric perspective" refers to the perspective embodied by the self, or else representing the self in a depicted or imagined environment. The egocentric perspective often coincides with one's sensorimotor perspective – the perspective that encodes self-to-object relations in the immediate environment. But the egocentric perspective can also be dissociated from the sensorimotor perspective, in space and time, and instead stand for a representation of the self in a disembodied environment. This is, for example, the case when experiencing or apprehending the viewpoint of a moving avatar representing the self in a virtual environment displayed on an interface (as in a video game), or when reasoning about prior sensorimotor experiences (as when recollecting the view out of the window of our childhood home). In non-spatial tasks, the "egocentric perspective" can be defined more broadly to capture one's egocentric knowledge – knowledge derived not only through vision, but also through other modalities, including language – as well as one's beliefs, attitudes, and emotions.

The view that the egocentric perspective is easier to adopt than other perspectives is compatible with the proposal that people use their own knowledge or perspective as a proxy for the knowledge or perspective of others. For example, priming is thought to facilitate how people achieve converging perspectives by supporting their alignment across various levels of linguistic representation (Pickering & Garrod, 2004). A related account of adaptation in dialogue is that the egocentric perspective (including egocentric knowledge and beliefs) is the default perspective during the early stages of processing. According to Keysar and colleagues (Horton & Keysar, 1996; Keysar, Barr, Balin, & Paek, 1998; Keysar, Barr, & Horton, 1998; Shintel & Keysar, 2009), language users initially default to using egocentric information and consider information about the partner only later, as when needing to repair a misunderstanding (see however, Brennan & Hanna, 2009; Metzger & Brennan, 2003; Ryskin, Brown-Schmidt, Canseco-Gonzalez, Yiu, & Nguyen, 2014).

There is evidence in support of this view in spatial tasks as well. In a computer-based task in which listeners responded to spatial instructions (e.g. "Give me the folder on the left") that were ambiguous in some visual contexts, listeners who responded predominately from the egocentric perspective (egocentric responders) were faster than those responding from the partner's perspective (other-centric responders) (Duran, Dale, & Kreuz, 2011). Moreover, other-centric responders experienced interference from the egocentric perspective, as evidenced by deviations of the mouse cursor toward the competitor "egocentric" object choice and other signatures of their mouse movements.

In contrast, other work provides evidence that adopting the perspective of a social partner need not incur a cognitive cost. Some studies, in fact, show that the social partner's perspective – e.g. what that person believes (Kovács, Téglás, & Endress, 2010) or what they can perceive (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010) – is hard to ignore, even when that perspective is not relevant to the task and may interfere with making judgments from the egocentric perspective. In a study involving visual perspective-taking, when participants had to judge the number of dots they could see while ignoring the perspective of an avatar, they were slower and made more errors on trials that involved a disparity in perspectives (i.e. when the avatar could see a different number of dots) (Samson et al., 2010). This has been taken as evidence that a social cue, such as another's visuospatial perspective, is processed automatically (see also Tversky & Hard, 2009). Moreover, Ryskin and colleagues (2014) have shown that listeners can readily appreciate their conversational partner's spatial perspective without showing any evidence of bias toward an egocentric interpretation, even when that perspective is counteraligned from their own.

In addition, in the domain of spatial memory, there is accruing evidence that, although the egocentric, initially-experienced viewpoint can determine how people organise and maintain spatial information in memory (Shelton & McNamara, 2001), that perspective can be overridden. Various contextual and even social factors, including the environment's geometry (Shelton & McNamara, 2001), the intrinsic symmetry of the spatial configuration (Li, Carlson, Mou, Williams, & Miller, 2011; Mou & McNamara, 2002), intrinsic features of the constituent objects of the configuration (Marchette & Shelton, 2010), and the relative alignment of the social partners with environmental features (Galati & Avramides, 2015) can also determine the perspective or axis around which spatial information is organised in memory – the so-called "organizing direction" for that information, which exhibits facilitation during spatial judgments (Mou, McNamara, Valiquette, & Rump, 2004).

Thus, it remains unclear how flexibly people can reason from non-egocentric spatial perspectives, during early processing at least. When storing spatial information in longer-term representations, there is compelling evidence that people show flexibility in taking social and environmental cues into account, as described above. However, at shorter timescales, such as during the course of interpreting a spatial description, it is possible that there is an initial egocentric default in processing (cf. Ryskin et al., 2014; Ryskin, Wang, & Brown-Schmidt, 2016).

If the egocentric perspective has precedence during the early processing of spatial language, one prediction

would be that the misalignment between the egocentric and the adopted non-egocentric perspective would increase cognitive cost. This prediction is informed by findings from non-interactive spatial perspective-taking tasks, illustrating that performance decreases in terms of speed and accuracy as the misalignment between egocentric and non-egocentric perspectives increases (Zacks & Michelon, 2005). Although the underlying mechanism that subserves the adoption of imagined perspectives is still debated – for example, whether it involves a mental rotation of the self (Keehner, Guerin, Miller, Turk, & Hegarty, 2006; May, 2004; Zacks & Michelon, 2005) or whether it involves visual matching, especially at lower angles of misalignment (Kessler & Thomson, 2010) – the main assumption is that adopting imagined perspectives at larger offsets is more cognitively demanding.

Collaborative spatial tasks have often manipulated the misalignment between conversational partners explicitly (e.g. Duran et al., 2011; Galati, Michael, Mello, Greenauer, & Avraamides, 2013; Mainwaring, Tversky, Ohgishi, & Schiano, 2003; Schober, 1993, 1995, 2009). The studies that have focused on the *interpretation* of spatial descriptions typically find evidence consistent with mental rotation, with language users experiencing a greater processing cost when the adopted viewpoint is at larger degrees of misalignment (Duran et al., 2011; Mainwaring et al., 2003; Ryskin et al., 2016). For example, in Duran and colleagues (2011), as the partners' misalignment increased (from 0° to 90° to 180°), there was a steeper increase in response times when listeners responded from the partner's perspective relative to when they responded egocentrically. However, studies focusing on the *production* of spatial language often don't support mental rotation as the process involved in reasoning from the partner's viewpoint. For instance, Schober (1995) found that speakers produced more other-centric spatial expressions when they were misaligned than aligned with their partner, but found no difference between being misaligned by 90° or 180°.

Whether the focus is on production or comprehension, it may be limiting to consider only orthogonal offsets between partners (i.e. 90°, 180°, 270°: the offsets aligned with the canonical axes of the language user or the language user's representation of the self in the task environment). Findings from spatial memory suggest that it is easier to reason from perspectives that are orthogonal vs. oblique (i.e. not aligned with the canonical axes) relative to the organising direction used to represent and maintain spatial information (McNamara, 2003). Thus, the orthogonal perspectives that are used in many of the previously described studies may be relatively privileged, insofar as they are

aligned with the canonical axes of the language user, and may obfuscate the cognitive cost of perspective transformation at varying offsets. Indeed, in some of our earlier work that has focused on the production of spatial descriptions, we found that adopting imagined perspectives at large oblique offsets is likely more demanding than adopting the maximum offset of 180°, at least as suggested by the speakers' increased egocentric descriptions: speakers who were offset by 135° from their partners were more likely to adopt their own (vs. their partner's) perspective in descriptions, compared to when offset by 90° or 180° (Galati et al., 2013).

In the present work we examine, for the first time to our knowledge, whether perspective-taking from imagined oblique (vs. orthogonal) perspectives is indeed more difficult during the *interpretation* of spatial descriptions. This undertaking can clarify the process by which people transform their egocentric perspective to another person's perspective in collaborative tasks.

In order to further probe into the potential precedence of the egocentric perspective, we also examine the contribution of sensorimotor information from the body. Specifically, we ask whether any advantage of the egocentric perspective is due to the sensorimotor encoding of information in terms of self-to-object relations (Avraamides & Kelly, 2008; De Vega, 2008). Does the egocentric perspective retain its advantage when it is abstracted and dissociated from the sensorimotor perspective of the language user? This question is especially timely, given that technological applications increasingly enable users to interact with one another remotely, through virtual environments or virtual shared spaces that offer different affordances and visualisations of the users' viewpoints. For example, in the context of a collaborative videogame representing perspectives on a 2D projection, the player's sensorimotor perspective can be dissociated both from her depicted egocentric perspective and her fellow players' perspective in the virtual environment. In that context, is the egocentric perspective still the easiest to adopt?

Some evidence suggests that the advantage of the egocentric perspective may, indeed, be due to its habitual coincidence with sensorimotor perspective. For example, individuals' performance can exhibit either sensorimotor facilitation or interference, depending on the alignment between their physical orientation in space and their imagined perspective (May, 2004). Similarly, the congruence between people's physical orientation in their immediate environment and their imagined perspective, as when reading about a protagonist moving in a described environment, influences the ease with which they make judgments about locations in the described environment (Hatzipanayioti, Galati, & Avraamides,

2016a, 2016b). These findings motivate the prediction that dissociating the egocentric and sensorimotor perspectives would influence the ease of adopting the egocentric perspective, and by extension perhaps make the selection of that perspective less likely.

So far, for expository purposes, we have assumed that, in a given task, people would select the perspective that is the least cognitively demanding. However, in social contexts, this is a false assumption, since language users routinely adopt imagined perspectives, including their conversational partner's, that have associated cognitive costs. According to Clark and Wilkes-Gibbs (1986), interlocutors assume shared responsibility for mutual understanding, and thus take into account the relative difficulty of perspective-taking—for themselves and for their conversational partners – selecting strategies that maximise their efficiency of communication while minimising their collective effort. By adhering to what has been termed as *the principle of least collaborative effort* (Clark, 1996; Clark & Wilkes-Gibbs, 1986), interlocutors may invest the cognitive effort to adopt another's perspective, if they believe that it will maximise their effectiveness in the joint task (e.g. Duran et al., 2011; Galati & Avraamides, 2015; Schober, 1995, 2009).

Specifically, when people perceive their conversational partner as being limited in terms of contributing to the task, they are more likely to be accommodating and to adopt the partner's perspective, despite its cognitive cost. In the yoga scenario described earlier, to the extent that the yoga instructor bears greater responsibility for mutual understanding – being the person responsible for directing the students through various movements, while they are more limited in how they can contribute to the design of the class – he is likely to provide verbal cues from the students' perspective (i.e. intending "turn the feet to the right" from the students' spatial perspective). Such adaptation is supported empirically. For example, when speakers plan routes for an imaginary addressee who is unfamiliar with the environment (vs. for themselves), they use different strategies, by elaborating their spatial descriptions more (using more words and details), simplifying the complexity of the planned routes (selecting routes along fewer, larger, and more prominent streets), and increasing their salience (referring to more landmarks) (Hölscher, Tenbrink, & Wiener, 2011). Moreover, speakers are more likely to use descriptions from the partner's perspective (e.g. "to your left") and less likely to use egocentric ones (e.g. "to my right") when describing spatial configurations to an imaginary partner, as opposed to a real one, with whom they can interact contingently (Schober, 1993). Similarly, listeners are more likely to interpret ambiguous spatial descriptions from the

partner's perspective when they believe the partner does not know their viewpoint, whereas they are more likely to interpret such descriptions egocentrically when they believed their partner is real (vs. simulated) (Duran et al., 2011). These findings underscore that social cues about the partner can provide pragmatic motivation to override the egocentric perspective, affecting language users' perspective strategy.

### **The present study**

In the present set of experiments, we tease apart the following factors that may influence the difficulty of perspective-taking in interactive spatial tasks: perspective strategy (egocentric vs. other-centric), the misalignment between the egocentric and the partner's perspectives, and the misalignment between the egocentric perspective as depicted in the task environment and the individual's sensorimotor perspective.

We use an experimental paradigm similar to that of Duran et al. (2011), in which listeners responded to spatial instructions from a conversational partner to select an object from a top-down view of a table. As we have noted earlier, in that set of experiments, instructions were ambiguous in some visual contexts (e.g. "Give me the folder on the left"), and based on the listeners' object choices on those ambiguous trials, listeners were classified as egocentric, other-centric, or mixed responders. Across experiments, Duran and colleagues (2011) examined whether the listeners' attributions about the speaker influenced their perspective strategies (the attributions being that the speaker was a simulated partner who knew their position around the table, Study 1; that the speaker was a simulated partner who didn't know their location around the table, Study 2; and that the speaker was another Amazon's Mechanical Turk worker, Study 3). As we have reported at the end of the previous section, the researchers did indeed find evidence that social attributions about the partner's role or beliefs influenced perspective strategy.

In the present studies, we use similar materials as Duran and colleagues (2011): listeners here also follow instructions to select an object from a tabletop, and some of these instructions are ambiguous, permitting the classification of listeners into different types of responders. As with Duran and colleagues (2011), one of our undertakings is to investigate whether, during the interpretation of spatial descriptions, adopting another person's perspective is, indeed, more cognitively demanding than the egocentric perspective. To compare the processing of these perspectives, given Duran et al.'s (2011) finding that the degree of egocentric and other-centric responding depends on simple attributional

cues about the partner, we manipulated social attributions about the partner (Experiments 1 and 2) in order to elicit different distributions of egocentric and other-centric responses (i.e. object selections based on the egocentric perspective vs. the partner's perspective). However, whereas Duran and colleagues (2011) led participants recruited over Amazon's Mechanical Turk to believe that they were either interacting with a real remote partner (another worker) vs. a simulated partner, here, we lead participants to believe that the partner with whom they are interacting in the laboratory is another participant vs. the experimenter. We predict that when participants believe that their partner is more limited and has real informational needs (i.e. is another participant), other-centric responding will be higher.

