

**Social and configural effects on the cognitive dynamics of perspective-taking**

Alexia Galati<sup>1, 2, 3</sup>, Rick Dale<sup>1, 3</sup>, Nicholas D. Duran<sup>4</sup>

<sup>1</sup> Cognitive and Information Sciences, University of California Merced, USA

<sup>2</sup> Department of Psychology, University of Cyprus, Cyprus

<sup>3</sup> Department of Communication, University of California Los Angeles, USA

<sup>4</sup> School of Social and Behavioral Sciences, Arizona State University, USA

Corresponding author:

Alexia Galati

5200 N. Lake Road  
Cognitive and Information Sciences  
University of California Merced  
Merced, CA 95343  
USA

agalati@ucmerced.edu

Word count: 28,993 words

### Abstract

We examined how perspective selection and its unfolding dynamics are shaped by the alignment of configural cues with social perspectives (the egocentric and the conversational partner's perspectives). Listeners received instructions (e.g., "Give me the folder on the right / left / front / back") from a simulated partner, whose depicted position around a table varied. Listeners selected from pairs of objects that were always aligned with themselves (ego-aligned, in Experiment 1a) or with the speaker (other-aligned, in Experiment 1b). In Experiment 2, listeners selected objects from triangular 3-object configurations whose orientation varied across trials, such that it was aligned with the egocentric (ego-aligned), the speaker's (other-aligned), or neither perspective. When the configural cue was paired with the speaker's perspective (whether consistently, Experiment 1b, or inconsistently, Experiment 2), listeners were more likely to adopt that perspective. Despite this preference, listeners who adopted the other-centric strategy did not experience facilitation from the other-aligned configural cue, as revealed by their mouse-trajectories. However, they did stabilize on that strategy more quickly compared to when the cue was ego-aligned (in Experiment 1b vs. 1a). When the configural cue varied continually (in Experiment 2), other-centric responders did exhibit sensitivity to that cue, making longer and more complex trajectories on neither-aligned configurations. Varying the alignment of the configural cue continually in Experiment 2 also had a drastic impact on the interpretation of the spatial terms *front-back*, which were interpreted according to a reversed mapping compared to the previous experiments. Two follow-up experiments further clarify how the configural cues in Experiments 1 and 2 were likely perceived. Altogether, our findings indicate that configural cues can promote

the likelihood of other-centric responding, contribute to the stabilization of perspective strategy, influence the interpretation of spatial language, and can have a selective impact on the dynamics of language users' responses.

*Keywords:* perspective-taking, audience design, spatial instructions, spatial cognition, mouse-tracking, cognitive dynamics

## 1. Introduction

During various tasks, from navigation to social interaction, humans may consider different perspectives. A perspective can bear on the self, known as the *egocentric perspective* (“left” = “my left”), and in most tasks this perspective coincides with one’s sensorimotor perspective—the perspective capturing self-to-object relations in the immediate environment. But humans can also accommodate a task partner, and take that partner’s perspective (“left” = “their left”). This is sometimes referred to as the *other-centric perspective*. In this paper, our goal is to uncover how different cues shape the cognitive processes involved in perspective-taking. To do so, we adapt a perspective-taking task that tracks computer-mouse movements during perspective choice. By tracking the *dynamics* of perspective choice, in the streaming x, y coordinates of computer-mouse movements during the task, we aim to refine our understanding of how cue integration works.

Many critical cues about perspective reside in the environment. Humans, like all navigating animals, use environmental cues to orient themselves and act in the world. Such configural cues include global features concerning the environment’s geometry and symmetry (Shelton and McNamara, 2001; Tverksy, 1981), its salient axes (e.g., those formed by prominent streets, Werner & Schmidt, 1999), and the slope of its terrain (Nardi et al., 2012; Weisberg & Newcombe, 2014), among many more. Beyond such global configural cues, relevant information about how to interact with the world also comes from more local configural cues, such as the direction and orientation of objects in the environment (e.g., Burigo & Sacchi, 2013), the internal elongated axes of the objects

(Quinlan & Humphreys, 1993; Sekuler & Swimmer, 2000), and the affordances of those objects (Gibson, 1979; Costantini et al., 2010).

Importantly, as social animals, we also use cues about the location and orientation of others in space to guide our language use and actions (e.g., Galati et al., 2013; Özyürek, 2002). In countless everyday scenarios, the alignment of social perspectives with configural features can provide useful information about others' likely intentions, given the function and affordances of objects, and even cultural convention. For example, when sharing a meal with others, we may take into account the perspective of a dinner guest and place a serving spoon in a location appropriate for their taking a turn with a food item: with the spoon's handle turned towards our guest and the implement's "head" located near or in the food item in question. This example underscores that we routinely take into account the orientation of others and the orientation of objects in space. However, the way in which we do so remains underexplored.

In the current research, we examine the potential interaction of configural and social cues on perspective selection in a task where the linguistic descriptions of space are ambiguous. Specifically, participants receive verbal instructions from a social partner (albeit simulated) to select one of two or three candidate objects in a common visual space. In target trials, it is ambiguous whether the target object should be selected based on the partner's perspective ("other-centric) or from the participant's perspective ("ego-centric"). Processing ambiguous spatial descriptions in everyday language use is not uncommon, given the multiplicity of options for how spatial terms can be mapped onto space, including the availability of relative/person-centered and absolute/geocentric terms (e.g., Levinson, 2003). Moreover, given constraints of our task, where participants cannot

explicitly ask about intended perspective, participants must make a spontaneous choice. What we are most interested in is how this perspective choice, as well as the accompanying cognitive difficulty in making the choice—reflected in the participants' mouse movements—is influenced by subtle changes in the configural organization of the objects. Specifically, we investigate the effects of the convergence between configural (directional or geometric) features of the scene and the participant vs. the task partner's perspective.

In what follows, we review evidence concerning the influence of configural cues on spatial reasoning and spatial language use (section 1.1). We then consider how the extant literature addresses the potential integration of such configural cues with social cues about the task partner (section 1.2). As we will point out, with the exception of a few studies using mouse-tracking and eye-tracking methods (section 1.3), little is known about the cognitive dynamics of that integration process. For example, one of the questions is whether configural cues that are spatially aligned with the task partner's perspective facilitate responses from that perspective, as indicated by increased preference for that perspective and more efficient processing. The current study, which we describe in more detail in section 1.4, addresses precisely these questions in a mouse-tracking paradigm.

### **1.1. The role of configural cues on spatial reasoning**

As we have suggested, configural properties contribute to perspective selection when reasoning about previously experienced scenes. In the domain of spatial memory, configural cues—such as the environment's geometry and geometric properties of objects in the environment—have been shown to influence the ease with which people reason

from imagined perspectives about spatial relationships in that environment. In the absence of configural cues, people are typically fastest and most accurate to make spatial judgments from their initially experienced, egocentric viewpoint (Shelton & McNamara, 2001). But in the presence of salient configural cues, the egocentric preference can be overridden, with non-egocentric perspectives exhibiting facilitation instead. For example, people are fastest or most accurate to reason from non-egocentric perspectives when those perspectives are reinforced by the axis of the environment's geometry (Shelton & McNamara, 2001), by the orientation the configuration's constituent objects that have intrinsic axes (i.e., the objects having an intrinsic *front-back*, Marchette & Shelton, 2010), and by the intrinsic axis of the spatial configuration arising from its symmetry (i.e., the symmetrical shape formed by the objects, Mou & McNamara, 2002; Li et al, 2011) or from its orthogonality (i.e., the number of right angles in that array, Richard & Waller, 2013). These findings have been taken to suggest that people maintain spatial relations in memory around a preferred perspective or preferred "organizing direction" (e.g., McNamara, 2003; Mou, McNamara, Valiquette, & Rump, 2004).

Beyond the domain of memory, the contribution of configural cues to spatial reasoning has also been examined in the domain of "reference frame" selection during spatial language interpretation and production (e.g., Carlson, 1999). A reference frame, in its broadest characterization, is thought to be a representation of a coordinate system for organizing spatial relations, consisting of a set of axes that define space and including parameters such as an origin, scale, direction, and orientation (Logan & Sadler, 1996). A confluence of evidence suggests that geometric properties of objects that are part of spatial configurations play an important role in how people use and interpret spatial

descriptions (Burigo & Sacchi, 2013; Burigo, Coventry, Cangelosi & Lynott, 2016; Carlson & Van Deman, 2008; Carlson-Radvansky & Irwin, 1994; Carlson-Radvansky & Logan, 1997). For instance, when objects are presented in non-canonical vs. canonical orientations (e.g., an “upside-down” pumpkin), language users take longer to formulate descriptions of those scenes or to respond to instructions (e.g., “the pumpkin is above the strawberry”) by placing objects at the correct location (Burigo & Sacchi, 2013). Recent computational modeling work further underscores the contribution of contextual information in the environment to reference frame selection, including configural cues about the axes of symmetry and the geometric properties of objects (e.g., about their direction and orientation) (Schultheis & Carlson, 2017).

## **1.2. Attributional cues and integration with configural cues**

Despite the evidence presented so far that configural cues individually contribute to perspective selection, little is known about how configural cues might interact with other cues, including social ones. Social cues—such as the partner’s viewpoint or social attributions about the partner’s ability to contribute to the task—have also been shown individually to influence memory for spatial arrays (Galati, Michael, Mello, Greenauer, & Avraamides, 2013; Shelton & McNamara, 2004), the interpretation of spatial expressions (Duran, Dale, & Kreuz, 2011; Mainwaring et al, 2003) and the production of spatial expressions (Schober, 1993, 1995, 2009).

For example, social attributions about the task partner serve as a contextual cue that modulates the listeners’ perspective strategy. Duran, Dale, and Kreuz (2011) demonstrated that listeners who responded to spatial instructions (e.g., “Give me the folder on the left”) that were ambiguous in some visual contexts, were more likely to



interpret these instructions from the partner's perspective when they believed that the partner did not know their viewpoint, whereas they were more likely to interpret such descriptions egocentrically when they believed their partner was real (vs. simulated) and thus had the capacity to adopt a perspective other than their own.

This is compatible with findings that speakers adapt their spatial descriptions, by including more spatial details or by being more likely to adopt their partner's perspective, when they perceive the partner to be limited in terms of their ability to contribute to the task (e.g., when the partner is unfamiliar with the environment, Hölscher, Tenbrink, & Wiener, 2011; not able to interact contingently, Schober, 1993; or has worse spatial abilities than they do, Schober, 2009). Collectively, these findings suggest that social cues can provide pragmatic motivation for language users to override the egocentric perspective, despite the presumed associated cognitive cost of adopting the partner's perspective. Such perspective strategies are compatible with the proposal that interlocutors assume shared responsibility for mutual understanding, and take into account the relative difficulty of perspective-taking—for themselves and for their conversational partners (Clark & Wilkes-Gibbs, 1986; Clark, 1996). In doing so, they select strategies that are intended to maximize their efficiency of communication and minimize their collective effort.

Beyond demonstrations that configural cues and social cues individually influence perspective selection, to our knowledge, only recent work by Galati and Avraamides (2015) has systematically parameterized both types of cues to assess their joint contribution to perspective selection. In that study, speakers had to describe a configuration with an axis of symmetry to a partner. Critically, that axis of symmetry was

aligned with the speaker's viewpoint, their partner's viewpoint, or neither viewpoint. The speaker's linguistic choices during the description of the configuration, as well as their memory performance (prior to descriptions) were examined. The findings revealed that the speakers' spatial judgments about the previously studied configuration were influenced by the convergence of social and configural cues. When the configuration's axis was aligned with the egocentric perspective, that perspective exhibited facilitation during spatial judgments. When the configuration's axis was aligned with the partner's perspective and this was known in advance, the partner's perspective showed facilitation relative to other headings. Similar patterns were observed for the perspective from which speakers described the configurations to the partner, with speakers using more frequently other-centric descriptions (e.g., "to your left") relative to the egocentric ones (e.g., "to my right") when the configuration was aligned with the partner.

The work by Galati and Avraamides (2015) demonstrates that multiple sources of information factor into perspective selection (see also Galati & Avraamides, 2013), at least over a relatively long timescale, as when speakers encode and maintain spatial information in memory and later describe that information (from memory) to a partner. Nevertheless, this work—along with much of the work described so far—is limited by its focus on speakers' overall distributions of perspective choices during language or spatial reasoning. These studies do not reveal how that perspective choice unfolds and stabilizes over time. Although people may be influenced by both social and configural cues when maintaining spatial information to guide their end-point perspective selection, how that selection process unfolds—e.g., during the interpretation of a spatial description—is generally unknown.

### 1.3. Cognitive dynamics of perspective-taking and cue integration

So far, only a few studies have taken into account the cognitive dynamics of spatial perspective-taking by using methods that permit the fine-grained sampling of behavior, such as mouse-tracking (Brennan, 2005; Duran, Dale, & Kreuz, 2011; Duran & Dale, 2014), or eye-tracking (Ryskin et al., 2014; Ryskin, Wang, & Brown-Schmidt, 2016; Galati, Diavastou, & Avraamides, 2018; Barr & Keysar, 2002; Hanna & Brennan, 2007; Wu et al., 2013). Studies examining the time-course of perspective-taking (whether spatial or non-spatial) often address explicitly a lively debate in psycholinguistics concerning the precedence of the egocentric perspective. According to some researchers, language users initially default to using egocentric information, and consider information relevant to the partner only later and as needed, by monitoring the partner's behavior (Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998; Keysar, Barr, Balin, & Paek, 1998; Shintel & Keysar, 2009; Kronmüller & Barr, 2007). In contrast, other researchers suggest that language users can take into account information about the partner during early processing (e.g., Brennan & Hanna, 2009; Metzing & Brennan, 2003; Ryskin, Wang, & Brown-Schmidt, 2016).

In spatial perspective-taking, some studies provide evidence consistent with early egocentrism, by demonstrating that the egocentric perspective is co-activated when adopting another's spatial perspective. For example, in the previously described study by Duran and colleagues (2011), listeners who responded egocentrically on ambiguous instructions (egocentric responders) were faster and made more direct mouse trajectories (to the egocentric option) than those responding from the task partner's perspective (other-centric responders). Importantly, other-centric responders experienced interference

from the egocentric perspective, as evidenced by signatures of their mouse movements, such as deviations of the mouse cursor toward the competitor “egocentric” object choice (also see Duran & Dale, 2014). Converging evidence from a similar task demonstrates that other-centric responders experience greater interference from the competing perspective than egocentric responders, as indicated by their first eye-gaze fixations on the configuration’s objects (Galati, Diavastou, & Avraamides, 2018).

In contrast, other spatial perspective-taking studies provide evidence that language users can readily appreciate their conversational partner’s perspective. For instance, Ryskin and colleagues (2014) showed that listeners need not exhibit a bias toward an egocentric interpretation, even when their partner’s perspective is counteraligned from their own. Related findings, also taken as evidence against an egocentric default, show that the other-centric perspective is at times processed readily (presumably automatically) even when it should be disregarded. In one such study, participants who had to judge the number of dots they could see while ignoring the perspective of an avatar 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), even though that perspective was irrelevant to the task (Samson et al., 2010). These findings suggest that another’s visuospatial perspective need not incur a cognitive cost, and that in fact may be hard to ignore (see also Tversky & Hard, 2009).

In our view, these seemingly contradictory sets of findings can be reconciled by shifting the focus away from the debate about the psychological reality of an early egocentric default, and towards uncovering how perspective selection stabilizes under different constraints, as reflected by signatures of the language users’ unfolding behavior.

This view accommodates findings where, in some contexts, the egocentric perspective exhibits initial precedence, as well as contexts in which it does not (for greater explanation, see Duran, Dale, & Galati, 2016). With these considerations in mind, we investigate the time-course of perspective resolution amidst varying social and configural cues.

#### **1.4. Current study**

The present work builds on the experimental paradigm of Duran, Dale, and Kreuz (2011). In that study, listeners received verbal instructions from a simulated partner to select an object from a configuration of objects displayed on a round table on their computer screen. Listeners then had to drag that object to their task partner's position, which was indicated around the table ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , or  $270^\circ$ ) along with their own ( $0^\circ$ ). Depending on where the partner was located and the wording of the instructions, the intended object was often ambiguous, requiring participants to interpret the instruction taking an egocentric or other-centric perspective (i.e., seeing the table from the partner's perspective to interpret the instructions from the partner's point of view). Critically, we extend this work by manipulating not only the partner's location, but also the alignment of the intrinsic axis of the to-be-selected target objects with the listener (Experiment 1a) or the partner (Experiment 1b). For the case of orientation with the listener, objects were aligned with the egocentric perspective ("ego-aligned"; a direct replication of Duran et al., 2011, Study 1), and for the case of orientation with the partner, the objects were aligned with the other-centric perspective ("other-aligned"). In a second experiment (Experiment 2), rather than manipulating the orientation of individual objects between participants, we manipulated, within participants, the orientation of the

objects taken as a coherent array. This array was formed by arranging the three objects displayed on the table in a triangular configuration. By doing so, the intrinsic axis of the configuration (based on the axis of symmetry formed by the isosceles triangle), could be oriented either with the listener, the partner, or neither depending on trial. Two follow-up experiments were also conducted—one for each main experiment—to examine whether individuals indeed apprehend the alignment of configural cues with either task partner in the way we had intended.

In our main Experiments (1a, 1b, and 2), we assessed the time course of perspective selection and the listeners' response strategies, by examining both their computer-mouse trajectories during the interpretation of spatial instructions and their aggregate perspective choices.

We should note that we take “perspective choice” to be equivalent to “reference frame selection” in this spatial perspective-taking task. However, we opt for the former term and for a broader conceptualization of “perspective” for two reasons. First, the candidate reference frames most relevant to our predictions are both relative reference frames (i.e., person-centered), with either the listener or the speaker as the origin point or “relatum” (i.e., the object or person relative to which other objects are described, Tenbrink, 2007). Since our predictions focus on the competition of these two relative reference frames, it is more parsimonious to refer to them simply as the “egocentric” and “other-centric” perspective. Second, as our literature review suggests, we consider the present work to inform not only spatial perspective-taking but conversational perspective-taking more broadly. In this view, we take the predictions laid out below to extend to other perspective-taking contexts, where different kinds of cues may be reinforcing the

egocentric or other-centric perspective. In those contexts, “perspective” could refer to not only spatial viewpoint, but also to (shared vs. non-shared) visual access, knowledge, and so on.

**1.4.1. Predictions.** Based on the work by Galati and Avraamides (2015), we predicted that, in terms of the listeners’ overall perspective strategies, listeners would be more other-centric when a configural cue reinforced that perspective (Experiment 1b) compared to when it did not (Experiment 1a). Beyond this shift in strategy, we anticipated that other-centric responding might also show facilitation when it is reinforced by a configural cue. That is, we expected that other-centric responders in Experiment 1b might exhibit faster and more direct mouse-trajectories than other-centric responders in Experiment 1a.

In Experiment 2, we expected that, despite the varying alignment of the configural cue, listeners would still stabilize on a consistent strategy, rather than switching strategies across trials depending on whether objects are aligned with the self or the partner. Although in some circumstances, particularly when language users can interact freely, switching perspectives can be effective for coordinating (e.g., Tversky, Lee, & Mainwaring, 1999; Brennan & Clark, 1996), language users generally adhere to a consistent spatial perspective or conceptual strategy (Garrod & Anderson, 1987; Duran, Dale, & Kreuz, 2011). Regardless of the overall distribution of egocentric and other-centric responders in Experiment 2, we anticipated that the configural cue—the orientation of the entire array of objects—would facilitate responses when it was aligned with the perspective from which the listener responded. In line with our predictions about Experiments 1a and 1b, we expected that other-centric responders might show facilitation

when the alignment of the configuration supported the other-centric perspective (i.e., other-aligned configurations vs. ego- or neither-aligned). Similarly, we expected that egocentric responders might also exhibit facilitation on ego-aligned configurations, or alternatively might be impervious to the alignment of the configural cue insofar as both the configuration's orientation and the speaker's position are irrelevant to their perspective strategy.

Together, the findings of these experiments can help clarify whether egocentric and other-centric perspective strategies involve processing asymmetries, and how these are modulated by contextual cues, including configural features of the task.

## **2. Method: Experiments 1a and 1b**

### **2.1. Participants**

We collected data from 200 users on Amazon Mechanical Turk<sup>1</sup>. For Experiment 1a, 38 users reported their gender as female, 50 as male, 2 as non-binary, and 10 did not report their gender; for Experiment 2b, 39 users reported their gender as female, 55 as male, and 6 did not report it. The participants' mean age was 36.87 ( $SD= 9.62$ ) for Experiment 1a (range: 19–61), and 36.07 ( $SD= 10.73$ ) for Experiment 1b (range: 20–73).

Data from 12 users were discarded, as they appeared to be repeated sessions based on the users' worker ID or IP address; for these cases, only the first session was included.

---

<sup>1</sup> As we describe in section 2.2., instructions to participants did not mention mouse-tracking and did not require participants to use a mouse. The majority of participants nevertheless used a mouse (Experiment 1a: 75%; Experiment 1b: 82%; Experiment 2: 85%). The response movements of those using trackpads (and those who did not specify their response mode) were sampled in the same way as those using a mouse. Given the same sampling procedures and the fact that relatively fewer people used trackpads, it is common practice for studies examining action dynamics to include both mouse and trackpad users in the analyses. Moreover, there is some evidence that the users' response mode is not associated with significant differences in the interpretation of their action dynamics: in a dual-response task, the action dynamics of MTurk participants using mouse or trackpads (which restrict movement in 2D) and of participants in the lab using Nintendo Wiimote (which allows movement in 3D) revealed negligible differences and provided converging interpretations (Duran, 2011).



This resulted in 188 unique participants: 93 for Experiment 1a, and 95 for Experiment 1b. The task was available to participants who were over 18 years old and located in the United States. Participants were compensated \$1.50 for approximately 15 minutes of effort.

## **2.2. General setup**

At the start of the experiment, participants were presented with an initial screen with information about the study, through which informed consent was also obtained. Participants were told that the data collected would remain confidential and that they would earn payment for their participation according to the rules and regulations of Amazon Mechanical Turk.

On the following screen, participants were informed that they would be receiving instructions over their audio speakers and that these instructions would be from a “simulated task partner”. They and their partner would be “seated” around a simulated table, with their respective positions marked by labels (“You” and “Partner”), and that during the task the partner would move around the table while they remained stationary. Furthermore, they were told that the partner’s instructions would be a request for one of two objects (simulated folders) laid out on the table. The instructions did not mention that participants’ mouse movements would be tracked, and there was also no mention of a time pressure to respond.

At the beginning of each trial, participants only saw the top-down view of a circular table. To initiate the display of the folders and partner’s position, participants used their mouse cursor to click on the word “GO” at the bottom-center of the screen. The folders that appeared were arranged in one of three ways: horizontally (across the

table), vertically (across the table), or diagonally (with a folder on the bottom left and the other on the top right, or a folder on the top left and the other on the bottom right, as in Figure 1). The participant's (i.e., the listener's) perspective was indicated by the word "You" and was fixed at  $0^\circ$ , whereas the partner's (i.e., the speaker's) perspective changed across trials ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ ) and was indicated by the word "Partner." Next to the "Partner" label was also a small box at which the listeners had to deposit one of the two folders depending on the partner's instructions. (When the speaker's perspective was at  $0^\circ$ , the words "Partner" and "You" were displayed at the bottom of the table on either side of the box.) Importantly, for participants in Experiment 1a, the intrinsic axis of the folders, specifically the top-bottom axis of the folders, was always aligned with the listener, facing  $0^\circ$ ; and for participants in Experiment 1b, the intrinsic axis of the folders was always aligned with the speaker, facing  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  depending on the speaker's location.

Once participants clicked to proceed, they heard a male's voice ask for one of the folders, always beginning with the phrase, "Give me the folder on the ..." and ending with one of four possible spatial directions, "right," "left," "front," or "back." Participants could then select and drag one of the two folders using their mouse cursor to the speaker's location around the table. Trajectories (beginning from a fixed mouse position at the bottom-center of the screen) captured the entire flow of the mouse movement, yielding behavioral measures that have been shown to reflect the dynamics of cognition (see section 2.4 below).

After the completion of all trials, participants were surveyed about their experience in the task environment, responding to questions about task difficulty (Q1),

about whether they believed they were interacting with a real person (Q2), and about whether they found the chat environment to be interactive (Q3)<sup>2</sup>. Finally, they were fully debriefed regarding the purpose of this study.

### 2.3. Stimuli

There were a total of 40 experimental trials, with 20 control and 20 critical trials. The control 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. Conversely, the critical trials were constructed such that interpretation from the listener's and the speaker's perspective would result in different object choices. This ambiguity was elicited by a combination of the speaker's instructions, the speaker's position around the table, and the orientation of the folders. For example, when a speaker is at 90° and the folders are arranged diagonally as in Figure 1a, an instruction of "Give me the folder on the right," would be ambiguous in terms of whose perspective should be taken (critical trial), whereas an instruction of

---

<sup>2</sup> Responses to the questions about beliefs about the partner (Q2) and the interactivity of the environment (Q3) confirmed that participants believed that the instructions came from a simulated partner, as opposed to a real partner interacting with them in real time. In Q3, some participants stated explicitly that they believed that the partner was a real person, but that they were not interacting with him in real time; some participants stated that they treated the simulated partner as a real partner. Q1, which concerned task difficulty, afforded some insights into how participants reasoned about the task. Participants verified that the task was not difficult, but their majority *did* note the challenge of selecting a response strategy, given the ambiguity of the instructions (e.g., mentioning that instructions were "confusing", or that their intended perspective could not be clarified without being able to ask a question). A number of participants identified their strategy (i.e., egocentric or other-centric) and some even justified their choice (e.g., "In real life I assume the partner would structure his demands based on my view", or "Since I could not communicate, I went as though the instructions were from the other person's perspective."). In addition, a few participants noted the difficulty of reasoning from the speaker's position (e.g., stemming from the fact that the speaker moved around the table, or that the speaker on some trials was counter-aligned with them), as well as difficulty with interpreting *front-back* instructions (e.g., whether "front" refer to "closer" or "farther"). As we will show, our results are broadly consistent with the participants' metacognitive reports about task difficulty. The survey responses are contained at the end of the raw data file of each participant, which are available in our GitHub and OSF repositories (see section 3.3).

“Give me the folder at the front,” would result in the same object choice because both perspectives coincide (control trial).

Of the critical trials, 12 involved cases for which the speaker was at 180° relative to the listener, 8 involved cases for which the speaker was at 90° offset to the right of left (i.e., either to the right at 90° or to the left at 270°). Of the control trials, 12 involved cases for which the speaker’s perspective was the same as the listener’s (at 0°), and another 8 involved cases for which the speaker was at a 90° offset (at 90° or 270°). Thus, trials with the speaker at a 90° offset could be control or critical trials, depending on the accompanying instruction, whereas those with the speaker at 0° were always control trials and those with the speaker at 180° were always critical trials.

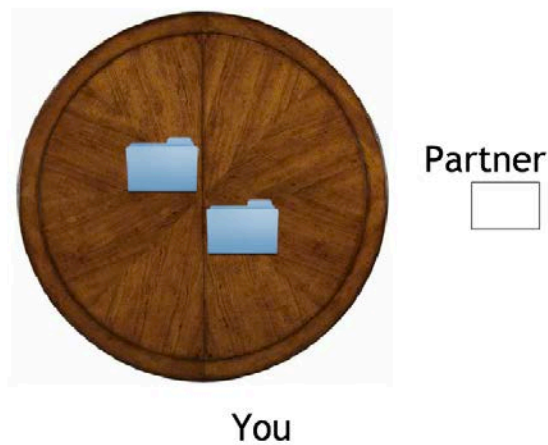
There were 10 trials from each type of spatial instruction (i.e., “Give me the folder on the ... [*right / left / front / back*]”), which were represented equally among control and critical trials. The order in which the trials were presented was determined by one of 5 pseudo-randomized lists; participants were assigned randomly to a particular list. Each list was constructed such that there were no more than three consecutive trials of the same type (critical or control). There was a single practice trial, which was identical across all lists (a control trial, with the speaker at 0°, a diagonal arrangement of the folders, and a *right* instruction).

In Experiment 1a the intrinsic axis of the folders (i.e., top-bottom axis of the folders) was always aligned with the perspective of the listener at 0°, exactly as in Duran, Dale, and Kreuz (2011). In contrast, in Experiment 1b, the intrinsic axis of the folders was always aligned with the perspective of the speaker. In other words, the folders changed orientation across trials, depending on whether the speaker’s position was at 0°,

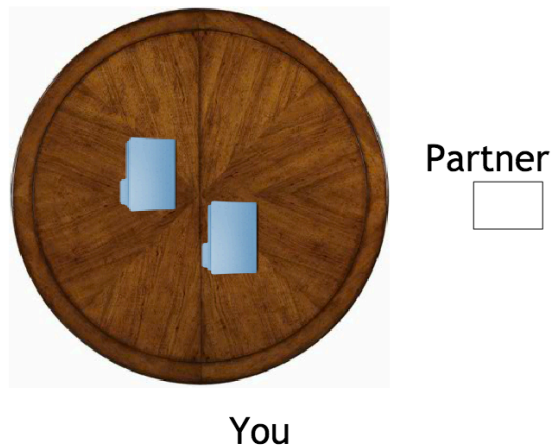
90°, 180°, or 270°. Beyond the change of orientation, the locations of the objects and the accompanying spatial instructions were otherwise identical in the two experiments.

*Figure 1. Examples of trials from Experiments 1a and 1b. The participant's perspective was always indicated by the word "You" at 0°, and the participant received verbal instructions to choose a folder from an imagined task partner, the speaker. In these examples, the speaker's perspective is at 90°, indicated by the word "Partner". These configurations would result in a critical trial when paired with the instruction "Give me the folder on the right" and in a control trial with the instruction "Give me the folder at the front." The configuration in Figure 1a is from Experiment 1a, where the folders were always aligned with the listener (ego-aligned), whereas the one in Figure 1b is from Experiment 1b, where the folders were always aligned with the speaker (other-aligned).*

**1a.**



**1b.**



#### 2.4. Behavioral tracking using computer-mouse trajectories

For each trial, the x, y coordinates of the participants' cursor were sampled at approximately 25 ms (a sampling rate of 40 Hz), from the moment they clicked in the “GO” region until they clicked their initial selection of a folder<sup>3</sup>. This sampling choice ensured that mouse movements, for all participants and all trials, had a common spatial region as the onset. This choice also disregarded any mouse movements that may have occurred while listening to the spatial instruction. As we have described in section 2.2, prior to clicking “GO” and while participants heard the speaker's instruction, only a blank table was displayed on the screen. Since the locations of the folders and the speaker's position were unknown, mouse movements during this timeframe are unlikely to reflect participants' response strategy.