Given these perspective choices, we examine how egocentric and other-centric perspectives differ in terms of their associated cognitive demands. Duran and colleagues (2011) did so by sampling the listeners' mouse trajectories and deriving measures, in addition to response times, that reflected the cognitive cost of perspective-taking (e.g. directional shifts and acceleration components in the mouse movements taken to reflect decision complexity due to interference from the competing object or due to hesitation). Here, in addition to response times, we consider measures obtained through eye-tracking; specifically, the listeners' first gaze fixation. Gaze fixations can clarify whether, during a particular response, the competing perspective is co-activated and causes interference during the early moments of processing. For example, when responding other-centrally, listeners may still fixate the egocentric option first.

Our second undertaking and a novel contribution of the present work is to probe the processes involved in perspective transformation and maintenance by considering the role of the misalignment between the egocentric and other-centric perspectives, thus going beyond a broad replication of Duran et al.'s (2011) work. In particular, we examine whether the pattern of response times of other-centric responders (i.e. those selecting objects based on the other-centric for the majority of the trials) is fully accounted for by mental rotation: is the size of angular disparity all that matters (predicting slowest other-centric responses at the maximum orthogonal offset of 180°), or does the type of angular disparity and its alignment with one's canonical axes matter instead (i.e. whether the partner is at an oblique vs. orthogonal offset)? In contrast to prior studies (e.g. Duran et al., 2011; Mainwaring et al., 2003), our design materials include perspective-taking situations in which partners are depicted to be misaligned both

by oblique offsets (135°, 225°) and orthogonal offsets (90°, 180°, 270°).

In Experiment 2B, we seek to unconfound the misalignment between the partners' positions and the number of objects in the configuration, in order to corroborate that the patterns observed in Experiments 1 and 2 are indeed due to the misalignment of perspectives and not due to the complexity of the configuration.

A final novel contribution of the present work is that we examine the degree to which any advantage of the egocentric perspective relies on its habitual alignment with the sensorimotor perspective: in Experiment 3, we dissociate the egocentric and sensorimotor perspectives by manipulating their misalignment. As we have noted, although in many contexts the sensorimotor perspective and the egocentric perspective are identical, involving the encoding of information in the immediate environment in terms of self-to-object relations, in other contexts – including those involving the representation of perspectives in remote or depicted scenes – the sensorimotor and egocentric perspectives can be dissociated. Thus, by manipulating the misalignment of these perspectives in Experiment 3, we can assess whether this misalignment increases the cognitive cost of adopting the egocentric perspective, and by extension influences the likelihood of adopting that perspective.

## Experiments 1 & 2

In order to induce egocentric and other-centric responding to different degrees, in Experiments 1 and 2, we manipulated the listeners' social attributions about the speaker (i.e. what they believed about who the speaker was and therefore what the speaker's informational needs were), using a cover story similar to Duran et al. (2011). In Experiment 1 listeners were led to believe that they were interacting with the experimenter, whereas in Experiment 2 they were led to believe that they were interacting with a naïve participant with real informational needs. We predicted that, insofar as the partner's informational needs are salient and relevant, listeners would be more likely to interpret spatial descriptions from their partner's viewpoint when they believed that those needs were relatively substantial vs. more trivial (Experiment 2 vs. Experiment 1).

Within these anticipated distributions of perspective choices, we were interested in investigating the role of the partners' misalignment on the ease of responding, as reflected by participants' response times. We reasoned that other-centric responders (but not egocentric ones) would show longer response times at the offsets at which perspective-taking is most difficult. If spatial perspective-taking is more difficult at oblique offsets (e.g.

调查者的反应时间所反映出的合作伙伴的错位在任务内容程度上的作用。推断：以他人为中心的参与者会在换位思考表现出更长的反应时间。

如果在斜线偏移处理进行空间换位比较困难，那么在偏移处的响应时间是最长的。如果换位视角随角度差视差的增加而增加，那么应在180度时响应时间最长。

McNamara, 2003), response times would be the longest at those offsets. If perspective-taking increases with angular disparity, as predicted by mental rotation, response times would be the longest at the maximum offset (180°).

We also set out to examine whether egocentric responding would increase at the offsets at which processing is most difficult (whether the oblique offsets, or the maximum offset). In other words, would people fluctuate in their perspective selection depending on the difficulty of the trial, or would they stabilise on an invariant perspective strategy across headings (one that is modulated by attributional cues about the partner)?

Finally, we asked whether during a particular mode of responding (egocentric or other-centric), there would be evidence of interference from the competing perspective. In particular, to the extent that listeners might default to an egocentric interpretation (e.g. Keysar, Barr, & Horton, 1998), we examined whether they would fixate the egocentric object first, before making an other-centric choice.

## Method

### Participants

In Experiment 1, 26 undergraduate students from the University of Cyprus served as listeners (24 female). In Experiment 2, 26 undergraduate students from the University of Cyprus served as listeners (22 female); none had participated in Experiment 1. They participated for course credit, or else as unpaid volunteers.

### Stimuli

The configurations and prerecorded instructions used were the same in both experiments.

### Configurations

Each computer-based trial displayed the top-down view of a circular table with a configuration of two or three objects (CDs). The participant's (i.e. the listener's) perspective (0°) was indicated by a red arrow, whereas the speaker's perspective, which changed across trials (0°, 90°, 135°, 180°, 225°, 270°), was indicated by a human figure (see Figure 1(a,b)).

Of the 56 experimental trials,<sup>1</sup> 20 trials were constructed such that the interpretation of the speaker's request would result in the same object choice from both the listener's perspective and the speaker's perspective (control trials), and 36 trials were constructed such that interpretation from the listener's and the speaker's perspective would result in different object

choices (ambiguous trials). The 10 practice trials were all control trials.

Configurations with 2 objects involved three types of arrangements: a horizontal arrangement of the two objects around the table's centre, a vertical arrangement, and a diagonal arrangement (with CDs on the bottom left and top right, or on the top left and bottom right, as in Figure 1(a)). In configurations with 3 objects, the CDs always formed a right angle shape, either with a hypotenuse along the central horizontal axis of the table (i.e. with two CDs horizontally around centre and one on the vertical axis, either at the top or bottom), or with a diagonal hypotenuse (e.g. with one CD on the top left, and the others at bottom right and bottom left, as in Figure 1(b)). Ambiguous trials from the orthogonal offsets (90°, 170°, 180°) were two-object trials, whereas those from the oblique offsets (135°, 225°) were 3-object trials. All control trials were two-object trials.

Of the control trials, 12 involved cases for which the speaker's perspective was the same as the listener's (i.e. 0°), and another 8 involved cases for which the speaker was at 90° or 270° (see Figure 1(a)). Of the ambiguous trials, 12 involved cases for which the speaker was at 180° (2 vertical, 2 horizontal, and 8 diagonal arrangements), 8 involved cases for which the speaker was at 90° or 270° (all were 2-object configurations with a diagonal arrangement; see Figure 1(a)), and 16 involved cases in which the speaker was at 135° or 225° (all were 3-object configurations, half with a horizontal and half with a diagonal hypotenuse; see Figure 1(b)).

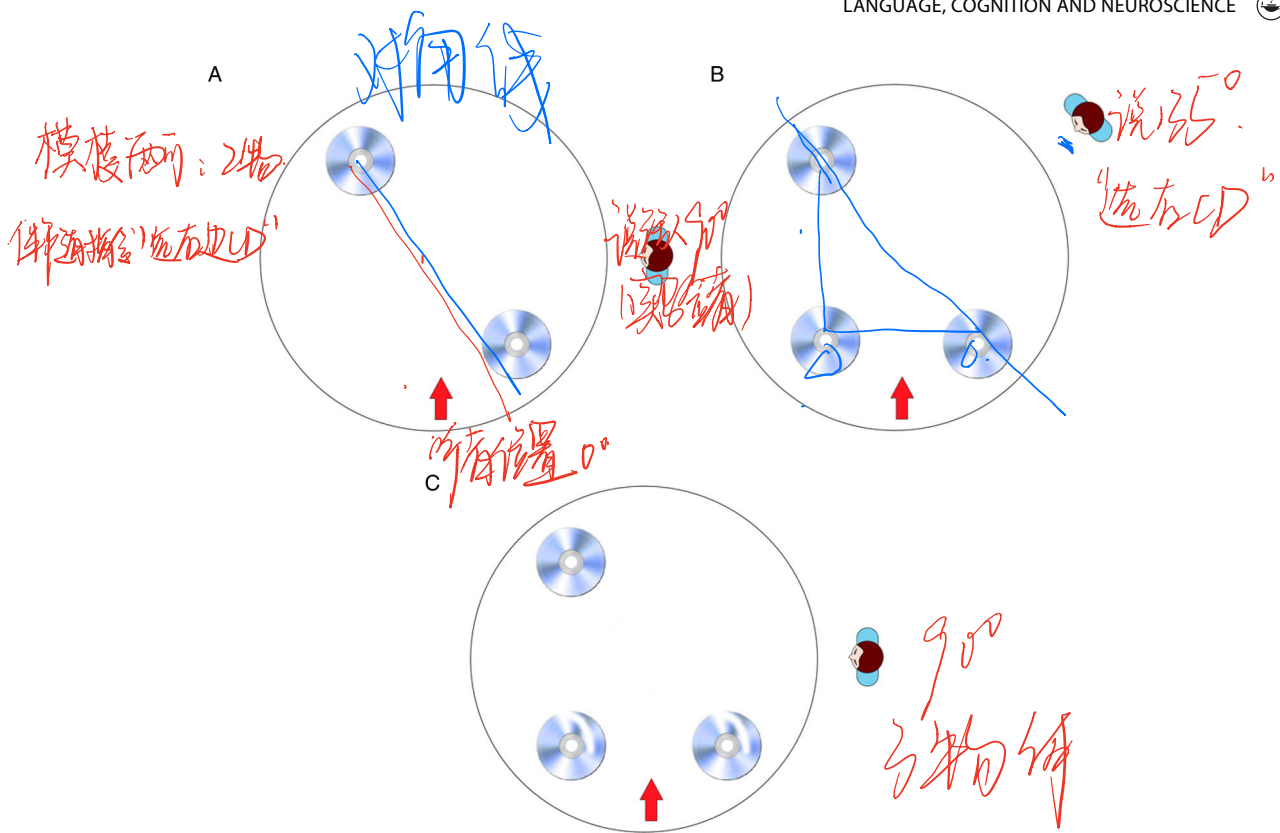
The order in which the trials were presented was randomised for each participant.

### Prerecorded instructions

Prerecorded instructions included one of four directional terms, having the form: "Select the CD that is on the right / left / top / bottom" (in Greek, *Dialekse to CD pu ine deksia / aristera / pano / kato*). There were 14 trials for each type of directional term (right, left, front, back) in the experimental trials: 28 trials with spatial terms on the lateral axis (i.e. left-right) and 28 with terms on the sagittal axis (i.e. top-bottom). The female experimenter produced nine different recordings of each instruction, uttered in a naturalistic way, to support the cover story that she spontaneously gave instructions in real time during the experiment. The instructions were controlled to have exactly the same duration (2s).

### Apparatus

Eye movements were measured with a Tobii T120 near infrared eyetracker with accuracy 0.5 and sampling rate



**Figure 1.** Examples of trials from Experiments 1–2, and 2B. The configuration in (a) illustrates an ambiguous 2-object trial when accompanied by the instruction “Select the CD on the right”. It can also illustrate control trial when accompanied by the instruction “Select the CD on the top”. The speaker is depicted by the human figure to be at 90°, whereas the arrow indicates the listener’s position at 0°. The configuration in (b) illustrates an ambiguous 3-object trial, with the speaker depicted at 135°, when accompanied by “Select the CD on the right”. The configuration in (c), illustrates an ambiguous 3-object trial under the same instruction with the speaker at 90°, used in Experiment 2B.

60 Hz (Tobii, Stockholm, Sweden). The stimuli were presented on a 19” monitor. A standard five-point calibration was used (Gredebäck, Johnson, & von Hofsten, 2010). Trials were presented in a different random order to each participant using E-Prime 2.0 (Psychology Software Tools, Inc).

## Procedure

### Procedure of Experiment 1

When participants arrived at the lab, the experimenter (E1) obtained informed consent for participation. Participants were told that they would have to select objects from a configuration presented on the computer screen, according to instructions they would receive from the experimenter, who would be stationed in an adjacent room. Participants were informed that, in the configurations, the experimenter would be represented as a human figure around the top-down view of a table, and could occupy either the same or a different position from their own, which would be indicated by an arrow. Participants were told that, while their own position would remain constant across trials, their

partner’s, as indicated by the human figure, would change across trials.

Before the experiment began, participants were taken to the adjacent room, where they were shown the experimenter’s networked computer screen, and their headset and microphone. Participants were informed that the experimenter would be looking at the exact same configuration on her screen as they did and that she would be giving them instructions to select an object with their mouse. It was made clear to participants that, on each trial, by virtue of viewing the same configuration, the experimenter would know their (the participant’s) location around the table.

At this point, the cover story was introduced: participants were told that, due to a technical problem with the networked audio equipment, they would be able to hear the experimenter providing instructions but that the experimenter would not be able to hear them or provide any feedback on their responses.

Upon returning to the testing room, the participants sat at a comfortable distance from the computer monitor and the eyetracker, which was situated under the raised monitor. To reinforce the cover story and



establish a pretext for one-way communication, participants were guided through a staged sound-check. After putting their headset on, participants waited for a few moments for the experimenter to move to the adjacent room, and then pressed the space bar of their keyboard. They then heard the experimenter saying, "Can you hear me?". When they pressed the spacebar to respond affirmatively, they heard the experimenter say, "Can you hear me OK?", and pressed the spacebar again. Both of these messages were pre-recorded.

The experimenter then returned to the testing room and guided participants through the eyetracker calibration. Participants were told that the eyetracker's purpose was to ensure that they would always fixate on the centre of the screen at the beginning of each trial. Once calibration was successfully completed, experimental instructions were provided in more detail. Participants were told that in order to initiate a trial, they had to fixate a cross that appeared on the centre of the screen, inside a red flashing square. When the red square stopped flashing, it would disappear and be replaced on the screen with the top-down view of an empty round table. They would then listen to an instruction. Then, two or three CDs would appear on the table, along with a human figure indicating the experimenter's position around the table and an arrow indicating the participant's perspective. Participants were told that, on the basis of the instruction, they had to move their cursor to the CD of their choice and left-click with their mouse.

After 10 practice trials, the experimenter returned to the testing room to ask whether the participant had any remaining questions. If participants asked explicitly which perspective they were supposed to adopt on the trials, the experimenter appeared ignorant, responding with "This is not my experiment so I don't know. Do what you think is best."

The experimenter then moved to the adjacent room and initiated the experimental trials. All participants heard the same prerecorded instructions during the experiment. For each trial to be initiated, participants had to fixate on the cross at the centre of the screen. During the experimental trials, the participants' eye movements were tracked and their response times were recorded. After the participants completed the 56 experimental trials, participants were then debriefed, and were informed about the cover story for one-way communication and about the tracking of their eye-movements. Experimental sessions lasted approximately 40 min.

### Procedure of Experiment 2

The procedure was nearly identical to that of Experiment 1, except for the fact that the experimenter (E1) of

Experiment 1 now served as a confederate and was introduced as a naïve participant by a new experimenter (E2). Thus, after participants provided informed consent for participation, they were told by E2 that they would have to select objects from a configuration presented on the computer screen, following the instructions they would receive from the other participant (E1), who would be in the adjacent room. As with Experiment 1, participants were informed that in the configurations the other participant would be represented as a human figure around the top-down view of a table, whose position could change across trials, whereas their own position would be indicated by an arrow and would remain fixed across trials.

As with Experiment 1, participants were taken by E2 to the adjacent room to see their partner's set-up, and were told that their partner would be viewing identical configurations on her screen and that she would be giving them instructions to select an object with their mouse. The same cover story of a technological problem was introduced to account for the one-way communication between E1 and the participant. E2 guided the participant through the same "sound check" procedure with E1 to further reinforce this cover story. E2 then guided the participant through the eyetracker calibration, and gave the same instructions as in Experiment 1, with the exception that the human figure was now meant to represent the position of the presumed naïve participant (E1).

The experimental trials were identical to those in Experiment 1. E2 remained outside the testing room during the experimental trials. After their completion, E2 returned to the testing room and participants were debriefed. The debriefing included a disclosure about the use of a confederate, the cover story for one-way communication, and the tracking of participants' eye-movements.

### Design and analyses

To summarise our experimental design, the social attribution for the speaker was manipulated across experiments (i.e. the speaker was introduced as the Experimenter in Experiment 1 vs. as another participant in Experiment 2); in each experiment participants were classified as a certain kind of responder based on their proportion of egocentric responses on ambiguous trials (i.e. *egocentric responder* for participants selecting the egocentric object on 70% or more of ambiguous trials, *other-centric responder* for participants selecting the egocentric object on 30% or fewer of ambiguous trials, or *mixed responder* for the remainder of participants); all participants responded to instructions that varied in

terms of trial type (ambiguous vs. control), speaker position (for ambiguous trials: 90°, 135°, 180°, 225°, 270°; for control trials: 0°, 90°, and 270°), and type of spatial instruction (on a lateral axis: *left-right*, on a sagittal axis: *top-bottom*). For each trial, the obtained data concerned whether the listener made an egocentric vs. other-centric object selection, whether the listener's first gaze fixation was on the egocentric vs. other-centric object, and the listener's response time for that trial.

In order to assess the listeners' performance on the task, we used the lme4 library (Bates, Maechler, et al., 2015) in R (R Core Team, 2016) to construct separate generalised linear mixed effects models for performance on ambiguous and control trials. In the Results section, next, we focus only on the results of ambiguous trials, and provide the full summary of those models in Appendix A. We present the results for control trials in Appendix B, since control trials served primarily as a check that participants were not responding randomly, given that there was only one "correct" CD option that corresponded to both the egocentric and other-centric choice. Ambiguous trials, on the other hand, were the trials of theoretical interest, since they presented the opportunity for listeners to make an egocentric or other-centric object choice, and in aggregate reflected the listener's perspective strategy.

For models with perspective choice on ambiguous trials as the dependent measure we started by including, as fixed factors, the listeners' social attribution about the partner (i.e. Experiment as a factor: Experiment 1, Experiment 2), the type of offset of the speaker's position (orthogonal: 90°, 180°, 270°, oblique: 135°, 225°), and instruction type (sagittal: *top-bottom*: lateral: *left-right*). For models with response time as the dependent measure, we also included perspective preference (egocentric, other-centric, mixed) as a fixed effect, along with its interaction with the other factors, unless specified otherwise.

Regarding our decision to condense the four spatial instructions (*top*, *bottom*, *left*, *right*) into two instruction types based on their axes (i.e. sagittal vs. lateral), we did so not only to simplify the complexity of the models but for theoretical reasons as well, as spatial expressions belonging to the two axes could differ in this task. This is in light of evidence that mapping *left* and *right* to appropriate regions of space is slower than *front-back* (or *top-bottom*), perhaps due to the fact that the lateral axis is highly symmetric (relative to a person's body), which makes it more difficult to differentiate *left-right* relative to linguistic terms associated with the sagittal axis (e.g. Avraamides & Sofroniou, 2006; Franklin & Tversky, 1990).

As random effects in the mixed effects models, we had intercepts for participants, and following the recommendations of Barr, Levy, Scheepers, and Tily (2013), we started with the full random effect structure, which included random slopes for type of offset for ambiguous trials (or the speaker's exact position for control trials), instruction type, and their interaction (Perspective preference was not used as a nested random slope term, since it was consistent within participants – a between-participants factor.). When the models did not converge, we simplified them by removing terms from the random effect structure, starting with the higher order terms (see the recommendations of Bates, Kliegl, Vasishth, & Baayen, 2015), until the most complex model that converged was obtained.

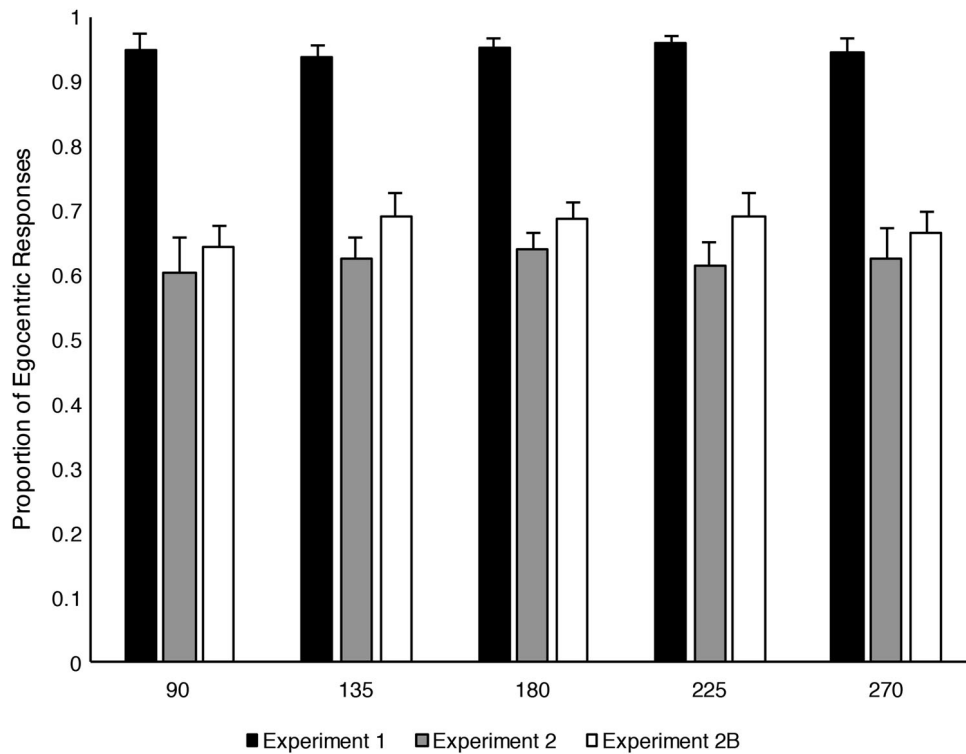
When visual inspection of residual plots for the models revealed deviation from normality and homoscedasticity, the dependent variable was log-transformed; this was the case for response times. For binary dependent variables (i.e. the selection of the egocentric object choice for ambiguous trials, and of the correct object choice for control trials), we used mixed logistic regression models with binomial error structure (Jaeger, 2008).

## Results

### Perspective choices

As illustrated in Figure 2, across experiments the degree of egocentric responding differed according to the listeners' attributions about the speaker: the proportion of egocentric object selections differed significantly across the two experiments, ( $\beta = -5.03$ ,  $SE = .40$ ,  $z = -12.73$ ,  $p < .001$ , see Table 1 of Appendix A). In Experiment 1, where listeners believed they were interacting with the experimenter, they responded overwhelmingly from their own perspective ( $M = 95\%$ ,  $SD = 22\%$ ), being more likely to interpret ambiguous instructions egocentrically than from their partner's perspective (see Figure 2). In Experiment 2, where listeners believed that the conversational partner was a naive participant, they adopted their own perspective less frequently than in Experiment 1, as illustrated in Figure 2. Overall, they responded egocentrically on 63% ( $SD = 48\%$ ) of the ambiguous trials.

Across the two experiments, the position of the speaker did not influence the listeners' perspective choices: the type of heading (orthogonal vs. oblique) at which the speaker was depicted was not a significant predictor of perspective choice ( $\beta = .13$ ,  $SE = .22$ ,  $z = .59$



**Figure 2.** Proportion of egocentric object selections on ambiguous trials across the speaker's position in Experiments 1, 2, and 2B.

$p = .55$ ). As shown in Figure 2, in each experiment, the listeners' aggregate perspective choices were consistent across trials. The type of instruction (lateral *left-right* vs. sagittal *top-bottom*) did not influence the listeners' perspective choices either ( $\beta = .15$ ,  $SE = .31$ ,  $z = .58$ ,  $p = .63$ ). The number of objects<sup>2</sup> in the configuration resulted in comparable distributions of egocentric and other-centric responses, as it can be observed in Table 1.

Finally, we examined the consistency of listeners' responses across ambiguous trials, classifying listeners as egocentric, other-centric, or mixed responders. For each participant, we computed the proportions of egocentric and other-centric responses on ambiguous trials. Following Duran and colleagues (2011), if proportion scores exceeded .70 for one of the two person-centred perspective categories, the participant was classified as member of that category; otherwise they were classified as a mixed responder. In Experiment 1, when listeners believed their partner to be the experimenter, only one participant was classified as an other-centric responder; the other 25 were all egocentric responders. In Experiment 2, when listeners believed their partner to be a naïve participant with real informational needs, there were 8 other-centric responders, 4 mixed responders, and 14 egocentric responders. The distribution of the responders across the two experiments was shown to be significantly different,  $\chi^2(2) = 12.55$ ,  $p < .01$ .

### Gaze fixations

In addition to perspective choices, we analysed participants' gaze fixations to determine whether they had considered the alternative perspective before responding either egocentrically or other-centrally. For each ambiguous trial we looked at the first fixation that fell on either the egocentric or the other-centric response option.<sup>3</sup> All nine possible locations at which objects could appear (i.e. the cells of a  $3 \times 3$  grid superimposed on the circular table) were defined as Areas of Interest (AOIs) and all gaze durations of 50 ms or more to an AOI were considered a fixation. For gaze analyses, we considered separately trials on which listeners made an egocentric vs. an other-centric choice, as we wished to examine the degree to which these choices were associated with the listeners' first gaze fixations.

For Experiment 1, we found that for trials in which listeners responded egocentrically, the first fixation was on the egocentric response choice in 93% of the trials, compared to 7% of the trials in which the first fixation fell on the other-centric response,  $t(24) = 20.01$ ,  $p < .001$ . In contrast, when listeners responded other-centrally, they looked first equally often to the egocentric (43%) and the other-centric response option (57%),  $t(24) = 1.06$ ,  $p = .30$ . These findings suggest that in Experiment 1, when listeners responded egocentrically, which was the majority of the time, they rarely considered the

**Table 1.** Means (and standard deviations) of the proportions of egocentric and other-centric perspective choices and response times (in ms) across two-object and three-object configurations for ambiguous and control trials in Experiments 1, 2, and 2B.