<sup>3</sup> Our choice for sampling trajectories excludes any cases where participants picked up a folder and then dropped it to select the other folder. This behavior, however, was rare: it amounted to 2.55% of the trials in Experiment 1a, 1.69% in Experiment 1b, and 2.63% in Experiment 2. Although also indicative of perspective competition, these folder switching events were of such low incidence so as not to alter dramatically our conclusions about perspective choice or the dynamics of the mouse trajectories (especially since the same perspective competition could still be evident in the initial trajectory to the first folder choice).

These mouse movements yielded the dependent measures on which we focus our analyses. We chose three mouse-movement measures: response time, movement distance, and directional shifts (or “x-flips”). These will be used for all experiments, and so we summarize them in detail here.

First, *response time* was simply the duration of time needed to render the perspective choice. Because this is a mouse-tracking task, response time is not a button press, but rather an explicit movement of the computer mouse to one of the candidate objects. We measured response time as the *milliseconds* required to make this first choice.

Second, *distance* can serve as a measure of trial difficulty. If participants vacillate on a given trial, they may show a more divergent series of x,y coordinates. Such difficulty will be indicated by higher total distance measured in *pixels*; this is the measure we used in Experiments 1a and 1b. Note that the position of a given object in the task may inflate or deflate this pixel measure simply by virtue of its position, and not cognitive dynamics. In Experiment 2, this issue was relevant as we used different configuration types that inadvertently involved differences in the pixel distance to the egocentric and other-centric object choice. To control for this, in Experiment 2, we use instead the *divergence from optimal distance*, taking distance in pixels to be how much more extended a trajectory is relative to an assumed straight line to the object that was chosen.

Finally, we added an additional measure of selection conflict, namely the number of changes of direction seen in a trajectory of x,y coordinates. This is sometimes known

as “x-flips,” and is a *count* variable of the number of times that the mouse cursor went back and forth on the task environment along the x-axis.

Though we use these three mouse-movement-derived dependent variables, we do not wish to imply that they should be interpreted independently. In fact, each is a “signal” of the decision dynamics that marks the manner in which cue combination is impacting cognitive processes. The measures will, of course, correlate positively: response time will correlate positively with trajectory distance. While we conduct statistical analysis on each measure separately, we expect that the three measures, interpreted *together*, will serve as a portrait of cognitive processes that sheds light on perspective choice.

### **3. Statistical Analysis: Experiments 1a and 1b**

#### **3.1. Data preparation**

For purposes of analysis, we are primarily interested in responses to critical trials, as these involved ambiguous instructions that presented an opportunity for listeners to make an egocentric or other-centric object choice. In aggregate, responses to these critical trials reflect the listener’s perspective strategy. Control trials served primarily as a check that participants were not responding randomly, given that there was only one “correct” folder option that corresponded to both the egocentric and other-centric choice. Unless noted otherwise, all models are based on critical trials, with analysis involving control trials reported in Appendix A.

There are four dependent variables analyzed in Experiment 1a and 1b. Three variables correspond to response behaviors based on trajectory movements: response time, total distance, and directional shifts (x-flips). We also created an additional dependent variable that classified participants into egocentric, other-centric, or mixed



responders based on how they primarily interpreted the ambiguous critical trials. For classification, we computed the proportions of egocentric and other-centric responses of each listener on critical trials. Following Duran, Dale, and Kreuz (2011), if the proportion scores exceeded .70 for one of the two perspective categories, the listener was classified as member of that category; otherwise they were classified as a mixed responder.

We also had theoretical motivation to condense the four instruction types (*front*, *back*, *left*, *right*) into two axes (i.e., sagittal vs. lateral), given evidence that mapping *left* and *right* to appropriate regions of space is slower than mapping *front-back*, perhaps due to the fact that the lateral axis is highly symmetric (relative to a person's body), which makes it harder to differentiate *left-right* relative to linguistic terms associated with the sagittal axis (e.g., Avraamides & Sofroniou, 2006; Franklin & Tversky, 1990). We also had theoretical motivation to condense speaker's position at 90° and 270° into one level, simply labeled 90°. This was done because there was no theoretical reason to expect a difference between a mental rotation or transformation that was in a clockwise vs. counterclockwise direction (a 90° offset in both cases) to adopt the speaker's perspective.

We also removed trials that were excessive in temporal duration. Accordingly, we removed trials that took more than 6000 ms from the initiation of the trial to selection of an initial folder or over 1000 pixels of total distance. These values reflect long tails of the distribution, well over 3 SD of their respective distributions. We chose to omit these trials because these radical divergences from their general distribution may reflect moments when the participant is not fully engaged in the task or understanding the trial. This corresponded to 2.34% of the data (3.10% of critical trials and 1.60% of control trials). Lastly, we also visually inspected the residual plots of the statistical models built for each

dependent variable (see below) to determine deviations from normality and homoscedasticity. If such deviations were pronounced, the dependent variable was log-transformed; this was the case for response times and total distance. It should be noted that investigation of models with untransformed variables revealed effects consistent with the results reported here.

### 3.2. Statistical models

We built separate linear mixed effects models for each of the dependent variables, combining the data from Experiment 1a and 1b. For the model with egocentric perspective choice as dependent variable, the fixed factors included the orientation of the folders (“Orientation”: Experiment 1a - ego-aligned vs. Experiment 1b - other-aligned; a between-subjects factor), the offset of the speaker’s position (“Position”: 180° vs. 90° offset; within-subjects), instruction type (“Instruction”: on a sagittal *front-back* axis vs. a lateral *left-right* axis; within-subjects), and their interactions.

We built an omnibus statistical model to examine the interaction between Orientation, Position, and Instruction, with planned contrasts that compared the proportion of egocentric choices on ego-aligned vs. other-aligned folders (ego-aligned = -0.5, other-aligned = 0.5), on the two speaker positions (90° = -0.5, 180° = 0.5), and on the two types of instruction (sagittal = -0.5, lateral = 0.5). Given that the egocentric perspective choice as a dependent variable is a binary variable, we used a logistic regression model.

For the separate omnibus statistical models with response time, total distance, and directional shifts as dependent measures, the same fixed effects structure of Orientation, Position, and Instruction factors were included, but now also including perspective

preference (“Perspective”: egocentric, other-centric, vs. mixed responder types; between-subjects) and its interaction with the other factors. Perspective preference was coded in terms of planned contrasts that compared the ease of responding of egocentric relative to other-centric responders (egocentric = -0.5, other-centric = 0.5), and of egocentric relative to mixed responders (egocentric = - 0.5, mixed = 0.5).

Moreover, in an exploratory set of analyses, we also included trial order and its interaction with Orientation; these analyses aimed to examine whether object orientation (i.e., Experiment identity) predicted how changes in response behavior evolved over time.

All models shared an initial random effects structure, with intercepts for participants, and following the recommendations of Barr, Levy, Scheepers, and Tily (2013), a full random effect structure that included random intercepts and slopes for Position and Instruction (and their interaction). For the models that also included trial order, random intercepts and slopes were also included. If 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.

In order to assess the listeners’ performance, we used the lme4 library (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2016). For testing the statistical significance of relevant contrasts and interactions, we also employed the R multcomp package (Bretz, Hothorn, & Westfall, 2010). We report an overall measure of captured variance, the coefficients of the predictors, their *t*-values, and indicate their *p*-value significance. Captured variance of overall models is reported as Conditional  $R^2$  variance

explained by fixed and random factors together, which was computed using the MuMIn R statistical package (Johnson, 2014). To compute  $p$ -values from these  $t$ -values, we assume they lie on a two-tailed  $z$ -distribution in the manner described in Mirman (2014).

### 3.3. Data and code sharing

Following the current best practices for open science, we have made the data and code for this project freely available. The de-identified raw data files, along with the code for preparing the data, specifying planned contrasts, and testing the statistical models, are available through our OSF (<https://osf.io/56rqw/>) and GitHub repositories for the project (<https://github.com/alexiagalati/social-configural-dynamics>).

## 4. Results: Experiments 1a and 1b

### 4.1. Perspective choice

First, we compared the distribution counts of egocentric, other-centric, or mixed responders across Experiments 1a and 1b using a Chi-square test. Across the two experiments, the distribution of responders differed significantly,  $\chi^2(2) = 6.93, p < .05^4$ .

In Experiment 1a, where folders were ego-aligned, there were 43 other-centric responders, 33 egocentric responders, and 17 mixed, whereas in Experiment 1b, where folders were other-aligned, there were 59 other-centric responders, 18 egocentric responders, and 18 mixed responders.

Next, we examined listeners' egocentric responses on critical trials by creating an omnibus linear mixed effects model to assess the relationship between folder orientation

---

<sup>4</sup> The distribution of responders in Experiment 1a did not differ significantly from that of Study 1 in Duran, Dale, and Kreuz (2011),  $\chi^2(2) = 2.62, p = .27$ , which used identical materials and procedure.

(Experiment 1a vs. 1b), speaker position (180° vs. 90°), and type of instruction (lateral vs. sagittal). The overall variance captured by the model reported in Table 1 was 80%.

Consistent with the distribution of types of responders, as illustrated in Figure 2, the distribution of egocentric perspective choices differed according to the orientation of the folders manipulated across the two experiments. As reflected in the significant contrast between the two experiments reported in Table 1, the proportion of egocentric responses on critical trials was greater in Experiment 1a, in which the folders were always ego-aligned ( $M = 47\%$ ,  $SD = 50\%$ ) than in Experiment 1b, in which the folders were always aligned other-aligned ( $M = 32\%$ ,  $SD = 47\%$ ).

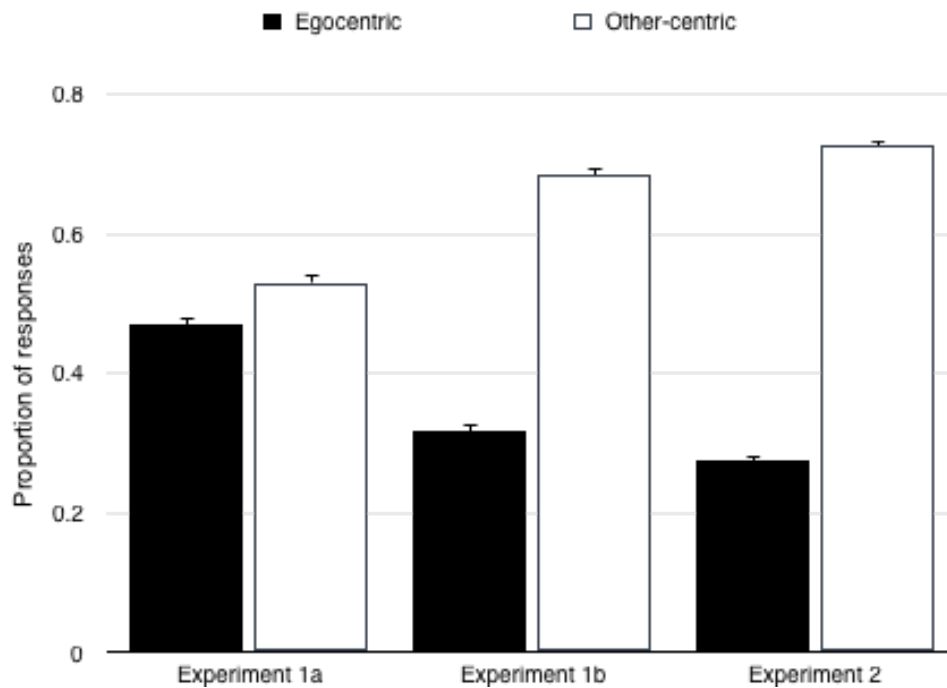
Listeners were consistent in their response strategy regardless of the speaker's position. The type of offset at which the speaker was depicted (90° vs. 180°) did not significantly predict the listeners' egocentric perspective choices. Importantly, this held for both experiments, as suggested by the non-significant interaction of the Orientation and Position contrasts in Table 1.

Indeed, when we examined the overall two-way and three-way interactions of the three factors (folder orientation, speaker position, and type of instruction), by performing likelihood ratio tests between the omnibus model and models that excluded each of the critical two- or three-way interactions, the overall two-way interaction between folder orientation and speaker position was not significant ( $p = .33$ ). Only the overall two-way interaction of folder orientation and type of instruction was significant ( $\chi^2(1) = 25.07$ ,  $p < .001$ ).

As shown in Table 3, the type of instruction (lateral vs. sagittal) influenced the listeners' perspective choices and interacted significantly with the orientation of the

folders. Listeners were overall more likely to make an egocentric choice on *left-right* ( $M = 50\%$ ,  $SD = 50\%$ ) than *front-back* instructions ( $M = 35\%$ ,  $SD = 48\%$ ), and this difference was greater in Experiment 1a, where folders were ego-aligned (lateral: 56% vs. sagittal: 38%) than in Experiment 1b, where the folders were other-aligned (lateral: 35% vs. sagittal: 28%). We unpack this interaction next.