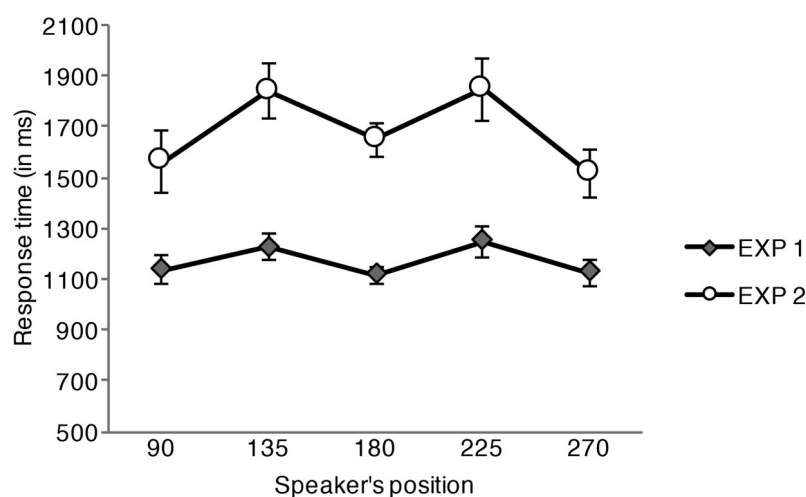
	Ambiguous trials			Control trials	
	Egocentric response	Other-centric response	RT	Correct Response	RT
Experiment 1					
2-object	.95 (.22)	.05 (.21)	1119 (596)	.98 (.13)	1086 (519)
3-object	.95 (.22)	.05 (.21)	1237 (798)	–	–
Total	.95 (.22)	.05 (.21)	1173 (698)	.98 (.13)	1086 (519)
Experiment 2					
2-object	.63 (.48)	.37 (.48)	1607 (1111)	.98 (.13)	1324 (729)
3-object	.62 (.49)	.37 (.48)	1844 (1684)	–	–
Total	.63 (.48)	.37 (.48)	1715 (1406)	.98 (.13)	1324 (729)
Experiment 2B					
2-object	.76 (.43)	.22 (.42)	1376 (1133)	.98 (.14)	1114 (620)
3-object	.76 (.43)	.23 (.42)	1391 (1075)	.99 (.08)	1124 (565)
Total	.76 (.43)	.22 (.42)	1383 (1108)	.99 (.12)	1118 (600)

other-centric perspective. Moreover, even when they responded other-centrally, they still considered the egocentric choice first nearly half of the time.

The gaze fixation results of Experiment 2 were similar to those of Experiment 1. When listeners responded egocentrically, they were much more likely to fixate first on the egocentric (92%) than the other-centric option (8%),  $t(25) = 6.64$ ,  $p < .001$ . In contrast, when responding other-centrally, listeners were more likely to fixate first on the other-centric option (57%) than the egocentric one (43%),  $t(25) = 2.91$ ,  $p < .01$ . Thus, even though in Experiment 2 listeners responded relatively more other-centrally than in Experiment 1, when they did give an other-centric response, they still frequently considered the egocentric perspective, nearly half of the time. And when they responded egocentrically, they still largely ignored the other-centric perspective as indicated by their gaze fixations.

### Response times

The listeners' social attributions about the speaker, which differed across Experiments, significantly predicted their response times in making an object selection ( $\beta = -.09$ ,  $SE = .03$ ,  $t = -3.26$ ,  $p < .01$ ; see Table 1 of Appendix A). In Experiment 1, listeners took on average 1173 ms ( $SD = 698$ ) to respond on ambiguous trials compared to 1715 ms ( $SD = 1406$ ) in Experiment 2. This difference is contextualised by the findings on perspective choices and gaze fixations reported above, under the assumption that adopting another's perspective is more cognitively taxing than responding egocentrically. As we saw, in Experiment 1, listeners responded overwhelmingly from their own perspective and rarely considered the partner's perspective, as indicated by their gaze fixations, whereas in Experiment 2 they were less likely to respond egocentrically, and when they responded other-centrally

**Figure 3.** Response times (in ms) on ambiguous trials across the speaker's position in Experiments 1 and 2.

they still frequently considered their egocentric perspective first.

Indeed, the listeners' perspective preference influenced response times significantly, as it significantly improved the fit of the model ( $\chi^2(8) = 305.59, p < .001$ ). Across both experiments, egocentric responders were more than one second faster ( $M = 1162, SD = 761$ ) than other-centric responders ( $M = 2282, SD = 1716$ ) and mixed-responders ( $M = 2306, SD = 1281$ ) on ambiguous trials. The differences between egocentric responders and the other responders were reliable, as shown in Table 1 of Appendix A (for other-centric responders vs. egocentric:  $\beta = .75, SE = .04, t = 15.10, p < .001$ , and for mixed responders vs. egocentric:  $\beta = 1.09, SE = .09, t = 11.93, p < .001$ ). As expected, other-centric responders were significantly influenced by whether the speaker was depicted at an oblique or orthogonal offset, as illustrated by the significant interaction of other-centric preference (vs. egocentric preference) and type of offset ( $\beta = -.12, SE = .05, t = -2.39, p < .05$ ).

In contrast to our findings about perspective choice, the speaker's position around the table – and specifically the type of offset the speaker was depicted at – did not significantly predict response times ( $\beta = -.06, SE = .02, t = -2.58, p < .001$ ). As observed in the sawtooth pattern of performance shown in Figure 3, at the oblique offsets of 135° and 225°, listeners were slower to respond than at the orthogonal offsets (90°, 180°, 270°). Although our analyses focus on the distinction between orthogonal and oblique offsets, in Figure 3, we present response times across all five of the speaker's positions, in order to illustrate visually the difference between oblique and orthogonal offsets, and underscore the point that responding on trials with the speaker at 180° (the maximum offset) did not take longer than on trials with the speaker at the oblique offsets (135° and 225°). This sawtooth pattern suggests that performance at 180° was not obfuscated by its grouping with the smaller orthogonal offsets (90°, 270°) – the effect of type of offset was not merely driven by faster responses at 90° and 270° alone.<sup>4</sup>

It is perhaps surprising that this difference between oblique and orthogonal offsets was observed in the response latencies of Experiment 1 ( $\beta = -.07, SE = .02, t = -3.18, p < .01$ , in a model of response times from Experiment 1 only), considering that listeners in Experiment 1 responded mostly egocentrically and fixated their gaze almost exclusively on the egocentric option first. This may suggest that listeners' automatically processed the speaker's position in Experiment 1, even though it was irrelevant to their perspective strategy.

However, what makes the interpretation of the effect of the speaker's position problematic, here, is that trials

with the speaker at oblique offsets (135° and 225°) always involved 3-object configurations, whereas those at orthogonal offsets (90°, 180°, 270°) always involved 2-object configurations, with listeners being overall 428 ms slower to respond on 3-object trials than 2-object ones. As shown in Table 1, there was a numerical difference in response latencies between 2-object and 3-object trials, independently of the attributions listeners made about their speaker across the two experiments. It is therefore possible that the observed cost in processing on trials with the speaker depicted at oblique offsets, even in Experiment 1, is due to the processing difficulty of 3-object trials relative to 2-object trials. In Experiment 2B, we aim to uncounfound whether the difficulty of 3-object trials was due to the number of objects or the speaker's position by including 3-object configurations when the speaker is at orthogonal offsets and 2-object configurations when the speaker is at oblique offsets. This would clarify whether listeners responding egocentrically processed the speaker's perspective, even when it was irrelevant.

Finally, we found that although the type of instruction (sagittal: *top-bottom*, vs. lateral: *left-right*) did not influence listeners' perspective choice, it did influence their response times significantly ( $\beta = -.10, SE = .02, t = -4.29, p < .001$ ), with sagittal trials being faster lateral ones. Other-centric responders were influenced by the effect of the type of instruction more than egocentric-responders ( $\beta = -.15, t = -2.85, p < .01$ ).

## Discussion

So far, our findings suggest that: (a) responding from another's perspective is cognitively more taxing than egocentric responding, yet people are more likely to do it when provided appropriate attributional cues about their partner, (b) during other-centric responding, adopting an imagined perspective at a large oblique offset is more cognitively taxing than at an orthogonal offset, and (c) surprisingly, even when responding egocentrically, the partner's perspective may still be processed, as reflected by increased response times when the partner was at oblique offsets. However, the statements in (b) and (c) are only tentative at this juncture, since in Experiment 1 and 2 the increased response times when the speaker was at oblique offsets could be due to the number of objects in those trials.

With respect to the first point, our findings demonstrate that social attributions about the conversational partner influence perspective-taking, in line with previous work of Duran et al. (2011). In that study, when listeners thought that their simulated conversational partner was limited (e.g. did not know their viewpoint)

other-centric responding increased, whereas when they thought that they were interacting with a real partner who knew their viewpoint, egocentric responding increased. Similarly, here, when listeners believed that their conversational partner was the experimenter, they were more likely to place the communicative burden on her, holding her responsible for ensuring mutual understanding, especially as she was the one making the requests. In this scenario, preference for the egocentric perspective was extremely strong, with all but one participant being classified as egocentric responders. On the other hand, when the same person's ability to contribute to the task was presented as being more limited, by being introduced as another participant,<sup>5</sup> other-centric responding increased.

These findings are compatible with the *principle of least collaborative effort* (Clark, 1996; Clark & Wilkes-Gibbs, 1986), which posits that language users select the perspective strategies that maximise their efficiency of communication. By assuming shared responsibility for mutual understanding, language users don't merely select the least cognitively demanding perspective, but rather are willing to invest the cognitive effort to adopt the other's perspective, if that is thought to maximise their efficiency of coordination (see also Duran et al., 2011; Schober, 1995, 2009). In this view, adopting the partner's perspective in Experiment 2, as several participants had done, made sense despite its associated cognitive cost. This cognitive cost was evident in the longer response times of other-centric and mixed responders relative to egocentric responders, and in the overall difference in response latencies of participants on ambiguous trials across Experiments 1 and 2.

The gaze fixation results can also be viewed as compatible with the idea that responding from another's perspective is costly, since the egocentric choice was often fixated before making an other-centric response (nearly half of the time) in both experiments. In contrast, the other-centric choice was rarely fixated before making an egocentric response. This asymmetry in the distribution of fixations suggests that the egocentric perspective introduces competition when responding other-centrally, whereas the reverse does not seem to be the case. The competition introduced by the egocentric perspective during other-centric responding is compatible with proposals that in many contexts perspective-taking involves an egocentric default in early processing (e.g. Duran et al., 2011; Keysar, Barr, Balin, et al., 1998; Keysar, Barr, & Horton, 1998).

As we noted in points (b) and (c) above, the cognitive cost of responding wasn't constant across speaker's positions. In both experiments, the listeners' response latencies exhibited a sawtooth pattern, with slower

performance at the large oblique offsets (135° and 225°) relative to the orthogonal offsets; the type of offset was a significant predictor of response times. This finding is difficult to interpret at this point due to a flaw in our design: the speaker's position and the complexity of the configurations were confounded, with 3-object configurations always depicting the speaker at an oblique offset (135° and 225°) and 2-object configurations always depicting the speaker at an orthogonal offset (90°, 180°, 270°). It is therefore unclear whether the difference in performance between the two types of offsets was due to always processing the partner's viewpoint (even in Experiment 1), or due to the complexity of the configuration.

If indeed the difference in response times between orthogonal and oblique offsets was exclusively due to the speaker's position, this would mean that adopting an imagined perspective that is at an oblique offset is more difficult than adopting one at an orthogonal offset, in line with previous proposals (Galati et al., 2013; McNamara, 2003). Since response times, here, were not the longest at the largest offset (180°), this may suggest that oblique offsets are harder to maintain or compute responses from than orthogonal offsets, even if all offsets were initially adopted through mental rotation. The increased processing cost at large oblique offsets is also consistent with findings from an interpretation task in which switching perspectives at large offsets was shown to be more cognitively costly than switching perspectives at small oblique offsets (Ryskin et al., 2016).

At this stage, the interpretation of the patterns in the response times is indeterminate. To clarify the source of the processing cost of those trials with the speaker at oblique perspectives, we tease apart these confounded factors (perspective and number of objects) in Experiment 2B.

## Experiment 2B

In Experiment 2B we sought to clarify whether, in Experiment 1 and 2, the longer response times obtained when the speaker was depicted at oblique offsets (135° or 225°) relative to orthogonal offsets, was due to the difficulty of adopting the speaker's oblique perspective or due to the difficulty of processing 3-object configurations. We therefore included trials with 3-object configurations when the speaker was depicted at orthogonal offsets (0°, 90°, 180°, 270°) and 2-object configurations when the speaker was depicted at the oblique offsets (135°, 225°).

The same procedure as Experiment 2 was followed. We chose the social attribution condition of Experiment 2 since it had resulted in a distribution of perspective preference that included both egocentric and other-centric

responders, and could therefore enable us to examine the behaviour of both types of responders in Experiment 2B. The listeners' response times on 2 and 3-object configurations at orthogonal and oblique offsets here should clarify the cause of the increased response times at the oblique offsets in the earlier experiments.

## Method

### Participants

Twenty-five undergraduate students from the University of Cyprus served as listeners (23 female). All participants received course research credit for their participation; none had participated in Experiments 1 and 2.

### Stimuli

The 56 trials of Experiment 1 and 2 were included in Experiment 2B. In addition to these, another 24 3-object trials with the speaker at orthogonal offsets and another 16 2-object trials with the speaker at oblique offsets were constructed, resulting in a total of 96 trials.

Of the new 3-object trials, 12 were 3-object ambiguous trials with the speaker depicted at the orthogonal offsets of 90°, 180°, and 270°. There were 4 such trials from each of these orthogonal offsets, one for each type of instruction (*left, right, top, bottom*). Figure 1(c) illustrates one such ambiguous 3-object trial, with the speaker at an orthogonal offset.

The remaining 12 3-object configurations were control trials with the speaker depicted at 0° (3 for each of the four different types of instructions: *left, right, top, bottom*). The orthogonal offsets of 90° and 270° did not lend themselves to creating 3-object control trials, as the addition of a third object to any of the eight 2-object control trials from these offsets from Experiment 1 and 2, would permit a felicitous interpretation only from the listener's or from the speaker's perspective. The 24 new 3-object trials from orthogonal offsets supplemented the 16 3-object ambiguous trials from Experiments 1 and 2, with the speaker depicted at oblique offsets.

All of the 16 new 2-object trials were ambiguous; half with the speaker depicted at 135° and half at 225°, two for each type of instruction (*left, right, top, bottom*) from each offset.

Thus, of the 96 trials of Experiment 2B, 56 involved 2-object configurations and 40 involved 3-object configurations. Of the 56 2-object configurations, 36 were ambiguous and 20 were control trials; of the 40 3-object configurations, 28 were ambiguous trials and 12 were control, resulting in a total of 64 ambiguous and 32

control trials. There were 24 trials of each instruction type (*left, right, top, bottom*), 16 ambiguous and 8 control trials.

The order in which the trials were presented was randomised for each participant.

### Procedure

The procedure was identical<sup>6</sup> to Experiment 2.

### Design and analyses

The experimental design of Experiment 2B contained the same factors as Experiment 2, with the addition of the type of configuration (2- vs. 3-object). We built the mixed effect models in the same fashion as in Experiments 1 and 2, with the addition of type of configuration as a fixed factor. For a more parsimonious presentation of the models (summarised in Table 2 in Appendix A), based on model comparisons, we did not include as predictors the interactions between object number with offset type and instruction type, as these interactions were not theoretically motivated and did not significantly improve the fit of the models. In the random effect structure, we included a random slope for number of objects in the configuration; we simplified the models as needed, using the same procedure as before until the models converged.

In terms of our expository presentation in the Results section below, we first consider listeners' perspective choices on ambiguous trials and compare the distribution of responders (egocentric, other-centric, and mixed) across Experiments 2 and 2B. In doing so, we establish whether the cover story used about the partner (i.e. about the partner being another participant) had a comparable impact on the two experiments, yielding similar distributions of responses. Then, for the measures indicating the listeners' difficulty in processing (i.e. gaze fixations and response times), we focus on performance in Experiment 2B. For these measures, we don't make direct comparisons of the two experiments (especially, given our interest in the influence of the speaker's position on response times), since in Experiment 2 speaker position and the number of objects in the configuration were confounded.

## Results

### Perspective choices

First, we wanted to establish that the distribution of egocentric, other-centric, and mixed responders in Experiment 2B was similar to that of Experiment 2, despite the addition of new trials. As we reported earlier, in

Experiment 2, 8 participants were classified as other-centric responders, 4 as mixed responders, and 14 as egocentric responders. In Experiment 2B, there were 5 other-centric responders, 2 mixed responders, and 18 egocentric responders. The distribution of the three types of responders across the two experiments did not differ significantly,  $\chi^2(2) = 1.84, p = .40$ . This suggests that the manipulation of the status of the speaker (as a naïve participant) had a similar impact on listeners' responses in both experiments.

In Experiment 2B, as in Experiment 2, the position of the speaker's offset (orthogonal vs. oblique) did not influence significantly the proportion of egocentric responses ( $\beta = -.17, SE = .23, z = -.73, p = .47$ ; see Table 2 in Appendix B). As illustrated in Figure 2, listeners in Experiment 2B adopted the speakers' viewpoint in similar proportions across the speaker's positions. Importantly, the number of objects did not influence the listeners' perspective choice ( $\beta = -.07, SE = .23, z = -.30, p = .77$ ). Consistently with the previous experiments, the type of instruction (lateral vs. sagittal) did not influence perspective choice either ( $\beta = -.05, SE = .23, z = .23, p = .82$ ).

### Gaze fixations

As with the previous experiments, the eye-tracking analyses of the ambiguous trials of Experiment 2B showed that when listeners responded egocentrically they were more likely to fixate first on the object compatible with the egocentric (94%) than the other-centric choice (6%),  $t(24) = 9.55, p < .001$ . In contrast, when selecting the other-centric response, participants first fixated

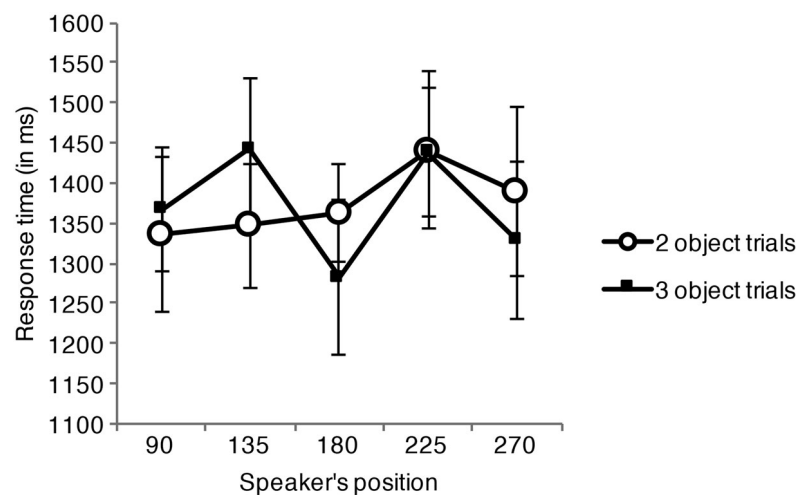
with comparable frequency the other-centric (57%) and the egocentric options (43%),  $t(24) = 1.66, p = .11$ .

### Response times

Our main undertaking in Experiment 2B was to establish whether the pattern of increased response times at oblique offsets, observed in Experiments 1 and 2, persisted after controlling for the number of objects in the configurations.

The type of offset (oblique vs. orthogonal) did indeed predict reaction times significantly ( $\beta = .05, SE = .02, t = 2.16, p < .05$ ; see Table 2 in Appendix B): on average listeners took 1417 ms (SD = 1219 ms) to respond when the speaker was depicted at oblique offsets and 1348 ms (SD = 984) on orthogonal offsets. Consistent with Experiments 1 and 2, offset type influenced the response times of other-centric responders (relative to egocentric responders) as indicated by a significant interaction of those terms ( $\beta = -.14, SE = .05, t = -2.65, p < .001$ ). Here, offset type influenced the response times of mixed responders (relative to egocentric ones) as well ( $\beta = -.32, SE = .08, t = -4.23, p < .001$ ).

The number of objects had an independent effect on response times ( $\beta = .06, SE = .02, t = 2.44, p < .05$ ). Importantly, the effect of the number of objects in the configuration did not interact with that of the type of offset ( $\beta = -.05, SE = .06, t = -.54, p = .25$ ). In Figure 4, the sawtooth pattern of performance is detected in both types of configurations. Even though this pattern is more visually apparent for 3-object than 2-object configurations, performance on the two types of configurations was closely overlapping, suggesting that the



**Figure 4.** Response times (in ms) on ambiguous trials from Experiment 2B, across the possible speaker's positions for 2-object and 3-object configurations.



increased processing cost of perspective-taking across the different headings did not depend on the number of objects in the configuration.

Consistent with our previous findings, listeners were faster on sagittal (*top-bottom*) verbal instructions compared to the lateral (*left-right*) ones, although the contribution of instruction type as a predictor did not reach significance ( $\beta = -.06$ ,  $SE = .03$ ,  $t = -1.89$ ,  $p = .06$ ).

To summarise: (a) in Experiment 2B, response times on trials with the speaker at oblique headings were slower than those with the speaker at canonical headings, and (b) the influence of the type of heading did not depend on the number of objects in the configuration. These findings collectively suggest that the sawtooth patterns observed in Experiments 1 and 2 were due to the difficulty of considering the oblique offsets rather than the number of objects in the displays.

## Discussion

The findings of Experiment 2B suggest that the sawtooth pattern across orthogonal and oblique offsets that we observed in listeners' response latencies in Experiments 1 and 2 was not merely due to the number of objects in the configurations. In Experiment 2B, the speaker's position – i.e. whether the speaker was located at oblique or orthogonal offset – had a reliable effect on the listeners' response times, independently of the number of objects in the configurations.

The difference in response latencies between oblique and orthogonal headings was driven by non-egocentric responders, including both other-centric responders, who by definition took the speaker's position into account when selecting an object, and mixed-responders, who adopted an other-centric orientation at least part of the time.

The fact that Experiments 2 and 2B yielded similar distributions of egocentric, other-centric, and mixed responders is reassuring, given that listeners were provided the same attributional cues about the speaker (i.e. that she was another participant). The persistence of the sawtooth pattern in the response latencies of Experiment 2B, specifically for non-egocentric responders, corroborates further the proposal that maintaining or computing responses from oblique perspectives is more difficult than from orthogonal offsets.

## Experiment 3

So far, we have found evidence consistent with the idea that adopting a non-egocentric strategy incurs a greater cognitive cost in a social spatial perspective-taking task. Making an egocentric response was faster than making

an other-centric response. Moreover, making an egocentric response rarely involved a first fixation to the other-centric object prior to object selection.

However, in these experiments, what has been referred to as the "egocentric perspective" in the task – i.e. the perspective representing the self / the listener – was an imagined perspective that was aligned with the listeners' sensorimotor perspective as they were facing the computer display. The participants' body orientation and facing direction were always aligned with and could be easily mapped onto that imagined egocentric perspective represented by the arrow (at 0°) in the top-down view of the table-top on their screen. Although some perspective transformation may have still taken place in these experiments in order to map the sensorimotor perspective onto the egocentric perspective at 0° on the tabletop scene, that transformation appears to have been fairly fast and automatic, as indicated by the listeners' response times and eye fixations when making an egocentric selection.

In Experiment 3, we wanted to examine whether the transformation of the sensorimotor perspective to non-zero "egocentric" perspectives would be more computationally demanding. We wanted to investigate whether, relative to the previous experiments, the processing advantage of the imagined egocentric perspective would persist. If not, this would suggest that the previously documented advantage of the egocentric perspective was due to its coincidence with the sensorimotor perspective. As we have noted in the Introduction, the sensorimotor perspective and the "egocentric" perspective representing the self may at times be dissociated, as when moving an avatar representing the self in virtual environments or when reasoning about our past experiences in remote environments.

In Experiment 3, to dissociate the sensorimotor from both the egocentric and other-centric perspectives, the participant's egocentric viewpoint varied from trial to trial to different headings (0°, 90°, 180°, 225°, 270°, 315°), while the speaker's other-centric viewpoint remained fixed at 90°. We reasoned that, in these circumstances, adopting the imagined egocentric perspective would be more difficult than in the previous experiments, as it would require mapping the interpretation of the relative spatial terms (*top*, *bottom*, *left*, *right*) on a trial-by-trial basis. Adopting the other-centric perspective – although also dissociated from the sensorimotor perspective – would involve a stable mapping of the spatial terms across trials on relevant sections of space. This prediction is also motivated by evidence that there is a cost associated with switching spatial perspectives across trials

when interpreting utterances from a social partner (Ryskin et al., 2014, 2016).

Under the same social attributional cue about the partner as Experiments 2 and 2B, in which the speaker was introduced as another participant, we wished to establish whether adopting the egocentric perspective here would indeed be more difficult compared to those experiments and to further examine whether this increased difficulty would be associated with a different distribution of perspective choices.

## Method

### Participants

Twenty-four undergraduate students from the University of Cyprus were the listeners (22 female). They all participated in exchange for course credit.

### Stimuli

As with the previous experiments, trials involved top-down views of a circular table with a configuration of two or three objects (CDs). In Experiment 3, the speaker's perspective, indicated by a human figure, remained stable across trials (at 90°), as shown in Figure 5. The participant's (the listener's) perspective changed across trials, with the position of the red arrow depicted at different headings across trials (90°, 180°, 225°, 270°, 315°, 0°). Notably, these headings correspond to those used in Experiments 1, 2, and 2B, but rotated by 90°.

Of the 80 trials, 16 trials were control trials, resulting in the same object choice from both the listener's and the

speaker's perspective, and 64 trials were ambiguous trials, resulting in different object choices. Half of the control trials were 2-object configurations (with the listener at 0° or 180°) and half were 3-object configurations (with the listener at 90°). Similarly, half of the ambiguous trials were 2-object configurations and half were 3-object configurations. Among ambiguous trials, there were 8 with the listener at 0°, 8 with the listener at 180°, 16 with the listener at 225°, 16 at 270°, and another 16 at 315°; these were evenly divided among 2- and 3-object configurations. The four types of instructions (*top*, *bottom*, *left*, *right*) were also equally represented in the stimulus set.

The practice trials ( $N=8$ ) were all control trials. The same prerecorded instructions as the previous experiments were used. The order in which the trials were presented was randomised for each participant.

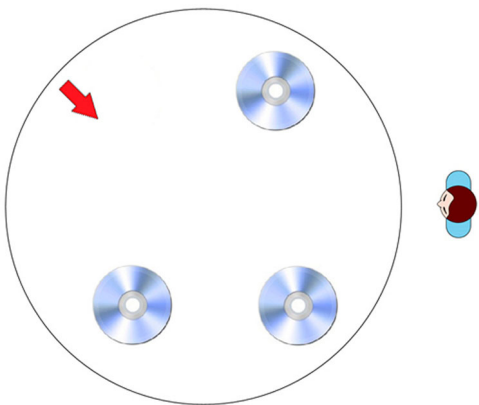
### Procedure

The procedure was identical to Experiment 2B. The only difference was that E2 explained to participants that the speaker's (E1's) viewpoint, represented by the human figure, would be stationary across trials at 90°, whereas their own viewpoint, represented by the arrow, would vary across trials.