*Figure 2. Proportion of egocentric and other-centric object selections on critical trials in Experiments 1a, 1b, and 2. Error bars indicate standard errors of the mean.*



*Table 1. Mixed-effects models for egocentric object choice, response times (log-transformed), total distance (log-transformed), and directional shifts, with fixed effects for Experiment (Experiment 1a: ego-aligned folders, Experiment 1b: other-aligned folders), perspective preference (ego, mixed, other)—for all but the first model, speaker offset (90°; 180°), and instruction type (sagittal: back-front; lateral: right-left), and their interactions, with the maximal random effect structure possible. For fixed effects and their interactions, we report the coefficient and its standard error, along with the associated z- or t-value and p-value.*

Predictor	Egocentric choice				Response times				Distance				Directional shifts			
	$\beta$	SE	z	p	$\beta$	SE	t	p	$\beta$	SE	t	p	$\beta$	SE	t	p
Intercept	-0.79	0.26	-3.08	0.00	7.16	0.03	268.18	0.00	5.65	0.01	450.48	0.00	1.93	0.07	27.56	0.00
Exp: 1b vs. 1a	-1.75	0.52	-3.39	0.00	-0.00	0.05	-0.02	0.99	-0.02	0.03	-0.65	0.52	0.18	0.14	1.32	0.19
Pref: mixed vs. ego	—	—	—	—	0.03	0.08	0.33	0.74	0.03	0.04	0.78	0.44	0.30	0.22	1.38	0.17
Pref: other vs. ego	—	—	—	—	0.23	0.07	3.57	0.00	0.09	0.03	2.78	0.01	0.23	0.17	1.36	0.18
Offset: 180° vs. 90°	-0.06	0.09	-0.59	0.55	0.09	0.01	6.65	0.00	0.04	0.01	4.22	0.00	0.08	0.05	1.66	0.10
Instruction: lateral vs. sagittal	1.33	0.10	13.35	0.00	0.07	0.02	3.68	0.00	0.02	0.01	1.49	0.14	-0.03	0.05	-0.50	0.62
Exp: 1b vs. 1a * Preference: mixed vs. ego	—	—	—	—	0.17	0.17	1.00	0.32	0.14	0.08	1.82	0.07	1.02	0.44	2.34	0.02
Exp: 1b vs. 1a * Preference: other vs. ego	—	—	—	—	-0.01	0.13	-0.04	0.97	-0.07	0.06	-1.15	0.25	-0.48	0.34	-1.39	0.17
Exp: 1b vs. 1a * Offset: 180° vs. 90°	0.19	0.19	1.04	0.30	0.01	0.03	0.19	0.85	0.02	0.02	0.93	0.35	-0.11	0.10	-1.15	0.25
Pref: mixed vs. ego * Offset: 180° vs. 90°	—	—	—	—	0.03	0.04	0.75	0.45	-0.01	0.03	-0.47	0.64	0.05	0.15	0.32	0.75
Pref: other vs. ego * Offset: 180° vs. 90°	—	—	—	—	0.11	0.03	3.51	0.00	0.10	0.02	4.09	0.00	0.32	0.12	2.78	0.01
Exp: 1b vs. 1a * Instruction: lateral vs. sagittal	-1.05	0.20	-5.30	0.00	-0.07	0.04	-1.94	0.05	-0.02	0.02	-0.85	0.40	-0.10	0.11	-0.93	0.35
Pref: mixed vs. ego * Instruction: lateral vs. sagittal	—	—	—	—	-0.06	0.06	-0.98	0.33	-0.07	0.04	-2.11	0.04	-0.08	0.17	-0.48	0.63
Pref: other vs. ego * Instruction: lateral vs. sagittal	—	—	—	—	0.42	0.04	9.58	0.00	0.13	0.03	4.65	0.00	0.53	0.13	4.06	0.00
Offset: 180° vs. 90° * Instruction: lateral vs. sagittal	-0.35	0.19	-1.83	0.07	0.03	0.03	1.03	0.30	0.04	0.02	2.16	0.03	0.19	0.09	2.05	0.04
Exp: 1b vs. 1a * Pref: mixed vs. ego * Offset: 180° vs. 90°	—	—	—	—	-0.08	0.08	-0.95	0.34	0.01	0.06	0.14	0.89	-0.23	0.30	-0.78	0.43
Exp: 1b vs. 1a * Pref: other vs. ego * Offset: 180° vs. 90°	—	—	—	—	0.03	0.07	0.51	0.61	0.00	0.05	0.03	0.98	0.41	0.23	1.75	0.08
Exp: 1b vs. 1a * Pref: mixed vs. ego * Instruction: lateral vs. sagittal	—	—	—	—	-0.17	0.11	-1.52	0.13	0.06	0.07	0.79	0.43	0.04	0.33	0.12	0.91
Exp: 1b vs. 1a * Pref: other vs. ego * Instruction: lateral vs. sagittal	—	—	—	—	0.25	0.09	2.80	0.01	-0.00	0.05	-0.06	0.95	0.13	0.26	0.51	0.61
Exp: 1b vs. 1a * Offset: 180° vs. 90° * Instruction: lateral vs. sagittal	-0.57	0.38	-1.50	0.13	-0.00	0.05	-0.09	0.93	0.00	0.04	0.10	0.92	0.30	0.18	1.67	0.10
Pref: mixed vs. ego * Offset: 180° vs. 90° * Instruction: lateral vs. sagittal	—	—	—	—	-0.03	0.08	-0.31	0.76	-0.03	0.06	-0.55	0.58	-0.40	0.28	-1.42	0.16
Pref: other vs. ego * Offset: 180° vs. 90° * Instruction: lateral vs. sagittal	—	—	—	—	0.10	0.07	1.51	0.13	0.01	0.05	0.15	0.88	0.55	0.22	2.49	0.01
Exp: 1b vs. 1a * Pref: mixed vs. ego * Offset: 180° vs. 90° * Instruct: lateral vs. sagittal	—	—	—	—	0.04	0.17	0.26	0.80	-0.00	0.12	-0.02	0.98	0.35	0.57	0.62	0.54
Exp: 1b vs. 1a * Pref: other vs. ego * Offset: 180° vs. 90° * Instruct: lateral vs. sagittal	—	—	—	—	-0.19	0.13	-1.46	0.14	-0.06	0.10	-0.65	0.51	-0.99	0.44	-2.24	0.03

#### 4.1.1. *Front-back distinction in perspective choice*

As described in the previous section, somewhat surprisingly, the type of instruction (i.e., whether the simulated partner gave a lateral “right/left” instruction vs. sagittal “front/back” instruction) interacted with the orientation of the folders, influencing perspective choice. Although we had a priori predictions that perspective choice would differ across experiments (with increased egocentrism for ego-aligned folders in Experiment 1a), we did not expect that the lateral/sagittal content of the instruction would be associated with different degrees of egocentrism depending on the orientation of the folders.

To unpack this observed interaction, we considered performance on control trials (see Appendix A), which provided insight into the mapping of lateral and sagittal terms in the two experiments. Folder selections on control trials suggested that the difference in egocentrism between lateral and sagittal critical trials across experiments was partly due to a small number of individuals who assigned a different mapping on the terms *front-back* than our assumed mapping. According to our coding, the folder “at the front” was assigned to be the folder closer to the person whose perspective was being adopted, and the folder “at the back” was assigned to be the folder that was farther. However, some participants used the reverse mapping. This alternative mapping is indicated by the fact that on *front-back* control trials in Experiment 1b, there were on average 10% “errors” ( $SD = 30\%$ ), most of which committed by individuals who were classified as mixed responders based on their performance on critical trials. As shown in Table 2, mixed responders from Experiment 1b selected the “incorrect” folder on nearly half the sagittal



control trials (46%). Mixed responders from Experiment 1a also selected the “incorrect” folder somewhat frequently on sagittal trials, but to a lesser degree (21%).

The fact that these high “errors” were limited to sagittal trials suggests that the individuals making these selections were not simply responding haphazardly to instructions, but were instead using systematically a different mapping of the terms *front-back*. For them, the term “front” was assigned to the folder farther from the person whose perspective was being adopted and the term “back” was assigned to the folder that was closer. On control trials, this resulted in the selection of our assumed “incorrect” folder. In Experiment 1b, 7 out of the 18 mixed responders exhibited consistently the reversed mapping of the *front-back* terms, on 70%-100% of the sagittal control trials. In Experiment 1a, 3 out of the 17 mixed responders used this mapping consistently.

Assuming that these mixed responders used the same mapping of sagittal terms on critical trials as well, this would suggest that they were classified as “mixed” not because they had actually switched egocentric and other-centric perspective strategies (or were responding randomly), but rather because their folder choices on critical sagittal trials were being coded as adhering to a different perspective than their responses on lateral trials, even though—by their coherent conceptualization—they were in fact adopting the same perspective on all critical trials. Since in Experiment 1b there were more mixed responders exhibiting this consistent mapping of the sagittal terms than in Experiment 1a (as ratified by their performance on control trials), this could have led to the interaction between folder orientation (i.e., Experiment) and type of instruction observed on critical trials.

Given this unexpected—but not unreasonable—mapping of the sagittal terms for a subset of participants, in Appendix B, we report a linear mixed effects model on egocentric choice selection that excludes mixed responders. In that model, whereas folder orientation (i.e., Experiment) and type of instruction continue to be predictive of perspective choice on their own, their interaction is no longer predictive of perspective choice. That is, listeners were still more likely to make an egocentric choice on lateral than sagittal trials, but this no longer depended on the folders' orientation.

As an additional exploratory analysis, given the alternative mapping of the terms *front-back* reflected in the performance of mixed responders on control trials, we reclassified egocentric, other-centric, and mixed responders based on their responses to critical trials with lateral instructions (*left-right*) only. The shift from egocentric to other-centric preference across the two experiments was even starker with this classification: in Experiment 1a, with ego-aligned folders, there were 42 egocentric, 29 other-centric, and 22 mixed responders, whereas in Experiment 1b, with other-aligned folders, there were 25 egocentric, 61 other-centric, and 9 mixed responders. This difference in the distributions was significant,  $\chi^2(2) = 21.12, p < .001$ . In Appendix B, we explore why the number of mixed responders decreases in Experiment 1b while it increases in Experiment 1a, and examine qualitatively the behavior of “true” mixed responders (i.e., those who were made 30-70% egocentric choices on both lateral and sagittal trials).

For the remaining analyses on the dynamics of the listeners' responses in the main text, we continue to use the original classification of responders, based on their object selections for all trials (i.e., including both lateral and sagittal instructions). This is for consistency with Duran et al.'s (2011) classification, and because the individuals using

the reversed mapping of sagittal terms are contained in the mixed responders' category in these two experiments. However, in Appendix B, we report supplementary exploratory analyses of models for all measures that exclude mixed responders.

*Table 2. Proportion of mean errors (and standard deviations) on control trials across the two experiments (1a and 1b), perspective preference (egocentric, mixed, other-centric) and verbal instruction (lateral: left-right, and sagittal: front-back).*

Experiment	Instruction	Perspective preference			
		Egocentric	Other-centric	Mixed	Average
1a	lateral	.03 (.16)	.01 (.11)	.05 (.21)	.02 (.15)
1a	sagittal	.04 (.21)	.03 (.17)	.21 (.41)	.07 (.25)
	Total	.04 (.19)	.02 (.14)	.13 (.34)	.05 (.21)
1b	lateral	.01 (.11)	.03 (.16)	.05 (.22)	.03 (.16)
1b	sagittal	.02 (.15)	.02 (.14)	.46 (.40)	.10 (.30)
	Total	.02 (.13)	.02 (.15)	.25 (.43)	.06 (.25)
Total		.03 (.17)	.02 (.15)	.19 (.39)	.06 (.23)

*Table 3. Means (and standard deviations) for response times (RT) in ms, total distance (for Experiments 1a and 1b) or divergence from optimal distance (for Experiment 2) in pixels (Distance), and the number of shifts in the direction of the mouse-trajectories (Directional Shifts) for egocentric, other-centric, and mixed responders across different conditions of alignment for the configural cue on critical trials (ego-aligned vs. other-aligned objects across Experiments 1a and 1b, and ego-aligned, other-aligned, neither-aligned, and both-aligned configurations in Experiment 2) for critical and control trials.*

Experiment	Dependent measure	Configural cue	Critical trials			Control trials		
			Perspective Preference			Perspective Preference		
			Egocentric	Other-centric	Mixed	Egocentric	Other-centric	Mixed
1a	RT	Ego-aligned	1286 (765)	1642 (904)	1439 (908)	1243 (741)	1456 (756)	1273 (670)
1b		Other-aligned	1231 (750)	1636 (907)	1666 (1181)	1096 (555)	1402 (669)	1467 (1021)
1a	Distance	Ego-aligned	288 (106)	327 (139)	293 (95)	277 (87)	294 (104)	284 (93)
1b		Other-aligned	270 (76)	305 (108)	312 (115)	265 (78)	286 (95)	295 (103)
1a	Directional Shifts	Ego-aligned	1.69 (1.50)	2.08 (1.50)	1.75 (1.33)	1.54 (1.30)	1.71 (1.28)	1.67 (1.28)
1b		Other-aligned	1.60 (1.34)	2.04 (1.46)	2.37 (2.17)	1.58 (1.22)	1.746 (1.29)	2.06 (1.90)
2	RT	Ego-aligned	1299 (611)	1630 (814)	1413 (844)	1371 (811)	1910 (1132)	1444 (867)
2		Other-aligned	1365 (801)	1574 (1011)	1491 (770)	1506 (792)	1786 (913)	1710 (1046)
2		Neither-aligned	1333 (809)	1795 (1076)	1584 (945)	1331 (687)	1599 (807)	1451 (887)
2		Both-aligned	—	—	—	1573 (942)	1871 (958)	1596 (935)
2	Distance	Ego-aligned	77 (80)	110 (109)	91 (82)	75 (89)	98 (106)	88 (76)
2		Other-aligned	85 (84)	111 (107)	112 (102)	99 (111)	136 (136)	150 (146)
2		Neither-aligned	85 (93)	135 (129)	117 (114)	104 (140)	123 (127)	122 (112)
2		Both-aligned	—	—	—	79 (84)	96 (105)	92 (103)
2	Directional Shifts	Ego-aligned	1.76 (1.21)	2.23 (1.22)	1.91 (1.49)	1.50 (1.28)	1.99 (1.61)	1.65 (1.23)
2		Other-aligned	1.90 (1.35)	2.13 (1.40)	2.03 (1.44)	1.79 (1.13)	2.03 (1.25)	2.04 (1.69)
2		Neither-aligned	1.64 (1.34)	2.41 (1.51)	2.08 (1.61)	1.64 (1.23)	2.04 (1.52)	1.83 (1.42)
2		Both-aligned	—	—	—	1.84 (1.56)	2.31 (1.68)	2.08 (1.41)

## 4.2. Response times

Table 1 includes a summary of the omnibus statistical model built to examine the relationship between folder orientation (Experiment), perspective preference, speaker position, and instruction type, separating out the results for each of the planned contrasts and their interactions. The overall variance captured by the omnibus model (with both fixed and random effects) was 55%.

Before interpreting the statistical significance of the contrasts of interest in Table 1, we examined the overall two-way, three-way, and four-way interactions of the four factors. Toward that end, likelihood ratio tests were performed between the omnibus model and models that excluded each of the critical two-, three-, or four-way interactions. The two-way interactions of perspective preference with instruction type ( $\chi^2(2) = 85.67, p < .001$ ) and with speaker position ( $\chi^2(2) = 18.95, p < .001$ ) were significant. Of the remaining interactions, only the overall three-way interaction between perspective preference, folder orientation, and instruction type was significant ( $\chi^2(2) = 7.76, p = .02$ ). We qualify these in the context of the results in Table 1, below.

As shown in Table 1, the orientation of the folders was not predictive of the listeners' response times overall: the latencies of Experiments 1a than 1b did not differ significantly. The listeners' perspective preference, on the other hand, was predictive of response times, in that egocentric responders were faster than other-centric responders. However, egocentric responders were not significantly faster than mixed responders (see Table 1).

Despite our expectation that other-centric responders would be faster when the folders were aligned with the speaker's perspective (in Experiment 1b), the contrast for the folder orientation did not predict the response times of other-centric relative to egocentric responders (see Table 1). Indeed, the overall two-way interaction between folder orientation (Experiment) and perspective preference was not significant,  $\chi^2(2) = 1.25, p = .54$ . As reflected in the mean response times reported in Table 3, other-centric responders were comparably fast to select a folder from the speaker's perspective, regardless of whether the folders were aligned with the speaker (Experiment 1b) or with their own perspective (Experiment 1a).

The speaker's position significantly predicted response times, with responders being slower when the speaker was at 180° than at 90°. This difference seem to hold for other-centric responders (relative to egocentric ones), but not for mixed responders (see Table 3). Whereas egocentric responders were similarly fast at the two offsets (at 180°:  $M = 1282, SD = 767$  ms; at 90°:  $M = 1261, SD = 753$  ms), other-centric responders exhibited a decrement over 200 ms at 180° (180°:  $M = 1733, SD = 938$  ms; 90°: 1501,  $SD = 835$  ms). As we have noted earlier, the overall two-way interaction between perspective preference and speaker position was significant.

A similar pattern was observed for the type of spatial instruction (see Table 1), as listeners were overall slower on lateral ( $M = 1636$  ms,  $SD = 1001$  ms) than sagittal ( $M = 1384, SD = 771$  ms) instructions. This difference between lateral and sagittal instructions predicted the response times of other-centric relative (relative to egocentric responders), but not of mixed responders. Again, this further qualifies the significant two-way interaction between perspective-preference and instruction type reported earlier.

To decompose the significant three-way interaction between folder orientation, perspective preference, and instruction type, we conducted a follow-up simple effects analysis by holding each level of the instruction type factor (i.e., lateral or sagittal) constant. This analysis revealed that the interaction between perspective preference and folder orientation was robust when all trials were considered together, as it was not significant for models with lateral trials ( $p = .43$ ) or sagittal trials ( $p = .30$ ) on their own—that is, it was not the case that it held uniquely for one type of instruction.

### 4.3. Total distance

The overall variance captured by the omnibus model reported in Table 1 was 31%. Again, we examined the overall two-way, three-way, and four-way interactions of the four factors (folder orientation, perspective preference, speaker position, and type of instruction). The overall two-way interaction of perspective preference with type of instruction and with speaker position were significant ( $\chi^2(2) = 20.59, p < .001$ , and  $\chi^2(2) = 18.66, p < .001$ , respectively). The interaction between speaker position and type of instruction was significant as well,  $\chi^2(1) = 4.63, p = .03$ . None of the remaining overall interactions were significant.

As shown in Table 1, consistent with response times, the folders' orientation was not predictive of the total distance of the listeners' mouse trajectories. Perspective preference, on the other hand, mattered, as other-centric responders made longer trajectories relative to egocentric responders. As with response times, egocentric and mixed responders did not differ reliably.

Again, despite our predictions, other-centric responders did not make significantly shorter trajectories when the folders were other-aligned (Experiment 1b) than ego-aligned

(Experiment 1a): the interaction between other-centric (vs. egocentric) preference and experiment identity did not predict the trajectories' total distance. Interestingly, mixed responders made trajectories that were longer in Experiment 1b than in 1a, compared to egocentric responders. This pattern can be observed in the mean total distances reported in Table 3.

As shown in Table 1, speaker's position (180° vs. 90° offset) also predicted the total distance of the mouse trajectories, although the type of instruction (lateral vs. sagittal) did not do so significantly. However, as with response times, these factors impacted other-centric responders: compared to egocentric responders, other-centric responders made longer trajectories when the speaker was at a 180° vs. a 90° offset, and when the speaker provided a lateral vs. sagittal instruction. This qualifies the overall two-way interactions reported above between perspective preference with speaker position and with instruction type. Moreover, as suggested by the overall two-way interaction of the speaker position and type of instruction, and evidenced by the significant interaction of the relevant contrasts in the omnibus model in Table 1, the difference between lateral and sagittal instructions was greater at the 180° than the 90° (21 vs. 6 pixels, on average).

#### **4.4. Directional shifts**

A similar pattern of performance was obtained for directional shifts. The overall variance captured by the omnibus model reported in Table 3 was 35%. As with response times and total distance, the overall two-way interaction between perspective preference with type of instruction ( $\chi^2(2) = 18.38, p < .001$ ) and with speaker position ( $\chi^2(2) = 10.83, p < .01$ ) were significant. And as with total distance, the interaction between speaker



position and type of instruction was significant as well,  $\chi^2(1) = 4.18, p = .04$ . None of the remaining overall interactions were significant.

As shown in Table 1, consistent with the previous measures, the folders' orientation was not predictive of the total distance of the listeners' mouse trajectories. In contrast to the previous mouse-tracking measures, other-centric responders did not make significantly less efficient trajectories (i.e., more directional shifts) than egocentric responders. Note, however, that when mixed responders were excluded, as we report in the analyses in Appendix B, the other-centric responders' decrement in performance (relative to egocentric responders) is observed in all measures, including directional shifts.

Again, there was no evidence that other-centric responders made more efficient trajectories when folders were other-aligned than ego-aligned (see Table 1): the interaction between other-centric (vs. egocentric) preference and experiment identity was not significant. As shown in Table 1, relative to egocentric responders, mixed responders made significantly more directional shifts when folders were other-aligned (in Experiment 1b) than when they were ego-aligned (in Experiment 1a). Nevertheless, the overall two-way interaction between perspective preference and folder orientation was not significant,  $\chi^2(2) = 5.46, p = .07$ .

Speaker position did not predict the number of directional shifts, and neither did the type of instruction. However, consistent with the previous measures, both factors predicted the number of directional shifts made by other-centric responders, who made more directional shifts relative to egocentric responders when the speaker was maximally misaligned or when responding to lateral instructions. This is indicated by the significant

interactions of the relevant contrasts in Table 1, which contextualize the overall three-way and two-way interactions reported above

#### 4.5. Dynamics across trials

In a final exploration, we considered whether listeners in the two experiments differed in terms of how readily they stabilized on their response strategy over time, by examining mixed effects models that included the order of the trial and its interaction with the other factors as predictors.

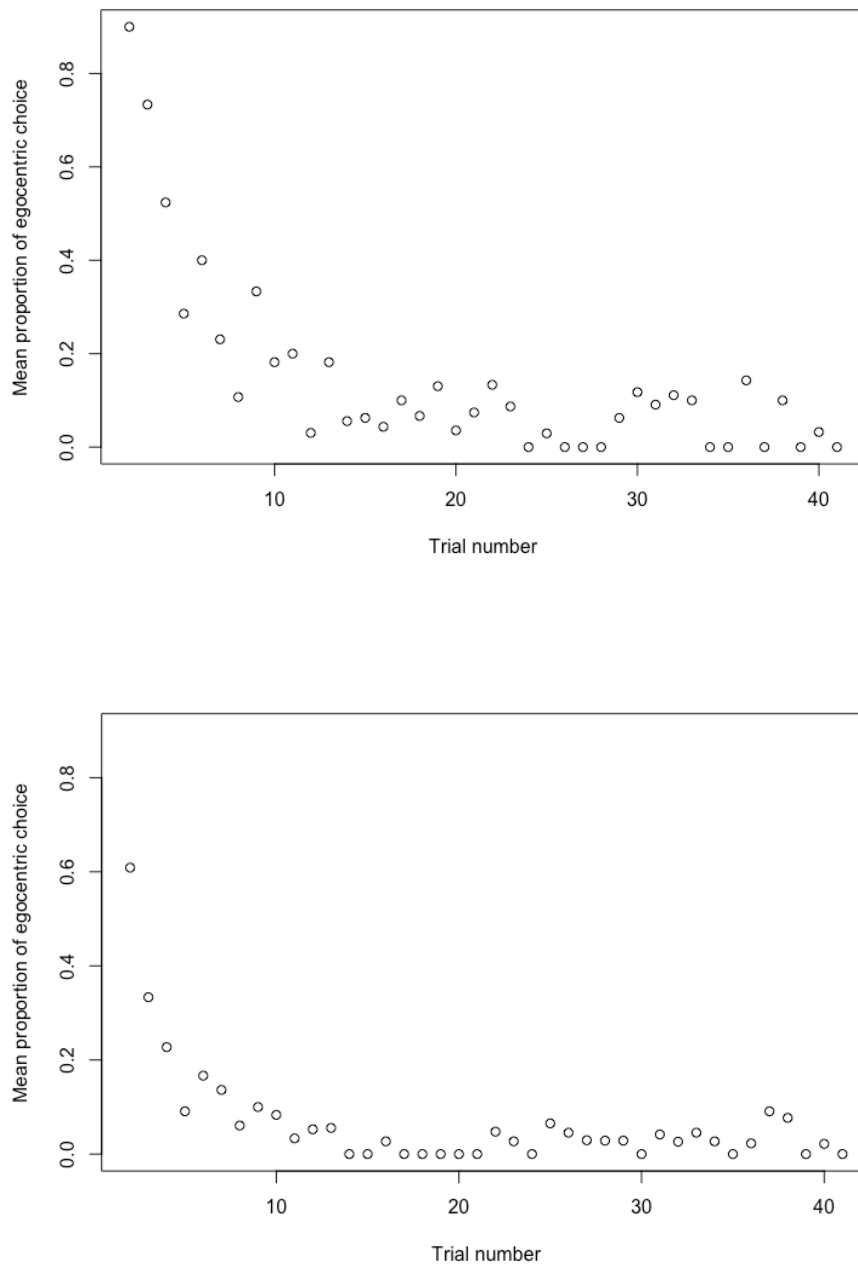
Trial order predicted the listeners' response times ( $\beta = -.06$ ,  $SE = .009$ ,  $t = -6.03$ ,  $p < .001$ ), but not the length of their mouse trajectories ( $\beta = .002$ ,  $SE = .007$ ,  $t = .28$ ,  $p = .78$ ) or the number of directional shifts they made ( $\beta = .05$ ,  $SE = .03$ ,  $t = 1.84$ ,  $p = .07$ ).

For response times, the impact of trial order differed according to the orientation of the configural cue, as the interaction between trial order and folder orientation (Experiment) was significant ( $\beta = -.07$ ,  $SE = .02$ ,  $t = -3.43$ ,  $p < .001$ ). This was not the case for the other two measures (total distance:  $\beta = -.02$ ,  $SE = .01$ ,  $t = -1.43$ ,  $p = .15$ ; directional shifts:  $\beta = -.02$ ,  $SE = .06$ ,  $t = -.04$ ,  $p = .70$ ).

Thus, relative to Experiment 1a, where folders were ego-aligned, in Experiment 1b, where folders were other-aligned, the listeners were faster to respond over time. This suggests that the other-aligned configural cue contributed to the stabilization of responses, which were predominately other-centric in Experiment 1b. We do not report the results of the model on the proportion of egocentric responses, as that mixed logistic regression model failed to converge, even when simplified with a backwards fitting approach. Nevertheless, as illustrated in Figures 3a and 3b, other-centric responders in

Experiment 1b stabilized on the other-centric perspective (i.e., approached 0% egocentric choice) with a sharper drop than other-centric responders in Experiment 1a.

*Figure 3. Mean proportion of egocentric responses across trial order for other-centric responders in Experiment 1a (top panel) and 2b (bottom panel).*



#### 4.6. Summary

The results of Experiments 1a and 1b suggest that a configural cue—such as the orientation of the configuration’s constituent objects—can have a significant impact on language users’ perspective strategy. This was indicated by the shift in the distribution of responders across the two experiments: in Experiment 1b, where objects were always aligned with the perspective of the partner, there were more responders who consistently adopted that perspective (other-centric responders) than in Experiment 1a, where objects were always aligned with the egocentric perspective. Moreover, in Experiment 1b, other-centric responders stabilized on that perspective choice more readily, as indicated by their performance across trials.

The impact of the configural cue is also evident in the fact that, for some participants, the cue’s convergence with the partner’s perspective contributed to an alternative interpretation of the sagittal terms *front* and *back*. As we observed on control trials, when the configural cue was aligned with the speaker (Experiment 1b) nearly half the responses of participants who were classified as “mixed responders” selected a folder based on the reverse mapping of the terms *front-back* compared to the rest of the responders, with many of these responders adhering to this mapping consistently (70-100% of the time). This may have occurred because the alignment of the folders, by continually varying across trials (while remaining aligned with the speaker), made more obvious the possibility that “front” could be interpreted as the “top of the table” as perceived from the heading of the person whose perspective was being adopted (i.e., farther away from that person). Some mixed responders from Experiment 1a also appeared to use that mapping, but they did so to a lesser degree.

Even though the configural cue had an impact on perspective strategy and on the interpretation of sagittal terms for a subset of the listeners, it did not generally influence the dynamics of listeners' mouse movements on critical trials, not even for other-centric responders. Despite our predictions, responding from the partner's (the speaker's) perspective was not facilitated by the alignment of that perspective with the configural cue: there was no evidence that other-centric responders in Experiment 1b made faster and more direct mouse trajectories compared to Experiment 1a.

Instead, the temporal and trajectory dynamics of the mouse movements were predicted by the listeners' perspective preference. Overall, other-centric responders made slower and longer trajectories than egocentric responders. Other-centric responders were also influenced by the speaker's position and by the type of verbal instruction more than egocentric responders, being slower when the speaker was depicted at 180° than at 90°, and on *left-right* than *front-back* instructions. This makes sense insofar as responding from another person's perspective requires taking into account that person's position and how relative spatial terms are mapped onto space, given that position.

Consistent with previous findings, other-centric responding incurs a greater associated cost when the partner is maximally misaligned (by 180°) vs. misaligned by an orthogonal 90° offset (e.g., Duran, Dale, & Kreuz, 2011; Mainwaring et al, 2003) or when the partner provides a lateral (vs. sagittal) instruction (e.g., Franklin & Tversky, 1990). Nevertheless, the influence of speaker position and spatial instruction did not depend on the orientation of the configuration's objects.

The lack of facilitation from the configural cue when other-centric responders responded to individual instructions could be because, in both Experiment 1a and 1b, the

speaker's position continually varied across trials, making the other-centric perspective strategy relatively taxing in both experiments. This possibility is consistent with findings that following spatial instructions from a new heading on each trial places demands on cognitive resources (Avraamides & Carlson, 2003; Ryskin et al., 2014; Ryskin, Wang, & Brown-Schmidt, 2016).

Although the dynamics of participants' responses do not reveal any facilitation for Experiment 1b, we do have some indirect evidence that, in some ways, folder orientation did impact the processing of other-centric responders. As we report in Appendix B, where we reclassify responders based on lateral trials only, a greater number of other-centric responders was reclassified as mixed responders when only lateral terms were considered in Experiment 1a ( $N=14$ ) than in 1b ( $N=1$ ). This may be taken to suggest that these otherwise other-centric responders were more likely to make errors (i.e., by making an egocentric choice) when interpreting *left-right* terms when the configural cue did not support the other-centric perspective, in Experiment 1a.

In Appendix C, we report the results of a follow-up experiment in which we investigate whether individuals perceive the alignment of the folders used here consistently with our classification of folders as ego-aligned and other-aligned. This undertaking was motivated by the fact that in Experiments 1a and 1b, the elongated axis of the folder reflected its width (i.e., it was aligned with the folder's left/right dimension), whereas the elongated axis of objects often serves as a cue for assigning the objects' "upright", top/bottom orientation (e.g., Quinlan & Humphreys, 1993; Sekuler & Swimmer, 2000). The results of that follow-up experiment suggest that the folders used here did offer a directional cue that was consistent with our classification of their

alignment relative to task partners, with participants classifying configurations as ego-aligned and other-aligned reliably above chance. We acknowledge that the choice of different objects in Experiments 1a and 1b could have provided clearer directional cues (e.g., by choosing objects whose elongated axis was aligned with its top/bottom orientation), and such cues could have potentially facilitated further other-centric responding when other-aligned. Still, the objects used here provided sufficiently a clear configural cue that was apprehended, in line with our intentions, as reinforcing a particular perspective.

In sum, the reorientation of a configural cue introduced a shift in response strategy across experiments, without influencing the dynamics of responding within trials when there was a conflict in the candidate perspectives. However, as we saw in our analyses that included trial order as a predictor, the configural cue did influence the time-evolution of perspective-taking during the course of the entire experiment: in Experiment 1b, with other-aligned folders, responders stabilized more quickly on their strategy, as their response times shortened more steeply over time. This was likely driven by other-centric responders, who in Experiment 1b stabilized more quickly and less variably on their preferred other-centric choice as illustrated in Figures 3a and 3b. Also, as we observed in the patterns of reclassification of responders on lateral trials only, other-centric responders may have experienced less difficulty interpreting lateral terms in Experiment 1b than in 1a.