### Design and analysis

To recapitulate the experimental design, participants were classified as a certain kind of responder (egocentric, other-centric, or mixed); they all responded to instructions that varied in terms of trial type (ambiguous vs. control), listener position (for ambiguous trials: 0°, 315°, 270°, 225°, 180°; for control trials: 0°, 90°, and 180°), and spatial instruction (*left*, *right*, *top*, *bottom*). As with the previous experiments, for each trial, we obtained data regarding whether the listener made an egocentric or other-centric object selection, whether the listener's first gaze fixation was on the egocentric or other-centric object, and the listener's response time for that trial.

As with the previous experiments, we focus on performance on ambiguous trials, and present the results of control trials in Appendix B. For models of perspective choice on ambiguous trials, we included the listener's position (0°, 315°, 270°, 225°, 180°) instruction type (sagittal: *top-bottom* vs. lateral: *left-right*), and number of objects (2 vs. 3) as fixed effects. For models of response time, we also included perspective preference (egocentric, other-centric, mixed) as a fixed effect. For a more parsimonious presentation of the models (summarised in Table 3, Appendix A), based on model



**Figure 5.** Example of a trial from Experiment 3. The configuration illustrates an ambiguous 3-object trial when accompanied by the instruction "Select the CD on the right", with the speaker depicted by a human figure at 90° (which remained stationary across trials) and the listener's position indicated by the arrow (which was variable across trials).

comparisons, we included as a predictor the interaction between perspective preference and listener position, but not the remaining interactions, as those were less theoretically motivated and did not significantly improve the fit of the models. In terms of the random effects, we started with the maximal random effect structure, and simplified using the same criteria as before until the models converged.

Note that in Experiment 3, we chose as a predictor the listener's position rather than the type of offset at which the listener was depicted (oblique vs. orthogonal), since we were primarily interested in the influence of the increasing misalignment (in 45° increments) between the egocentric and sensorimotor perspectives. Moreover, the cognitive cost of reasoning from oblique offsets is often documented to hold for large but not for small offsets (e.g. for 225° but not for 315° here; see for example Ryskin et al., 2016). Such findings suggest that grouping oblique offsets in this experiment may obfuscate the cost of perspective-taking under increasing sensorimotor dissociation.

We focus on the results of Experiment 3, but also make a comparison of the distribution of responders here with that of Experiment 2B to establish whether these distributions differed (due to the misalignment of the egocentric and sensorimotor perspectives here) or not (due to the same attribution about the speaker).

## Results

### Perspective choices

We had anticipated that adopting the egocentric perspective would be more difficult here, since the

egocentric perspective was variable across trials whereas the other-centric perspective was fixed. However, listeners still responded egocentrically to a large extent: they chose the egocentric object on .76 of ambiguous trials ( $SD = .43$ ); the same proportion of egocentric responses as in Experiment 2B.

Despite this preference, there was evidence that the difficulty of adopting the egocentric perspective increased as the misalignment between the egocentric and sensorimotor perspectives increased. With increasing misalignment, there were fewer egocentric responses and more other-centric responses. As shown in Table 2, egocentric responding decreased as the arrow representing the participant moved from 0° to 315° to the larger offsets (270° and 225°, and further so at 180°). This change in the proportions of responses across headings contrasts with what we had observed in Experiments 1 through 2B. In those experiments, the proportions of egocentric responses remained stable across speaker positions; as we have reported, the type of offset at which the speaker was depicted did not influence perspective choice. Here, as adopting the egocentric viewpoint became more difficult (corroborated by increasing reaction times, seen in Table 2 and discussed below), listeners were less likely to interpret spatial instructions from their own viewpoint. This is captured by the finding that the listener's position at the larger offsets (specifically at 270°, 225°, 180°) were all significant predictors in the model for egocentric perspective choice, as shown in Table 3 of Appendix A.

The number of objects in the configuration did not influence the selection of an egocentric choice ( $\beta = -1.63E-06$ ,  $SE = 1.85E-01$ ,  $z = .00$ ,  $p = .99$ ), and

**Table 2.** Means (and standard deviations) of the proportions of egocentric (Ego choice) and other-centric (Other choice) perspective choices, and response times (in ms) on ambiguous trials, across the types of response preference (egocentric, other-centric, and mixed) and across the different positions of the listener in Experiment 3. In Experiment 3, the other-centric perspective was stationary at 90°, whereas the egocentric perspective changed across ambiguous trials (0°, 315°, 270°, 225°, 180°) in 45° increments.

	Egocentric responders			Other-centric responders		
	Ego choice	Other choice	RT	Ego choice	Other choice	RT
0°	.97 (.18)	.03 (.16)	1738 (1277)	.00 (.00)	1.00 (.00)	1587 (732)
315°	.95 (.21)	.02 (.15)	1831 (1150)	.00 (.00)	1.00 (.00)	1631 (977)
270°	.94 (.23)	.03 (.18)	1700 (1052)	.02 (0.14)	.98 (.14)	1511 (744)
225°	.92 (.27)	.05 (.22)	2062 (1340)	.02 (0.14)	.98 (.14)	1560 (1410)
180°	.88 (.33)	.08 (.28)	2283 (1386)	.04 (0.20)	.75 (.44)	1516 (1252)
Total	.94 (.25)	.04 (.20)	1901 (1238)	.02 (0.20)	.96 (.20)	1564 (1056)
	Mixed responders			All responders		
	Ego choice	Other choice	RT	Ego choice	Other choice	RT
0°	.71 (.46)	.21 (.41)	3626 (3968)	.81 (.39)	.17 (.38)	1955 (1893)
315°	.69 (.47)	.27 (.45)	3083 (2315)	.80 (.40)	.18 (.38)	1962 (1396)
270°	.42 (.50)	.44 (.50)	3714 (4031)	.76 (.43)	.20 (.40)	1929 (1831)
225°	.38 (.49)	.38 (.49)	3323 (3305)	.76 (.44)	.21 (.41)	2157 (1776)
180°	.21 (.41)	.58 (.50)	3825 (3447)	.69 (.47)	.23 (.42)	2480 (1849)
Total	.48 (.50)	.37 (.48)	3462 (3378)	.76 (.42)	.20 (.40)	2053 (1733)

neither did the type of instruction (lateral vs. sagittal;  $\beta = -2.54E-01$ ,  $SE = 1.86E-01$ ,  $z = -1.37$ ,  $p = .17$ ).

We investigated further the relationship between perspective choice and listener position, by considering the three types of responders (egocentric, other-centric or mixed responders), which were classified according to the same criteria as before. In Experiment 3, 18 participants (75%) were classified as egocentric, 3 (12.5%) as mixed responders, and 3 (12.5%) as other-centric responders. The distribution of the three types of responders across Experiment 3 and Experiment 2B, both of which used the same attributional cue about the speaker, did not differ significantly,  $\chi^2(2) = .68$ ,  $p = .71$ .

As observed in Table 2, for egocentric and mixed responders, as the angular disparity between the egocentric and sensorimotor perspectives increased, the proportion of egocentric choices decreased. In addition, other-centric responders selected the other-centric choice fairly consistently across headings, except for 180°, where other-centric responses dropped (due to selecting the 3rd CD of the configuration, and not due to selecting the egocentric CD). This may reflect the confusability of conflicting perspectives in this scenario (with the sensorimotor at 0°, the imagined egocentric at the counter-aligned 180°, and the preferred other-centric at 90°). These patterns could not be evaluated statistically, as models of perspective choice that included as predictors perspective preference and its interaction with the listener's position did not converge, given the small number of other-centric and mixed responders.

### Gaze fixations

Analysis of the gaze fixations for the ambiguous trials revealed that when listeners responded egocentrically, they first fixated more frequently the egocentric (73%) than the other-centric choice (27%),  $t(23) = 7.70$ ,  $p < .001$ . Relative to the previous experiments, where listeners fixated on the egocentric choice more than 90% of the time, here, they experienced increased interference from the other-centric option during egocentric responding.

When participants responded other-centrally, they first fixated more frequently the other-centric choice (72%) than the egocentric choice (28%),  $t(23) = 2.21$ ,  $p < .05$ . Relative to the previous experiments, where the difference between proportions of fixations at the two objects was smaller, the proportion of other-centric first fixations increased. This suggests that listeners responding other-centrally here experienced less competition from the egocentric perspective.

### Response times

As we anticipated, the misalignment between the sensorimotor and imagined egocentric perspectives influenced response times. Table 2 shows that listeners were slower to respond as their depicted position deviated more from their sensorimotor perspective (0°). The listener's depicted position did indeed have a significant impact on response times, as evidenced by a significant improvement in the fit of the model with its inclusion,  $\chi^2(12) = 69.47$ ,  $p < .001$ . As shown in Table 3 of Appendix A, the difference between responding from 0° and the maximally misaligned heading (180°) was significant as a predictor of response time ( $\beta = .27$ ,  $SE = .05$ ,  $t = 5.54$ ,  $p < .001$ ), as was the difference between 0° and the next farthest offset (225°:  $\beta = .16$ ,  $SE = .04$ ,  $t = 3.74$ ,  $p < .001$ ). The differences between 0° and the closest headings of 270° and 315° were not significant predictors of response time ( $p = .78$  and  $p = .22$ , respectively).

Overall, other-centric responders were the fastest of the three groups numerically ( $M = 1564$ ,  $SD = 1056$ ), followed by egocentric responders ( $M = 1901$ ,  $SD = 1238$ ), and with mixed responders being the slowest ( $M = 3462$ ,  $SD = 3378$ ). Although the response times of these responders (other-centric and mixed) did not differ significantly from egocentric responders (see Table 3 of Appendix A), the inclusion of perspective preference in the model significantly improved its fit,  $\chi^2(10) = 26.84$ ,  $p < .001$ .

The effect of the listener's position on response times was driven by egocentric responders, as its contribution remained significant in a model for which only egocentric responders were considered,  $\chi^2(4) = 67.43$ ,  $p < .001$ . Egocentric responders took 1738 ms ( $SD = 1277$  ms) to respond when they were depicted at 0°, 1830 ms ( $SD = 1150$  ms) at 315°, at 1700 ms ( $SD = 1052$  ms) at 270°, 2062 ms at 225° ( $SD = 1340$  ms), and 2283 ms ( $SD = 1386$  ms) at 180°. They were significantly faster at 0° than 180° ( $\beta = .27$ ,  $SE = .05$ ,  $t = 5.88$ ,  $p < .001$ ) and 225° ( $\beta = .16$ ,  $SE = .03$ ,  $t = 3.97$ ,  $p < .001$ ). This is also captured in Table 3 in Appendix A, by the significant interaction terms between other-centric response preference (vs. egocentric preference) and each of these two listener positions: in contrast to egocentric responders, for other-centric responders the difference between 0° and these two headings was very small (0°:  $M = 1587$ ,  $SD = 732$  ms, 180°:  $M = 1516$ ,  $SD = 1252$  ms; 225°:  $M = 1560$ ,  $SD = 1410$  ms).

Consistent with the findings on perspective choice and with the results of Experiment 2B, the number of objects influenced response times here as well ( $\beta = .07$ ,  $SE = .02$ ,  $t = 3.37$ ,  $p < .001$ ), with listeners being

faster on 2-object configurations ( $M = 1930$ ,  $SD = 1451$ ) than 3-object ones ( $M = 2186$ ,  $SD = 1855$ ). The type of instruction also predicted response times significantly ( $\beta = -.16$ ,  $SE = .03$ ,  $t = -5.58$ ,  $p < .001$ ), with listeners being slower on *left-right* instructions ( $M = 2186$ ,  $SD = 1671$ ) than *top-bottom* instructions ( $M = 1922$ ,  $SD = 1785$ ).

## Discussion

The findings of Experiment 3 suggest that the alignment of the sensorimotor and imagined egocentric perspectives influences the ease of adopting that egocentric perspective. As the misalignment between the sensorimotor and the depicted egocentric perspective increased, listeners were slower to respond egocentrically.

In contrast to the previous experiments, egocentric responders here were not faster than other-centric participants; in fact, they were numerically slower, although the difference between the two groups was not reliable. Additionally, as reported in the control trials' results in [Appendix B](#), egocentric responders made numerically more errors on control trials than other-centric responders. In the previous experiments, when the sensorimotor and egocentric perspectives were aligned, performance on control trials was near perfect across all types of responders.

What's more, the listeners' first gaze fixation during egocentric responding exhibited more interference from the other-centric perspective than in the previous experiments, in which the other-centric perspective was rarely fixated first. And conversely, during other-centric responding here, the egocentric option exhibited less competition, as it was less frequently fixated first compared to the previous experiments.

Collectively, these findings suggest that, in the previous experiments, the alignment of the sensorimotor and egocentric perspectives was largely responsible for the ease of adopting the egocentric perspective relative to the other-centric one. In this scenario, where the sensorimotor perspective was dissociated from both the egocentric and other-centric perspectives, adopting the fixed perspective (the other-centric one) was relatively less cognitively taxing than the variable perspective (the egocentric one). This makes sense, insofar as interpreting spatial terms from a new heading on each trial could place demands on cognitive resources (Avraamides & Carlson, 2003; Ryskin et al., 2014, 2016).

What is remarkable is that, despite the relative demands of adopting the egocentric perspective in this experiment, egocentric responding was predominant. Listeners heard instructions from a speaker who was presumed to be a real, naïve participant, and who had

greater responsibility for mutual understanding in this informationally asymmetrical task, insofar as the speaker requested a targeted action from the listeners and could not provide feedback. The relative predominance of egocentrism over other-centricism, in this social setting, is pragmatically motivated.