## **5. Method: Experiment 2**

In Experiment 2, we take a different approach at examining the potential impact of a configural cue to perspective selection: we manipulate the alignment of a configural

cue—the intrinsic orientation of the configuration—*within* participants, such that it coincides with the egocentric perspective on some trials and with the other-centric perspective on others. This contrasts with Experiments 1a and 1b, where the alignment of the configural cue with a particular perspective was constant across trials (in 1a always ego-aligned, and in 1b always other-aligned). Under those consistent pairings, we did not find evidence of the configural cue influencing the dynamics of the listeners' responses on individual trials.

It is possible that a configural cue's *variable alignment* could influence the dynamics of responses, and that this influence may depend on the listener's perspective preference (egocentric vs. other-centric). In Experiment 2, we pursue this question by varying the alignment of a configural cue—the orientation of a 3-object configuration—across trials for all listeners. Here, all listeners were presented with displays in which, not only the speaker's position varied across trials, but the alignment of the configuration did as well. We examine the possibility that this increased variability in the alignment of task features makes listeners more susceptible to the influence of the configural cue during their responses on individual trials.

We created 3-object configurations that were ego-aligned and other-aligned, but we also included configurations that were aligned with neither the egocentric or other-centric perspective (neither-aligned). This methodological decision was motivated by a study by Galati and Avraamides (2013, 2015), which investigated perspective selection in memory and language use. In that study, pairs of Direction Givers (DGs) and Direction Followers (DFs) reconstructed a 7-object spatial configuration, whose axis of symmetry was ego-aligned (relative to the DG), other-aligned, or neither-aligned. The findings



suggested that reasoning from another's perspective was particularly challenging when the spatial configuration reinforced neither the egocentric or other-centric perspective. The neither-aligned condition involved longer interactions (in terms of conversational turns) compared to the other two conditions, and more egocentric (than other-centric) expressions in the DG's descriptions when the DG did not know the DF's perspective at the time of memorizing the configuration.

Although suggestive of a cognitive cost associated with neither-aligned cues, that work does not provide direct evidence of such a cost, insofar as it focuses on speakers' aggregate linguistic behavior (their number of turns and the distribution of their spatial expressions). In Experiment 2, through the use of mouse-tracking, we can examine not only whether other-centric responding is more difficult than egocentric responding, but also whether responding from a particular perspective varies in difficulty depending on the alignment of the configural cue.

We anticipate that, despite the varying configural cue, listeners will still stabilize on a consistent perspective strategy during the task (whether egocentric or other-centric). Still, to the extent that a configural cue can facilitate responding when it converges with the perspective being adopted, we predict that listeners may make faster and more direct mouse trajectories on trials on which the orientation of the configuration supports the listener's strategy. Other-centric responders would therefore be expected to exhibit facilitation on other-aligned configurations relative to the other types of configurations. A second, more exploratory prediction about other-centric responders, based on the findings of Galati and Avraamides (2013; 2015), would be that they may experience more difficulty making judgments on neither-aligned configurations than on the other two

types. Finally, egocentric responders may exhibit facilitation on ego-aligned configurations, or alternatively may ignore the configuration's alignment altogether, since both the configuration and the speaker's position are irrelevant to their perspective strategy: for egocentric responders in this task, the linguistic spatial terms map onto stable regions of space based on their fixed perspective.

### **5.1. Participants**

We collected data from 201 sessions on Amazon Mechanical Turk (66 female, 108 male, 27 not reported). As with Experiments 1a and 1b, the task was available to participants who were over 18 years old and located in the U.S., and participants were paid \$1.50 for the responses. The mean age of users was 35.51 years ( $SD=10.54$ ; range: 19–68). We excluded the data of 12 sessions that appeared to be duplicates based on the worker's ID or their IP address, including only the first session in the data analyses. Another 5 participants submitted incomplete data, resulting in the data of 184 users being submitted to data analyses.

### **5.2. General setup**

In Experiment 2, we manipulated the intrinsic orientation of the configuration, such that the triangular shape formed by the constituent objects varied across trials. This is in contrast to Experiments 1a and 1b, where the alignment of the configural cue remained constant across trials. Here, configurations could be aligned with the listener's perspective (ego-aligned), with the speaker's perspective (other-aligned), or with neither perspective (neither-aligned). The orientation of the folders remained fixed throughout the experiment, with folders always being aligned with the egocentric perspective.

The instructions and setup were otherwise identical to Experiments 1a and 1b.

### 5.3. Stimuli

Similar to Experiments 1a and 1b, each trial displayed the top-down view of a circular table with a configuration of three objects: two folders and a CD. A CD was chosen as the third object, as it permitted forming triangular configurations without introducing additional (and potentially competing) directional features. The participant's (i.e., the listener's) perspective, at  $0^\circ$ , was indicated by the word "You", whereas the speaker's perspective, which changed across trials ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ ), was indicated by the word "Partner" and a rectangular box in which participants would deposit the folder requested by the speaker (see Figures 4a-c).

Of the 48 trials, 28 trials were constructed such that interpretation from the listener's and the speaker's perspective would result in different object choices (critical trials). This was done in a manner similar to Experiments 1a and 1b, where the "Partner" label position induces an ambiguity in how lateral or sagittal terms are interpreted. In 8 of the 28 critical trials, the speaker was at  $180^\circ$ , while in the remaining 20 the speaker was either at  $90^\circ$  or  $270^\circ$ .

With respect to the configuration's alignment, there were three configuration types among critical trials: ego-aligned, other-aligned, or neither-aligned configurations. For person-aligned trials, the axis of symmetry of an isosceles triangle formed by the three objects was aligned with either the participant (ego-aligned) or with the speaker's perspective (other-aligned) (see Figures 4a and 4b). The top vertex of the triangle centered over the two bottom vertices, along the base, is generally regarded as a canonical position for triangles (e.g., Lupyan, 2017) and produces an alignment for the relevant label along the triangle's base. For neither-aligned trials, the configuration

formed an isosceles right-angle triangle, whose axis of symmetry was not aligned with either perspective (see Figure 4c). There were 8 ego-aligned, 8 other-aligned, and 12 neither-aligned trials.

Another 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). Of the 20 control trials, 8 involved cases for which the speaker's perspective was the same as the listener's (i.e.,  $0^\circ$ ), and another 12 involved cases for which the speaker was at  $90^\circ$  or  $270^\circ$ .

In terms of the configuration's alignment, those 8 trials with the speaker at  $0^\circ$  were always both-aligned. Among the remaining 12 control trials, 4 involved configurations with the axis of symmetry aligned with the listener (2 ego-aligned, 2 ego-reversed, with the apex of the isosceles triangle pointing toward  $0^\circ$ ), 4 involved configurations with the axis aligned with the speaker (2 other-aligned, 2 other-reversed), and 4 involved neither-aligned configurations.

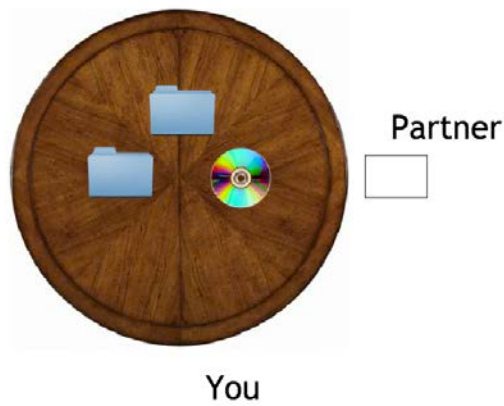
The same pre-recorded instructions as Experiments 1a and 1b were used. Since there were pre-recorded instructions on 10 trials for each spatial term (i.e., "Give me the folder on the ... [*right / left / front / back*]"), two tokens from each instruction type repeated with the constraint that these were never presented consecutively.

Trials were presented in one of 10 pseudo-randomized lists, to which participants were randomly assigned. The first 8 trials were always control trials. This was done because the visual task features of this experiment were more variable than in Experiments 1a and 1b—both speaker position *and* configuration orientation varied across trials here—and we wanted to allow participants to settle on their perspective

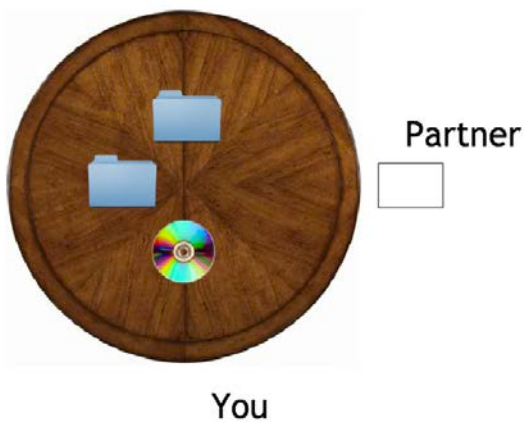
preference before facing ambiguity in the instructions. These initial control trials included 4 trials with the speaker at 0° (one from each of the front-back-left-right instructions) and 4 with the speaker at 90° / 270° (one from each of the front-back-left-right instructions); their order was pseudo-randomized across the 10 lists. The remaining 40 trials appeared in a pseudo-randomized order, with the constraints of avoiding consecutive trials with the same alignment type, the same instruction (*front, back, left, right*), or the same speaker position. Additionally, there were never more than two consecutive control trials or more than 4 consecutive ambiguous trials. With these constraints we created 5 pseudo-randomized orders for the 40 trials, which were then reversed to yield the final 10 lists.

Figure 4. Examples of trials from Experiment 2. The configuration in Figure 4a illustrates an ego-aligned trial, where the intrinsic axis of the triangular configuration is aligned with the listener's perspective, which is indicated by the word "You" at 0°. The speaker is depicted by the words "Partner" at 90°. This configuration constitutes a critical trial when accompanied by the instruction "Give me the folder that's in the front". It illustrates a control trial when accompanied by the instruction "Give me the folder on the right". Figure 4b illustrates an other-aligned configuration, with the triangular configuration aligned with the speaker's perspective at 90°. Figure 4c illustrates a neither-aligned configuration, with the speaker at 90°; the axis of symmetry of the right angle triangle formed by the objects is not aligned with either partner.

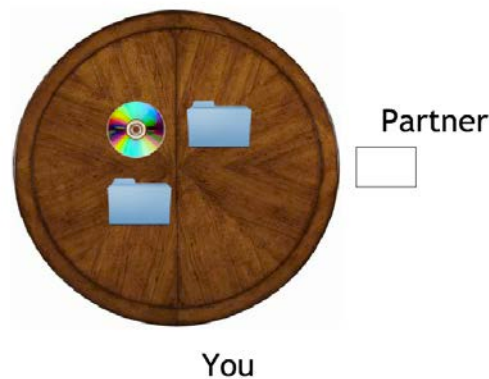
4a.



4b.



4c.



#### 5.4. Behavioral tracking using computer-mouse trajectories

As with Experiments 1a and 1b, we examined response times and directional shifts in motion. For Experiment 2, we carried out an artifact correction on the measure of distance. Instead of total distance in pixels we computed the divergence from the optimal distance to the selected folder. We took the optimal distance to be the distance in pixels from the trial initiation to their object selection: this straight line reflects an optimal distance because it is the minimal possible distance needed to reach that point. We then subtracted observed distance from this optimal distance. This was done because, in the three types of configurations (ego-aligned, other-aligned, and neither-aligned), the average distance from the start point of the trial to the selected folder differed<sup>5</sup>.

<sup>5</sup> On average, the optimal distance to folders in ego-aligned configurations was greater ( $M = 240$ ,  $SD = 28$  pixels) than in other-aligned ( $M = 209$ ,  $SD = 43$  pixels;  $\beta = -29.37$ ,  $SE = 1.69$ ,  $t = -17.36$ ,  $p < .001$ ) and neither-aligned configurations ( $M = 212$ ,  $SD = 42$  pixels;  $\beta = -22.46$ ,  $SE = 1.56$ ,  $t = -14.43$ ,  $p < .001$ ). In particular, the difference between ego-aligned and neither-aligned configurations was greater for the egocentric folder (a 40 pixel difference) than for the other-centric folder (a 23 pixel difference), as indicated by an interaction of configuration type and folder selection ( $\beta = -17.72$ ,  $SE = 2.96$ ,  $t = -5.98$ ,  $p < .001$ ).

## 6. Statistical Analysis: Experiment 2

As with Experiments 1a and 1b, we constructed separate linear mixed effects models for critical and control trials for each dependent measure. The results of control trials are reported in Appendix A.

In contrast to Experiments 1a and 1b, we did not include the offset from the speaker's position as a fixed effect for critical trials because ego-aligned and other-aligned configurations were not possible with the speaker at 180°: when the speaker was at 180°, the configuration's axis of symmetry would be aligned with both the speaker and the listener. (Critical trials with the speaker at 180° were therefore all neither-aligned.) In our materials, ego-aligned and other-aligned configurations occurred only when the speaker was at a 90° offset (in either direction) for critical trials. Similarly, for control trials, when the speaker was at 0° the configuration could only be both-aligned and could not occur in any of the other alignment conditions (ego-aligned, other-aligned, or neither-aligned); again, the speaker's position was not included in those models as a fixed factor.

For the proportion of egocentric choices, we built an omnibus statistical model that examined the interaction between Configuration type (ego-, other-, and neither-aligned), and Instruction type (sagittal = -0.5 vs. lateral = 0.5). Configuration alignment was coded in terms of planned contrasts that compared responses on ego-aligned vs. other-aligned configurations (ego-aligned = -0.5, other-aligned = 0.5) and of ego-aligned vs. neither-aligned configurations (ego-aligned = -0.5, neither-aligned = 0.5).

For response time, divergence from optimal distance, and directional shifts, our omnibus statistical models also included Perspective preference as a fixed effect (egocentric, other-centric, mixed), in order to examine the interaction between



Perspective preference, Configuration alignment, and Instruction type. As with Experiment 1, Perspective preference was coded in terms of planned contrasts that compared the ease of responding of egocentric relative to other-centric responders (egocentric = -0.5, other-centric = 0.5), and of egocentric relative to mixed responders (egocentric = - 0.5, mixed = 0.5).

We started with the full random effect structure, which included random slopes for Configuration type, Instruction type, and their interaction. When the model did not converge, we simplified the models with the same procedure as before.

Again, we log-transformed dependent variables when visual inspection of residual plots for the models revealed deviation from normality and homoscedasticity; this was the case for response time and divergence from optimal distance. Based on visual inspection of the initial data, we excluded from the trials on which participants took more than 6000 ms to respond or made trajectories longer than 1000 pixels, using parallel outlier procedures as in Experiments 1a and 1b. This resulted in 2.86% of the data being excluded (2.56% of critical trials and 3.28% of control trials). For binary dependent variables, we used mixed logistic regression models with binomial error structure.

We also adopted a parallel approach as in Experiments 1a and 1b to examine changes in response behavior over time, conducting an additional exploratory analysis that included trial order and its interaction with the other factors (perspective preference, configuration type, and instruction type) as fixed effects. Additional exploratory analyses that exclude mixed responders are reported in Appendix B.

## 7. Results: Experiment 2

### 7.1. Perspective choice

Before getting to the listeners' egocentric and other-centric choices on critical trials, we have to note that our initial analyses of control trials revealed listeners in Experiment 2 used consistently the alternative mapping of the terms *front* and *back*. When using our original assignment of the terms *front-back*—whereby the term *front* referred to the folder closer to the person whose perspective was being adopted—error rates on sagittal control trials were extremely high: 89% for *front* and 95% for *back* (compared to 4% and 5% for *left* and *right* instructions). The fact that high error rates were limited to sagittal control trials suggested that listeners in Experiment 2 used a reverse mapping of the sagittal terms than that predominately used by listeners in Experiments 1a and 1b, as well as in Duran, Dale, and Kreuz (2011).

We therefore recoded the assignment of these sagittal terms, both for determining the “correct” choice on control trials (see Appendix A) and for determining the egocentric/other-centric choice on critical trials, reported below. Through this recoding, the folder at the *front* was now assigned to be the folder farther from the listener or speaker (to determine the egocentric and other-centric choice, respectively) and the folder at the *back* is the one that was closer. In our discussion in section 7.6., we address why this remapping of the spatial terms may have been observed here.

With this revised assignment of the sagittal terms, we established that listeners in Experiment 2 responded predominately from the partner's perspective. As shown in Figure 2, their mean proportion of egocentric responses on critical trials was 27%, with 73% of responses being other-centric ( $SD = 45\%$ ).

An omnibus statistical model was built to examine the interaction between configuration type and instruction type on the proportion of egocentric responses. The overall variance captured by the omnibus model (with both fixed and random effects) was 80.54%. As indicated by the mean proportions of egocentric responses reported in Table 4, listeners were no more likely to make an egocentric choice on ego-aligned configurations compared to other-aligned configurations or neither-aligned ones (see Table 5).

The type of instruction did significantly predict egocentric responding, with listeners being significantly more likely to respond egocentrically on lateral *left-right* instructions (on 31% of the time,  $SD = 46\%$ ) than sagittal *front-back* instructions ( $M = 24\%$ ,  $SD = 42\%$ ). As shown in Table 4, the difference between the two instruction types was greater on neither-aligned configurations (a 12% difference) compared to ego-aligned ones (a 7% difference), and it was smaller on other-aligned configurations (a 3% difference) relative to ego-aligned ones. These patterns are substantiated by the interactions of the relevant contrasts for configuration and instruction types in Table 5.

The proportions of egocentric and other-centric responses were consistent with the listeners' overall perspective preference on critical trials. When classifying listeners as egocentric, other-centric, or mixed responders using the same criteria as before, there were 129 other-centric responders, 34 egocentric responders, and 21 mixed responders in Experiment 2. This distribution of other-centric, egocentric, and mixed responders was significantly different than that of Experiment 1a ( $\chi^2(2) = 15.18, p < .01$ ) and the original Duran et al (2011) study ( $\chi^2(2) = 11.55, p < .01$ ). It was not significantly different from that of Experiment 1b, which also included other-aligned cues,  $\chi^2(2) = 3.15, p = .21$ . As

with Experiments 1a and 1b, in Appendix B, we report an additional set of analyses—without mixed responders—given the fact that, despite our recoding of sagittal terms, some participants here still interpreted *front-back* the way the majority of participants in Experiments 1a and 1b did.

It is notable that, even though both Experiment 2 and Experiment 1a involved folders that were always ego-aligned, the proportion of egocentric responses in the two experiments in fact differed (27% vs. 47% egocentric choices on critical trials). As Table 4 shows, this difference is observed even when only the ego-aligned configurations of Experiment 2 are considered (also involving 27% egocentric responses).

Table 4. *Mean proportions of egocentric responses (and standard deviations) on critical trials across the different configural cue alignment conditions in Experiments 1a, 1b, and 2, and across verbal instruction (lateral: left-right, and sagittal: front-back).*

Experiment	Configural cue	Proportion of Egocentric Responses		
		Sagittal	Lateral	All instructions
1a	Ego-aligned	0.38 (0.49)	0.56 (0.50)	0.47 (0.50)
1b	Other-aligned	0.28 (0.45)	0.35 (0.48)	0.32 (0.47)
2	Ego-aligned	0.24 (0.43)	0.31 (0.46)	0.27 (0.45)
2	Other-aligned	0.25 (0.43)	0.28 (0.45)	0.27 (0.44)
2	Neither-aligned	0.22 (0.41)	0.34 (0.47)	0.28 (0.45)

*Table 5. Mixed-effects model for egocentric perspective choice, response times (log-transformed), divergence from optimal distance (log-transformed), and directional shifts, with fixed effects for perspective preference (ego, mixed, other) for all but the first model, configuration orientation (Configuration: ego-aligned, neither-aligned, other-aligned), and instruction type (sagittal: back-front; lateral: right-left), and with the maximal random effect structure possible. For fixed effects and their interactions, we report the coefficient and its standard error, along with the associated z- or t-value and p-value.*

Predictor	Egocentric choice				Response times				Distance				Directional shifts			
	$\beta$	SE	z	p	$\beta$	SE	t	p	$\beta$	SE	t	p	$\beta$	SE	t	p
Intercept	-2.32	0.30	-7.84	0.00	7.19	0.02	298.01	0.00	4.23	0.04	110.92	0.00	2.01	0.07	29.86	0.00
Pref: mixed vs. ego	—	—	—	—	-0.06	0.08	-0.74	0.46	0.10	0.13	0.76	0.45	-0.01	0.22	-0.05	0.96
Pref: other vs. ego	—	—	—	—	0.26	0.05	4.82	0.00	0.31	0.09	3.65	0.00	0.50	0.15	3.28	0.00
Config: neither- vs. ego-aligned	0.02	0.15	0.16	0.87	0.07	0.02	3.65	0.00	0.10	0.04	2.43	0.02	0.06	0.07	0.98	0.33
Config: other- vs. ego-aligned	-0.13	0.16	-0.82	0.41	-0.03	0.02	-1.75	0.08	0.03	0.04	0.61	0.54	0.02	0.07	0.24	0.81
Instruction: lateral vs. sagittal	1.00	0.19	5.17	0.00	0.12	0.02	7.04	0.00	0.12	0.03	3.69	0.00	-0.05	0.05	-0.98	0.33
Pref: mixed vs. ego * Config: neither- vs. ego-aligned	—	—	—	—	0.04	0.06	0.68	0.49	-0.07	0.13	-0.51	0.61	0.13	0.22	0.61	0.54
Pref: mixed vs. ego * Config: other- vs. ego-aligned	—	—	—	—	-0.01	0.06	-0.18	0.86	0.18	0.14	1.29	0.20	0.07	0.24	0.31	0.76
Pref: other vs. ego * Config: neither- vs. ego-aligned	—	—	—	—	0.14	0.04	3.47	0.00	0.23	0.09	2.56	0.01	0.51	0.15	3.38	0.00
Pref: other vs. ego * Config: other- vs. ego-aligned	—	—	—	—	-0.13	0.04	-2.81	0.00	-0.24	0.10	-2.48	0.01	-0.55	0.16	-3.35	0.00
Pref: mixed vs. ego * Instruction: lateral vs. sagittal	—	—	—	—	0.03	0.06	0.45	0.65	0.10	0.11	0.91	0.36	-0.04	0.16	-0.23	0.82
Pref: other vs. ego * Instruction: lateral vs. sagittal	—	—	—	—	0.16	0.04	4.20	0.00	0.05	0.07	0.71	0.48	0.43	0.11	3.85	0.00
Config: neither- vs. ego-aligned * Instruction: lateral vs. sagittal	1.35	0.30	4.56	0.00	0.04	0.04	0.98	0.33	-0.21	0.08	-2.57	0.01	-0.02	0.13	-0.13	0.89
Config: other- vs. ego-aligned * Instruction: lateral vs. sagittal	-1.06	0.32	-3.27	0.00	-0.02	0.04	-0.52	0.61	0.30	0.09	3.42	0.00	0.00	0.14	0.03	0.98
Pref: mixed vs. ego * Config: neither- vs. ego- * Instruction: lateral vs. sagittal	—	—	—	—	-0.13	0.12	-1.13	0.26	-0.75	0.26	-2.87	0.00	-0.79	0.43	-1.81	0.07
Pref: mixed vs. ego * Config: other- vs. ego- * Instruction: lateral vs. sagittal	—	—	—	—	0.03	0.13	0.23	0.816	0.57	0.29	2.00	0.05	0.50	0.47	1.06	0.29
Pref: other vs. ego * Config: neither- vs. ego- * Instruction: lateral vs. sagittal	—	—	—	—	0.40	0.08	4.92	0.00	0.91	0.18	5.04	0.00	0.93	0.30	3.12	0.00
Pref: other vs. ego * Config: other- vs. ego- * Instruction: lateral vs. sagittal	—	—	—	—	-0.12	0.09	-1.35	0.18	-0.56	0.20	-2.85	0.00	-0.72	0.33	-2.21	0.03

## 7.2. Response times

An omnibus statistical model was built to examine the three-way interaction between perspective preference, configuration alignment, and instruction type. Table 5 includes a summary of this model, separating out the results for each of the planned contrasts and their interactions. The overall variance captured by the omnibus model (with both fixed and random effects) was 38.87%.

Before interpreting the statistical significance of the contrasts of interest in Table 5, we examined the overall three-way interaction as well as the two-way interactions (involving the contrasts with perspective preference and configuration alignment, and the contrasts with perspective preference and instruction). The overall two-way interaction between perspective preference and configuration alignment was indeed statistically significant ( $\chi^2(4) = 22.22, p < .001$ ), as was that between perspective preference and instruction type ( $\chi^2(2) = 25.43, p < .001$ ), and the three-way interaction of these factors ( $\chi^2(4) = 28.03, p < .001$ ). The remaining two-way interaction between configuration alignment and instruction type was not significant ( $p = .61$ ).

Listeners were overall slower to select folders when the configuration was neither-aligned ( $M = 1683$  ms,  $SD = 954$  ms) than for the other two configuration types (ego-aligned:  $M = 1544$  ms,  $SD = 813$ ; other-aligned:  $M = 1525$  ms,  $SD = 811$ ), with the difference between neither-aligned and ego-aligned configurations being significant in predicting response times, as shown in Table 5. Listeners were also significantly slower on *left-right* ( $M = 1749$ ,  $SD = 960$  ms) than *front-back* ( $M = 1448$ ,  $SD = 761$  ms) instructions. Overall, egocentric responders were significantly faster than other-centric responders, taking on average 1333 ms ( $SD = 755$  ms) to select a folder compared to

1684 ms ( $SD = 874$  ms) for other-centric responders. Egocentric responders were not significantly faster than mixed responders ( $M = 1508$  ms,  $SD = 990$  ms) (see Table 5).

The two-way interactions between the perspective preference and configuration type contrasts in Table 5 suggest that, while other-centric responders took longer to respond than egocentric responders, this difference was *more* pronounced on neither-aligned configurations than ego-aligned ones (with the coefficient indicating a difference in the positive direction), and *less* pronounced on other-aligned configurations than ego-aligned ones (with a difference in the negative direction). The increased sensitivity of other-centric responders to the configural cue can also be observed in the mean response times presented in Table 3.

To decompose the significant overall three-way interaction between perspective preference, configuration alignment, and instruction type, reported above, we conducted a simple effects analysis by holding each level of the instruction type factor (i.e., lateral or sagittal) constant. This follow-up revealed that the interaction between perspective preference and configuration alignment was significant for lateral instructions ( $\chi^2(4) = 43.56$ ,  $p < .001$ ), but not for sagittal instructions ( $\chi^2(4) = 4.86$ ,  $p = 0.30$ ). In the model focusing on lateral instructions only, the two-way interactions between perspective preference (egocentric vs. other-centric) and configuration type were significant (for the ego-aligned vs. neither-aligned contrast:  $\beta = .34$ ,  $SE = .06$ ,  $t = 5.83$ ,  $p < .001$ ; for ego-aligned vs. other-aligned:  $\beta = -.18$ ,  $SE = .06$ ,  $t = -2.90$ ,  $p = .004$ ). This simple effects analysis further qualifies the 2-way interactions involving these effects shown in Table 5: it suggests that the processing cost that other-centric responders experienced relative to egocentric responders, which was more pronounced on neither-aligned configurations

and less pronounced on other-aligned configurations (relative to ego-aligned ones), was most evident on lateral instructions.

### 7.3. Divergence from optimal distance

The omnibus statistical model for divergence from optimal distance, reported in Table 5, captured 20.97% of the overall variance. The overall two-way interaction between perspective preference and configuration alignment was statistically significant ( $\chi^2(4) = 9.88, p = .04$ ), as was the interaction between configuration alignment and instruction type ( $\chi^2(2) = 12.99, p = .0015$ ), and the three-way interaction of these factors ( $\chi^2(4) = 26.14, p < .001$ ). The two-way interaction between perspective preference and instruction type was not significant overall ( $p = .25$ ).

Similar to the patterns in response times, listeners made overall longer trajectories when folders were neither-aligned ( $M = 123.30, SD = 123.00$  pixels) than for the other two configuration types (ego-aligned:  $M = 101.60, SD = 101.80$ ; other-aligned:  $M = 106.50, SD = 103.20$  pixels). The difference between neither-aligned and ego-aligned configurations significantly predicted the trajectories' divergence from optimal distance, as shown in Table 5. Listeners also experienced greater difficulty on lateral instructions than sagittal ones, making more divergent trajectories (lateral:  $M = 120.5, SD = 119.30$ ; sagittal:  $M = 104.10, SD = 103.80$  pixels). Again, as shown in Table 5, egocentric responder made more efficient mouse trajectories ( $M = 83.11, SD = 86.85$  pixels) than other-centric responders ( $M = 120.70, SD = 118.10$  pixels). Again, mixed responders did not significantly differ from egocentric responders in their trajectories ( $M = 108.40, SD = 102.50$  pixels).



Following the same pattern as response times, the two-way interactions between perspective preference and configuration type in Table 5 indicate that, while other-centric responders made more divergent trajectories than egocentric responders, this was more pronounced on neither-aligned configurations and less pronounced on other-aligned configurations (compared to ego-aligned configurations).

A follow-up simple effects analysis, while held each level of the instruction type factor constant, revealed—in line with response times—that the interaction between perspective preference and configuration alignment was significant for lateral instructions ( $\chi^2(4) = 31.32, p < .001$ ), but not for sagittal instructions ( $\chi^2(4) = 4.18, p = .38$ ). In the model focusing on lateral instructions only, the two-way interactions between perspective preference (egocentric vs. other-centric) and configuration type were significant (for the ego-aligned vs. neither-aligned contrast:  $\beta = .69, SE = .12, t = 5.35, p < .001$ ; for ego-aligned vs. other-aligned:  $\beta = -.52, SE = .14, t = -3.76, p < .001$ ). This qualifies the corresponding two-way interactions involving these effects (perspective preference and configuration alignment), shown in Table 5. Consistent with response times, this finding suggest that the processing cost experienced by other-centric responders relative to egocentric ones (which was more pronounced on neither-aligned configurations and less pronounced on other-aligned configurations) held primarily for lateral instructions.

#### **7.4. Directional shifts**

The omnibus statistical model for directional shifts is reported in Table 5; this model captured 21.92% of the overall variance. Following the same approach as with the previous measures, we established that the overall two-way interaction between perspective preference and configuration alignment was statistically significant ( $\chi^2(4) =$

23.38,  $p < .001$ ), as was the interaction between perspective preference and instruction type ( $\chi^2(2) = 19.13$ ,  $p < .001$ ), and the three-way interaction of perspective preference, configuration alignment, and instruction type ( $\chi^2(4) = 10.61$ ,  $p = .03$ ). The two-way interaction between configuration alignment and instruction type was not significant ( $p = .99$ ).