In conjunction with this possibility, another possibility at work may be that listeners may have reasoned that because the egocentric perspective was explicitly pointed out to them on each trial (with the moving arrow), they were expected (by the experimenter and/or the speaker) to adopt this perspective. We should note that in Experiments 1 through 2B, where the speaker's perspective was pointed out on each trial through the moving figure, we did not observe a uniform shift to other-centric responding (instead responding was modulated by social attributions, with overwhelmingly egocentric responding in Experiment 1). Still, it is possible that continually highlighting the participants' egocentric perspective (vs. the partner's perspective) resulted in different attributions about what strategy they were expected to adopt in the task.

A consistent observation across all experiments, is that responses to instructions from the sagittal axis (*top-bottom*) were generally faster than those from the lateral axis (*left-right*). This is in line with previous work showing that interpreting directions to move to the left and right of an imagined facing direction is generally slower than moving forward/up and backward/down (e.g. Avraamides & Sofroniou, 2006). This is most likely because (1) mapping left and right to appropriate regions of space requires defining first the *front-back* (or *top-bottom*) dimension, and that (2) the *left-right* axis is highly symmetric and contains no salient cues to differentiate left from right.

In sum, the findings of Experiment 3 underscore that in perspective-taking contexts that involve a dissociation of the sensorimotor and imagined perspectives, adopting an imagined perspective incurs a greater processing cost as it becomes increasingly misaligned from the sensorimotor perspective. Importantly, in perspective-taking contexts of a social nature, attributions about the social partner are predictive of the perspective strategy that people adopt (i.e. whether they choose to adopt the imagined perspective representing themselves vs. their partner), regardless of the cognitive demands of that strategy.

## General discussion

We set out to investigate some of the factors that influence the difficulty of perspective-taking in a social context where listeners interpreted spatial instructions.

We asked whether responding other-centrally incurs a greater processing cost than responding egocentrically, and whether increased misalignment between the egocentric and other perspectives (the other-centric or the sensorimotor perspective) increases processing cost. We also asked how the difficulty of perspective-taking, as reflected by response times and by the competition of perspectives during early eye-fixations, influences the listeners' perspective strategies.

Our findings underscore that social attributions about the partner have a strong influence on the perspective strategy that language users adopt. As we observed, in Experiment 1, listeners were almost exclusively egocentric under the belief they were interacting with the experimenter, whereas in Experiment 2, which used identical materials, their other-centric preference increased when the same speaker was introduced as another participant. Experiments 2B and 3, which used the same cover story as Experiment 2 resulted in comparable distributions of perspective preference. The preponderance of egocentrism we observed in Experiment 1 is compatible with other work demonstrating that speakers are less likely to adapt their behaviour (e.g. they are less likely to disambiguate their spatial descriptions), when they suspect that their conversational partner is a confederate and does not have real informational needs (Roche, Dale, & Kreuz, 2010).

The distributions of perspective strategies across experiments is consistent with the view that when speakers perceive their conversational partner's as being limited in some way (in terms of their knowledge, their informational needs, or their ability to interact with them), they are more likely to invest the effort to adopt the partner's perspective (e.g. Clark & Wilkes-Gibbs, 1986). As we've noted, listeners were relatively more accommodating when they perceived their partner to be a fellow participant vs. the experimenter. Similarly, in other work, language users were more likely to adopt their partner's perspective or make felicitous adjustments when they believed their partner to be imaginary (Schober, 1993), a child (Newman-Norlund et al., 2009), unfamiliar with the environment (Hölscher et al., 2011), or when they believed that their partner did not know (Duran et al., 2011) or did not share their viewpoint (Mainwaring et al., 2003; Schober, 1993, 1995).

Importantly, social attributions trumped the difficulty of adopting a particular perspective in predicting the overall perspective strategy used. The social attribution that the partner was another participant gave rise to similar distributions of egocentrism and other-centrism, even when adopting the egocentric perspective became more difficult than adopting the partner's

perspective in Experiment 3. Other-centric responding did not dramatically increase in that scenario, but remained comparable to Experiment 2B, which involved the same social attribution about the partner.

This view of social attributions – as factors that bias language users toward a particular perspective preference – is compatible with models that construe social attributions as “control parameters”, that is, as system variables that shape how language users settle on a particular perspective strategy. Such a dynamical model of perspective-taking has been proposed by Duran and Dale (2014), who described a bistable attractor model in which attributions about the partner were represented as a control parameter. Complex dynamical systems – including biological, cognitive, and social ones – are systems whose behaviour evolves over time and is non-linear, arising from a large number of interacting elements or components. In dynamical systems, control parameters are variables that create instabilities in the system's behaviour. These parameters, when changed, can significantly influence how the system evolves over time, and can represent some important external condition or factor that constrains the system's behaviour. By using a control parameter to represent social attributions about the partner in a dynamical model, Duran and Dale (2014) captured the fact that language users stabilise either on egocentric or other-centric orientation under different task constraints. Their model – with different values for the control parameter for the different attributions listeners had about the speaker – accounted well for the behavioural data of Duran et al. (2011), which used a similar paradigm as the one we have used here.

The view that social attributions about the partner can be represented as simple variables (as control parameters or as constraints) is also expressed in proposals that speakers track partner-specific information as simple (often binary) cues or distinctions. (Brennan, Galati, & Kuhlen, 2010; Brennan & Hanna, 2009; Galati & Brennan, 2010). Such “one-bit models” about the partner (e.g. *has my partner heard this before or not*), permit speakers to easily track or cue these relevant distinctions in a timely fashion that can have a drastic impact how they plan and process language (including their spatial perspective strategy). For example, a speaker's voice can constrain the interpretation of a temporarily ambiguous sentence when it is consistently paired with a visuospatial perspective (Ryskin et al., 2016); this association – between the speaker's identity and their position in space – is successfully deployed to the extent that it can be maintained in memory due to being reliable, relevant, and simple. Another related account is Butterfill and Apperly's (2013) “minimal

theory of mind”, which proposes that people track others’ propositional attitudes, including beliefs, by representing them as simpler relational mental states, such as action-directed goals. Together, these accounts of perspective-taking underscore that simple representations of relevant information about the partner (e.g. concerning shared “common ground” information, or the partner’s perceptions or beliefs) may be necessary in order to characterise language users’ nimble behaviour in dialogue. Simple models eliminate the cognitive demands associated with representing others’ perceptions or beliefs fully, and enable perspective-taking even for those with limited cognitive resources or sophistication (e.g. adults under cognitive load, infants, chimpanzees).

We should acknowledge that, regardless of the social attribution we elicited, egocentric responding was relatively high across experiments. This may be because, in all experiments, listeners ascribed greater responsibility for ensuring mutual understanding to the speaker, who was the requester an action and could not provide feedback. In light of these task features and constraints, the relatively high degree of egocentrism is not so surprising. Under other attributional scenarios – for example if listeners believe that the speaker does not know where they are located, as in one of the experiments in Duran et al. (2011) – the distribution of perspective preference may shift more dramatically toward other-centric responding.

In terms of the features of our experimental task, we should note that although – for the sake of experimental control – our task departed from several real-life dialogue situations where interlocutors interact freely and can explicitly negotiate the mapping of linguistic terms when faced with ambiguity (e.g. as in Galati & Avraamides, 2015; Schober, 1993, 1995, 2009), our task’s constraints are in fact common in real-life. In many everyday situations interlocutors have to coordinate with one another under conditions that are noisy, involve unidirectional communication channels, temporal lags, and other limitations in shared affordances that prevent contingent feedback. For instance, in a multiplayer video game, a player with a headset may be on a team with players without headsets; in this situation, contingent verbal feedback is not possible, linguistic communication is unidirectional and can potentially introduce ambiguity regarding the player’s intentions or the unratified goal of the collective at a given moment. These task constraints don’t necessarily thwart successful communication, but rather guide the perspective strategies of language users.

Even when communication is limited in some respect (in our study, involving unidirectional communication,

no feedback from the speaker, and an often ambiguous speech signal), language users still stabilise on strategies that are driven by task constraints, including social attributions about who the partner is. Indeed, with the exception of the few mixed responders in our studies, participants were generally consistent in their perspective strategy. In some circumstances, switching perspectives can be effective for coordinating, especially when conversational partners can interact freely (Brennan & Clark, 1996; Tversky, Lee, & Mainwaring, 1999). At the same time, language users generally abide to a consistent spatial perspective upon establishing an explicit or implicit conceptual strategy (Garrod & Anderson, 1987). As we discussed above, perhaps due to the task constraints, which precluded the negotiation of perspective strategies, perspective choice here was largely modulated by the social attributions about the partner.

Beyond investigating perspective strategy, another central aim of our study was to elucidate how the misalignment of perspectives influences the difficulty of perspective-taking, as reflected by response times. Across our studies, three potentially relevant perspectives could be misaligned: the participants’ sensorimotor perspective, their depicted egocentric perspective, and the depicted other-centric perspective.

First, let’s consider participants’ response latencies when the misalignment of the depicted egocentric and other-centric perspectives varied, while the sensorimotor and egocentric perspectives remained aligned (i.e. with listeners depicted at 0°, in Experiments 1, 2, and 2B). In these experiments, we found that when participants responded other-centrally – i.e. adopting a perspective that was misaligned from their egocentric/sensorimotor one – they were slower at the large oblique offsets (of 135° and 225°) than at the maximum offset of 180°. In Experiment 2B, we ruled out the possibility that the increased response times at the oblique offsets were due to the number of objects on those trials, and clarified that the difference between oblique and orthogonal offsets was driven by adopting an other-centric perspective (insofar as other-centric and mixed responders were influenced by offset type significantly more than egocentric responders).

These findings suggest that, although any given perspective may be initially adopted through a process of mental rotation, maintaining that perspective active in working memory or computing responses from that perspective is more difficult from oblique than from orthogonal offsets. This is in line with McNamara’s (2003) view that it is more difficult to maintain or reason from perspectives that are at oblique (vs. orthogonal) headings relative to the axis–or “organizing direction” – being used to encode that information. By including oblique offsets in

our design, our study extends and further qualifies the findings of previous studies that had included only orthogonal offsets (e.g. Duran et al., 2011; Mainwaring et al., 2003) or only oblique offsets (Ryskin et al., 2016), which had suggested that processing cost increases linearly with the misalignment between partners.

Next, let's consider participants' response latencies when the sensorimotor perspective was misaligned from the candidate imagined perspectives (the egocentric and other-centric). In Experiment 3, we dissociated the sensorimotor and depicted egocentric perspectives because we had hypothesised that the ease of processing from the egocentric perspective in Experiments 1, 2, and 2B was, at least in part, due to its alignment with the sensorimotor perspective. In Experiment 3, where most participants (75%) were egocentric responders, we observed that as the misalignment between the sensorimotor and the imagined egocentric perspectives increased, so did response times. Moreover, as that misalignment increased, egocentric responders were less likely to select the egocentric object choice, whether due to temporarily switching strategy or making an error. In contrast to the previous experiments, where perspective choice was unaffected by the speaker's variable position, in Experiment 3 perspective choice was significantly affected by the listener's variable depicted position.

Collectively, these results suggest that the alignment of the sensorimotor and imagined egocentric perspectives had a significant impact on the ease of perspective-taking, with the imagined egocentric perspective incurring a greater cognitive cost than the other-centric perspective, which remained stable throughout the task. Altogether, the findings are compatible with evidence that people use their sensorimotor framework to encode locations (De Vega, 2008; May, 2004). Such encoding is considered particularly useful for the updating of spatial relations during movement (Avraamides & Kelly, 2008). Indeed, many studies have provided evidence that people effortlessly keep track of the changing self-to-object relations in their immediate environment, presumably by relying on proprioceptive information and vestibular signals they receive during movement, even when vision is constrained (e.g. Loomis, Lippa, Klatzky, & Golledge, 2002; Rieser, Guth, & Hill, 1986).

In addition to the patterns in response times, the participants' gaze fixations also corroborated the increased difficulty of egocentric responding in Experiment 3: egocentric responses showed more interference from the other-centric choice (and conversely, other-centric responses showed less interference from the egocentric choice) compared to previous experiments; in those experiments, the other-centric choice was rarely fixated

first on egocentric responses, while the egocentric choice was fixated nearly half of the time on other-centric responses. These changes in the competition of perspectives, as indicated by participants' early eye-fixations across experiments, qualify earlier proposals that there is an egocentric default in early processing (e.g. Duran et al., 2011; Keysar, Barr, & Horton, 1998): they highlight that such a default is tied to the sensorimotor perspective, at least in contexts involving visuo-spatial perspective-taking.

This qualification may seem trivial, seeing that the sensorimotor and imagined egocentric perspectives usually coincide, but these perspectives become more commonly dissociated as we interact in virtual environments and other technologically-mediated settings. For instance, collaborative multiplayer videogames involving 2D projection are analogous to the experimental situation of Experiment 3, insofar as the sensorimotor perspective is dissociated from both the egocentric and other-centric perspectives represented on the screen.

It is important to acknowledge that there is arguably nothing intrinsically "egocentric" about the imagined egocentric perspective in Experiment 3. Whether represented by an arrow or a human figure, both the imagined egocentric and the imagined other-centric perspectives are representations of social perspectives in a top-down depiction of a joint task: both perspectives are dissociated from the sensorimotor perspective. What matters is the participant's interpretation of the arrow and figure, respectively. The questions of "Who am I?" and "Where am I?" are fundamental to human experience (Proulx, Todorov, Aiken, & de Sousa, 2016), and in this setting – in conjunction to the contribution of social attributions about the partner, on which we have focused – it is possible that the representation of the self was prioritised. Literature on self-referential processing suggests that we prioritise perceptual information that relates to the self: for example, people are faster to interpret a degraded stimulus when it has been previously associated with the self (Humphreys & Sui, 2015; Sui, He, & Humphreys, 2012) and to verify the matching of movements with labels, when the label refers to the self than to someone else (e.g. the mother or a stranger, Frings & Wentura, 2014). In Experiment 3, labelling a depicted perspective as "representing the self" promoted egocentric preference, whether due to continually highlighting that changing perspective or due to the presumed increased difficulty of the task for the listener. However, it did not eliminate the cost associated with reasoning from imagined perspectives that are continually changing and are dissociated from the sensorimotor perspective.

To summarise, in social perspective-taking tasks of a spatial nature, the egocentric perspective is easier to



adopt if it is aligned with the sensorimotor perspective. When the imagined egocentric perspective is dissociated from the sensorimotor perspective and it is continually variable, it incurs a processing cost: this is evidenced both by increased latencies and decreased egocentrism with increasing misalignment. With respect to misalignment, when the sensorimotor and egocentric perspectives coincide, adopting a misaligned other-centric perspective is particularly difficult at large oblique offsets compared to the orthogonal offsets. Still, it is the language user's social attribution about their partner – rather than the cost of adopting a particular perspective – that shapes their perspective strategies. Future work can further clarify the interaction of social cues and task specific features (including how the self and the other are represented in the task) on perspective-taking difficulty and perspective preference. Such work can have important implications for understanding and optimising behaviour in situations where people adopt a disembodied perspective, as for example when following a drone's path in a viewfinder or when interacting with others in shared virtual spaces.

## Notes

1. Two trials were removed from the analyses of both Experiments because the participants' responses were coded incorrectly in E-Prime due to experimenter error. Both trials had the figure representing the speaker be at 90° and involved a *back* instruction; one was a control trial and the other was an ambiguous trial.
2. We did not enter the number of objects as a fixed effect in the models, as it was confounded with offset type; including this factor would have introduced collinearity, since trials with the speaker at orthogonal offsets always involved 2-objects and trials with the speaker at oblique offsets always involved 3-objects. We address this issue in Experiment 2B.
3. We also analysed fixation dwell times to egocentric and other-centric response options prior to responding. As the results for all experiments converge with those for the first fixation, we omit those analyses for the sake of brevity.
4. To corroborate this point, in a supplementary analysis, we examined the sawtooth pattern of performance, observed in Figure 3, in an ANOVA framework by fitting the participants' aggregated response times at each of the five headings (90°, 135°, 180°, 225°, 270°) with a planned contrast with the following weights: -0.5, 0.75, -0.5, 0.75, -0.5, with maxima at 135° and 225° (for a related analytical approach, see Galati et al., 2013; Greenauer & Waller, 2008). This sawtooth contrast described adequately the listeners' response times in both experiments. For Experiment 1, the planned contrast was significant,  $F(1, 25) = 8.18, p < .05$ , accounting for 98% of the variance, and leaving a non-significant amount of the variance unaccounted for ( $p = .97$ ). For Experiment 2, this planned contrast was also significant,  $F(1, 25) = 5.48, p < .05$ , accounting for 91% of the variance, and leaving a non-significant amount of the variance unaccounted for ( $p = .89$ ). This sawtooth pattern was more pronounced in Experiment 2 (as seen Figure 3), qualifying the interaction between the speaker's position and the Experiment,  $F(4, 200) = 4.69, p < .01$ .
5. To gauge the degree to which the social attribution promoted by the cover story was successful, before debriefing, we had asked participants a series of questions about their beliefs about their task partner. These questions began with a broad framing ("Did you notice anything strange in the experiment?"), asked about their awareness of their perspective strategy and its consistency, and in Experiments 2, 2b, and 3 – in which we had used the cover story about the confederate being another participant – we asked explicitly about their beliefs about their task partner ("Did you think at any point that the person providing instructions next door may not have been a real participant?"). In these experiments, 44% of participants responded "Yes" to that question – i.e. reporting that they suspected at least at some point that their partner was not a real participant. Although this percentage may seem high, its interpretation is complicated by the issue of demand characteristics. Given the questions posed by the Experimenter leading up to the debriefing, it's unclear whether all participants who responded "Yes" actually suspected the task partner to be a confederate during the experiment, or whether instead they felt compelled to respond so because the possibility just occurred to them and they wanted to save face. Notably, participants rarely reported their suspicions about the use of a confederate in the preceding questions. Moreover, participants' response to this target question was not associated with their perspective-taking strategy; participants were not more likely to respond egocentrically if they had responded "Yes" (i.e. considering the confederate's informational needs as minimal). This could be either due unreliable responses to that question driven by demand characteristics, or due to participants behaving collaboratively – despite any suspicions – according to the information they were provided in the cover story. Even if not all participants bought the cover story we provided, remarkably, any suspicions about the confederate were not enough to trump the effect of the cover story, as evidenced by the increased other-centric orientation observed in Experiment 2 relative to Experiment 1.
6. The only procedural difference relative to Experiment 2 was that, during the experiment, E1 was in a room down the hall (vs. an adjacent room) and E2 remained in the same room as the participant, but in another cubicle and not visible to the participant. These changes in the set up could not be avoided, as the UCY Psychology Department had moved to a new building.

## Acknowledgements

We are grateful to Artemis Stefani for assistance with data collection and data processing, and to Edwin Dalmaijer for

assistance with the eye-tracking analyses. We are also thankful to two anonymous reviewers for their constructive comments.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the H2020 Marie Skłodowska-Curie Actions grant agreement No 705037 to A.G. and by the European Research Council under grant number 206912-OSSMA to M.A.

## ORCID

Alexia Galati  <http://orcid.org/0000-0001-7593-6629>

## References

- Avraamides, M. N., & Carlson, R. A. (2003). Egocentric organization of spatial activities in imagined navigation. *Memory & Cognition*, 31, 252–261. doi:10.3758/BF03194384
- Avraamides, M. N., Hatzipanayioti, A., & Galati, A. (2015). What's so difficult with adopting imagined perspectives? *Cognitive Processing*, 16, (Suppl. 1), 121–124. doi:10.1007/s10339-015-0728-3
- Avraamides, M. N., & Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cognitive Processing*, 9, 93–106. doi:10.1007/s10339-007-0188-5
- Avraamides, M. N., & Sofroniou, S. G. (2006). Spatial frameworks in imagined navigation. *Psychonomic Bulletin & Review*, 13, 510–515. doi:10.3758/BF03193878
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). *lme4: Linear mixed-effects models using Eigen and S4*, 2014. *R package version*, 1–1.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482–1493. doi:10.1037/0278-7393.22.6.1482
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 53, pp. 301–345). Cambridge, MA: Academic Press/Elsevier.
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1, 274–291. doi:10.1111/j.1756-8765.2009.01019.x
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28, 606–637. doi:10.1111/mila.12036
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39. doi:10.1016/0010-0277(86)90010-7
- De Vega, M. (2008). Levels of embodied meaning: From pointing to counterfactuals. In M. De Vega (Ed.), *Symbols and embodiment* (pp. 285–308). Oxford: Oxford University Press.
- Duran, N. D., & Dale, R. (2014). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas in Psychology*, 32, 131–146. doi:10.1016/j.newideapsych.2013.03.004
- Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, 121, 22–40. doi:10.1016/j.cognition.2011.06.009
- Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63–76. doi:10.1037/0096-3445.119.1.63
- Frings, C., & Wentura, D. (2014). Self-prioritization processes in action and perception. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1737–1740. doi:10.1037/a0037376
- Galati, A., & Avraamides, M. N. (2015). Social and representational cues jointly influence spatial perspective-taking. *Cognitive Science*, 39, 739–765. doi:10.1111/cogs.12173
- Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, 62, 35–51. doi:10.1016/j.jml.2009.09.002
- Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*, 68, 140–159. doi:10.1016/j.jml.2012.10.001
- Galati, A., Panagiotou, E., Tenbrink, T., & Avraamides, M. N. (2017). Dynamic strategy selection in collaborative spatial tasks. *Discourse Processes*, 23, 1–23. doi:10.1080/0163853X.2017.1293482
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition*, 27, 181–218. doi:10.1016/0010-0277(87)90018-7
- Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology*, 35, 1–9.
- Greenauer, N., & Waller, D. (2008). Intrinsic array structure is neither necessary nor sufficient for nonegocentric coding of spatial layouts. *Psychonomic Bulletin & Review*, 15, 1015–1021. doi:10.3758/PBR.15.5.1015
- Hatzipanayioti, A., Galati, A., & Avraamides, M. N. (2016a). The protagonist's first perspective influences the encoding of spatial information in narratives. *The Quarterly Journal of Experimental Psychology*, 69, 506–520. doi:10.1080/17470218.2015.1056194
- Hatzipanayioti, A., Galati, A., & Avraamides, M. N. (2016b). Updating spatial relations to remote locations described in narratives. *Memory & Cognition*, 44, 1259–1276. doi:10.3758/s13421-016-0635-6
- Hölscher, C., Tenbrink, T., & Wiener, J. M. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition*, 121, 228–247. doi:10.1016/j.cognition.2011.06.005
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117. doi:10.1016/0010-0277(96)81418-1
- Humphreys, G. W., & Sui, J. (2015). The salient self: Social saliency effects based on self-bias. *Journal of Cognitive Psychology*, 27, 129–140. doi:10.1080/20445911.2014.996156

- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007
- Keehner, M., Guerin, S. A., Miller, M. B., Turk, D. J., & Hegarty, M. (2006). Modulation of neural activity by angle of rotation during imagined spatial transformations. *Neuroimage*, *33*, 391–398. doi:10.1016/j.neuroimage.2006.06.043
- Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition*, *114*, 72–88. doi:10.1016/j.cognition.2009.08.015
- Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, *39*, 1–20. doi:10.1006/jmla.1998.2563
- Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric bias of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*, 46–49. Retrieved from <http://www.jstor.org/stable/20182500>
- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834. Retrieved from <http://www.jstor.org/stable/40986593>
- Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.
- Li, X., Carlson, L. A., Mou, W., Williams, M. R., & Miller, J. E. (2011). Describing spatial locations from perception and memory: The influence of intrinsic axes on reference object selection. *Journal of Memory and Language*, *65*, 222–236. doi:10.1016/j.jml.2011.04.001
- Loomis, J. M., Lippa, Y., Klatzky, R. L., & Golledge, R. G. (2002). Spatial updating of locations specified by 3-D sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 335–345. doi:10.1037/0278-7393.28.2.335
- 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*, 3–42.
- Marchette, S. A., & Shelton, A. L. (2010). Object properties and frame of reference in spatial memory representations. *Spatial Cognition & Computation*, *10*, 1–27.
- May, M. (2004). Imaginal perspective switches in remembered environments: Transformation versus interference accounts. *Cognitive Psychology*, *48*, 163–206. doi:10.1016/S0010-0285(03)00127-0
- McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Lecture notes in artificial intelligence: Spatial cognition III* (pp. 174–191). Berlin: Springer-Verlag.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Journal of Memory and Language*, *49*, 201–213. doi:10.1016/S0749-596X(03)00028-7
- Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 162–170. doi:10.1037/0278-7393.28.1.162
- Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 142–157. doi:10.1037/0278-7393.30.1.142
- Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A. C., De Ruiter, J. P., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition*, *111*, 46–54. doi:10.1016/j.cognition.2008.12.004
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 167–226. doi:10.1017/S0140525X04000056
- Proulx, M. J., Todorov, O. S., Aiken, A. T., & de Sousa, A. A. (2016). Where am I? Who am I? The relation between spatial cognition, social cognition and individual differences in the built environment. *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.00064
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rieser, J. J., Guth, D. A., & Hill, E. W. (1986). Sensitivity to perspective structure while walking without vision. *Perception*, *15*, 173–188.
- Roche, J., Dale, R., & Kreuz, R. J. (2010). The resolution of ambiguity during cognition: More than mere mimicry? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 206–211). Austin, TX: Cognitive Science Society.
- Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language*, *74*, 46–76. doi:10.1016/j.jml.2014.04.003
- Ryskin, R. A., Wang, R. F., & Brown-Schmidt, S. (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition*, *147*, 75–84. doi:10.1016/j.cognition.2008.12.004
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. A., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what others see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1255–1266. doi:10.1037/a0018729
- Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1–24. doi:10.1016/0010-0277(93)90060-9
- Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about location? *Discourse Processes*, *20*, 219–247.
- Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Eds.), *Spatial language and dialogue* (pp. 23–39). Oxford: Oxford University Press.
- Shelton, A. L., & McNamara, T. P. (2001). Visual memories from nonvisual experiences. *Psychological Science*, *12*, 343–347. Retrieved from <http://www.jstor.org/stable/40063643>
- Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, *1*, 260–273. doi:10.1111/j.1756-8765.2009.01018.x
- Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1105–1117. doi:10.1037/a0029792

- Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, 35, 371–391. doi:10.1006/jmla.1996.0021
- Tversky, B., & Hard, B.-M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, 110, 124–129. doi:10.1016/j.cognition.2008.10.008
- Tversky, B., Lee, P., & Mainwaring, S. D. (1999). Why do speakers mix perspectives? *Spatial Cognition and Computation*, 1, 399–412. doi:10.1023/A:1010091730257
- Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and Cognitive Neuroscience Reviews*, 4, 96–118.