Listeners made numerically more directional shifts on neither-aligned configurations ( $M = 2.23$ ,  $SD = 1.58$ ) than ego-aligned ( $M = 2.11$ ,  $SD = 1.43$ ) and other-aligned configurations ( $M = 2.08$ ,  $SD = 1.43$ ). However, these differences were not significant, as shown in Table 5.

Consistent with the previous measures, egocentric responders made more efficient mouse trajectories than other-centric responders (see Table 5), involving fewer directional flips (egocentric responders:  $M = 1.75$ ,  $SD = 1.31$ ; other-centric responders:  $M = 2.28$ ;  $SD = 1.53$ ). Again, mixed responders did not significantly differ from egocentric responders in terms of the directional flips they made ( $M = 2.02$ ,  $SD = 1.40$ ).

Following the same pattern as the previous measures, two-way interactions between perspective preference and configuration type in Table 5 indicate that, while other-centric responders made more directional shifts than egocentric responders, this was more pronounced on neither-aligned configurations and less pronounced on other-aligned configurations (compared to ego-aligned configurations).

As with the previous measures, we decomposed the overall three-way interaction with a follow-up simple effects analysis, keeping each level of the instruction type factor constant. As with the previous measures, the interaction between perspective preference and configuration alignment was significant for lateral instructions ( $\chi^2(4) = 28.92$ ,  $p <$

.001), but not for sagittal instructions ( $\chi^2(4) = 5.22, p = .26$ ). In the model focusing on lateral instructions only, the two-way interactions between perspective preference (egocentric vs. other-centric) and configuration type were significant (for the ego-aligned vs. neither-aligned contrast:  $\beta = .97, SE = .21, t = 4.59, p < .001$ ; for ego-aligned vs. other-aligned:  $\beta = -.91, SE = .23, t = -3.93, p < .001$ ). Again, this qualifies the two-way interactions of these factors reported in Table 5, suggesting that the patterns observed—whereby other-centric responders experience a processing cost relative to egocentric responders, which is accentuated on neither-aligned configurations and tempered on other-aligned configurations—is limited to lateral instructions.

### 7.5. Dynamics across trials

As with Experiments 1a and 1b, we explored how the listeners' responses stabilized over time, with an interest in whether responses on the three types of configurations stabilized differently across trials. We pursued this by including the order of the trial as a predictor along with its interaction with the other factors (perspective preference, configuration type, and instruction type) in linear mixed effects models on the dependent measures for critical trials.

Trial order predicted performance for response times ( $\beta = -.10, SE = .01, t = -9.30, p < .001$ ), but not for divergence from optimal distance ( $\beta = -.03, SE = .02, t = -1.70, p = .09$ ), or directional shifts ( $\beta = -.003, SE = .03, t = -.12, p = .90$ ). The interactions of trial order with the configuration type contrasts—the main factor of interest—were not significant predictors in any of these models, nor were the interactions of trial order with the perspective preference contrasts.

These analyses suggest that, beyond the fact that listeners were faster to respond over time, they did not reliably change over time the dynamics of the mouse trajectories as a function of the types of configurations they viewed or their perspective strategy. In contrast to Experiments 1a and 1b, where we found evidence that the configural cue—which was consistently aligned with a particular perspective—did influence the stabilization of the other-centric strategy (as illustrated in Figures 3a and 3b), in Experiment 2, the configural cue—which was inconsistent—did not impact responses over time.

Other-centric responders, who constituted the majority in Experiment 2, stabilized quickly on their perspective strategy (certainly by the first critical trial, which was the 9<sup>th</sup> trial in the session). We do not report the results of the model on the proportion of egocentric responses, as that mixed logistic regression model failed to converge even with a backwards fitting approach. Nevertheless, the response preference of other-centric responders did not appear to differ over time across the three types of configurations. As we had anticipated, participants adhered to a consistent strategy in Experiment 2, despite the variable configural cue.

## 7.6. Summary

In this experiment, we set out to examine whether a configural feature—namely the orientation of the configuration’s axis of symmetry—influences the dynamics of listeners’ responses when it is paired *inconsistently* with candidate social perspectives.

Overall, we found some modest evidence that the alignment of the configuration influences the dynamics of the listeners’ responses. Listeners made trajectories that took numerically the longest, diverged the most from optimal distance, and involved the

greatest number of directional shifts when configurations were neither-aligned than when they were aligned with a social perspective. The processing cost of neither-aligned configurations relative to ego-aligned configurations was reliable for response times and divergence from optimal distance.

As we had predicted, other-centric responders showed more sensitivity to the configural cue than the other responders. First, we had anticipated that other-centric responders would show facilitation on other-aligned configurations. Indeed, while other-centric responders experienced a cost in processing relative to egocentric responders, this cost was reduced on other-aligned configurations (relative to ego-aligned ones) for all mouse-tracking measures.

Moreover, we found evidence that, relative to egocentric responders, other-centric responders showed a decrement in performance on neither-aligned configurations (relative to ego-aligned configurations), making trajectories that were slower, more divergent from optimal distance, and with more directional shifts. The difficulty of neither-aligned configurations extends the earlier findings of Galati and Avraamides (2013, 2015), who had shown that in unconstrained interactions speakers do experience a decrement in performance on neither-aligned configurations (relative to ego-aligned or other-aligned ones), by using more egocentric descriptions and taking longer to coordinate while collaborating on the reconstruction of those configurations.

Since the patterns we observed pertain to other-centric responders, this may suggest that adopting another's perspective increases sensitivity to configural cues that have the potential to facilitate judgments from social perspectives. For example, listeners who are sufficiently motivated or attentive to adopt the partner's perspective may also be

more likely to pay more attention to other cues in the configuration, including those emerging from the orientation of the configuration.

Although the configuration's orientation had a selective impact on dynamics of participants' responses, the addition of this configural cue to the task originally used by Duran, Dale, and Kreuz (2011), resulted in some strikingly different patterns of performance. Similar to Experiment 1b, which also introduced an other-aligned configural cue, other-centric responding became more predominant than in Duran, Dale, and Kreuz (2011, Study 1) and its replication in Experiment 1a here, both of which used ego-aligned folders throughout the whole experiment. The difference in egocentric choices with Experiment 1a persists even when we consider only the ego-aligned trials of Experiment 2. Moreover, as we report, the distribution of other-centric, egocentric, and mixed responders in Experiment 2 differed from that of both the original study of Duran et al (2011) and of Experiment 1a, while it was comparable to that of Experiment 1b. We will return to this point in the General Discussion.

Another striking difference between Experiment 2 and the previous experiments (1a and 1b here, and those of Duran, Dale, and Kreuz, 2011) concerns the interpretation of the spatial terms *front-back*. Listeners in Experiment 2 reversed the mapping of the *front-back* terms onto space, relative to listeners from the previous experiments. In the previous experiments, the majority of the listeners interpreted the folder "at the front" as referring to the one closer to the person whose perspective was being adopted, whereas here it referred to the folder farther from the person whose perspective was being adopted. As we have discussed, in Experiments 1a and 1b, a subset of the responders classified as mixed had also adopted this alternative mapping of the sagittal terms.

However, in Experiment 2, the overwhelming majority of responders used this mapping, as indicated by performance on control trials: listeners used this mapping on 92% of *front-back* instructions.

This dramatic shift in the interpretation of *front-back* terms could have been induced by the nature of the configural cue used in Experiment 2: the orientation of a triangular multi-object configuration that could be construed as having an intrinsic “front” (or “top”), represented by the isosceles triangle’s top vertex. With the exception of four control trials (2 ego-reversed and 2 other-reversed configurations), the folder at the triangle’s “top” or “front” was always farther away from the position of the person with whom the configuration was aligned (whether the listener or speaker). That is, the triangle always “pointing away” from that person’s position, as shown in Figure 4a and 4b. This consistent configural pairing on person-aligned configurations (i.e., with the triangle’s top vertex consistently farther from that person’s position) could have motivated the interpretation of “front” as the “farther away” folder.

In sum, the alignment of the configuration, which varied within-participants here, had a modest and selective impact on the dynamics of the listeners’ responses. Other-centric responders did not experience facilitation per se on other-aligned configurations, but they exhibited sensitivity to the configural cue. While other-centric responders experienced an overall cost in processing relative to egocentric responders, this cost was lessened when the configuration’s alignment supported the other-centric perspective and was increased when the configuration’s alignment supported neither social perspective. Moreover, the introduction of this varying configural cue had a drastic impact on the listeners’ perspective preference and their interpretation of sagittal terms: it increased

other-centric responding relative to experiments where cues were always ego-aligned, and it caused listeners to use the opposite mapping of the spatial terms *front-back*.

We should note that the follow-up experiment we conducted to examine how individuals perceived the alignment of the 3-object configurations complicates, in some respects, the interpretation of some of our results here. That follow-up experiment, reported in Appendix C, addresses the possibility that the three objects may not have been perceived as a triangle that afforded a directional cue in the way we had intended. The classifications of the configurations (as “ego-aligned”, “other-aligned”, “neither-aligned”, or “both-aligned”) made by a new set of participants suggests that those participants likely perceived ego-aligned configurations as other-aligned, and vice-versa: we found that “ego-aligned” and “other-aligned” classifications experienced interference from the competing configuration type.

As we propose in Appendix C, this might be the case if participants use the orientation of the *base* of the triangle (rather than its perpendicular vertex) to make explicit judgments about the triangle's alignment. Using this geometric cue, our ego-aligned configuration in Figure 4a would indeed be classified as other-aligned, and the other-aligned configuration in Figure 4b as ego-aligned.

We cannot say definitively that participants in Experiment 2 also apprehended the alignment of the configurations in the same way as participants in the follow-up study, seeing that the Experiment 2 participants did not have to make explicit judgments about the configurations' alignment. Nevertheless, it is possible that they did also perceive the alignment of the configurations according to the orientation of the triangle's base. If that's the case, our finding that the mouse trajectories of other-centric responders in Experiment



2 showed facilitation on other-aligned configurations relative to ego-aligned ones should be reinterpreted as the reverse pattern. We return to this caveat in the General Discussion.

Despite complicating the interpretation of this particular finding, the results of the follow-up experiment do ratify that participants appear to interpret these 3-object configurations as triangles. Earlier in this section, we had relied on participants' conceptualization of the 3 objects as triangles to explain why, for the majority of participants in Experiment 2, the *front-back* terms were mapped in the opposite direction than in Experiments 1a and 1b. We proposed that this dramatic shift arose from mapping the spatial term "front" onto the triangle's intrinsic top (its top vertex).

Alternative conceptualizations of the three objects (e.g., as an L-shape or not as a triangle at all, given the violation of the gestalt continuity principle with the 3<sup>rd</sup> non-folder object) cannot easily account for this finding. If we were to assume that listeners apprehended configurations as Ls, the folder at the L's "top"/"front" would have a different identity than the folder typically selected by participants. For instance, a listener apprehending the configuration in Figure 4b as a triangle and responding other-centrally to an instruction to select "the folder at the front" should select the folder farther away from the partner (at the top/the vertex of the triangle), consistent with what we found. In contrast, a listener apprehending that same configuration as an L and responding other-centrally should select the folder closer to the partner (at the top of the L), which is generally inconsistent with the data. (Conceptualizing configurations as an L-shape also does not account well for the classification results of participants in the follow-up study, as we show in Appendix C: classifications based on that conceptualization were no better than chance.) Thus, any alternative explanations for how

participants apprehended the three objects should be able to account coherently for the results of Experiment 2, including the observed flip in the interpretation of the sagittal terms.

## 9. General Discussion

We set out to examine how configural cues in environmental scenes influence the ease and likelihood with which language users adopt a particular perspective. Specifically, we asked whether the social partner's perspective exhibits facilitation when it is supported by configural cues. In Experiment 1a, we used a configural cue—the orientation of the intrinsic axis of two objects on the tabletop scene—that did not reinforce the simulated partner's perspective, as it was always aligned with the egocentric perspective, reproducing the design of Duran et al. (2011, Study 1). In Experiment 1b, we used the same 2-object configurations as in Experiment 1a, except that the orientation of the objects was always aligned with the task partner's changing position (other-aligned). In Experiment 2, we used another type of configural cue—the axis of symmetry of a 3-object configuration—whose alignment we manipulated across trials. This manipulation yielded configurations that on critical trials were aligned with the participant (ego-aligned), the task partner (other-aligned), or neither perspective (neither-aligned).

Our findings demonstrate that introducing a configural cue to a task promotes an other-centric perspective strategy. When we introduced configural cues that were either consistently or inconsistently aligned with the other-centric perspective—i.e., in Experiments 1b and 2—we observed an increase in the listeners' preference for an other-centric response strategy relative to the absence of such other-aligned cues (i.e., in Experiment 1a and the original study by Duran et al., 2011). Moreover, as we observed

when examining object selections over time, listeners stabilized more readily on an other-centric perspective strategy when the configural cue consistently reinforced the other-centric perspective (in Experiment 1b) than when it did not (Experiment 1a).

Introducing a configural cue that converges with the partner's perspective may make that perspective more salient, or alternatively may highlight the difficulty of the task for the partner. For example, the changing orientation of the configural cue in Experiments 1b and 2 may have been attention grabbing enough to increase sensitivity to the partner's perspective and to other task features (including the partner's position), resulting in an increase in other-centric response strategy. Alternatively, by noticing the variability in the configural cue's orientation, listeners may have made the assessment—whether explicitly or implicitly—that the task was more difficult for the speaker (vs. for themselves), thus becoming more likely to adopt the speaker's perspective. This possibility is in line with the view that interlocutors share responsibility for mutual understanding (e.g., Clark & Wilkes-Gibbs, 1986; Clark, 1996), and adapt their language and behavior to help the partner when they perceive the partner as being more limited in contributing to the task (e.g., when they believe their partner is a child, Newman-Norlund et al., 2009, unfamiliar with the environment, Hölscher, Tenbrink, & Wiener, 2011, or has worse spatial ability than they do, Schober, 2009).

The impact of configural cues on the selection of perspective strategy is consistent with prior work showing that such cues can determine the preferred perspective or “organizing direction” of maintaining spatial information in memory, as reflected by the ease with which people reason from different imagined perspectives when making judgments about previously studied scenes (Marchette & Shelton, 2010; Mou &

McNamara, 2002; Richard & Waller, 2013). The present work also extends previous work on “reference frame selection” during the interpretation of individual spatial descriptions (e.g., Burigo & Sacchi, 2013; Burigo, Coventry, Cangelosi & Lynott, 2016; Carlson & Van Deman, 2008), by lending insight into the dynamics of the listener’s perspective selection within and across individual spatial instructions.

Beyond the observed differences in perspective strategy across Experiments 1a and 1b, there wasn’t strong evidence that the configural cue influences the response dynamics on individual trials. This could be because in the two experiments the pairing of the configural cue with social perspectives was consistent: the configural cue was always ego-aligned in Experiments 1a, and always other-aligned in Experiments 1b. In these experiments, the listeners’ response dynamics were predicted by their perspective preference instead.

In Experiment 2, where the pairing of the social perspectives with the configural cue varied continually across-trials, the impact of the configural cue on the listeners’ response dynamics was subtle. Overall, the listeners’ response dynamics on ego-aligned configurations did not differ from those on other-aligned configurations, although trajectories were more efficient—in terms of response time and divergence from optimal distance—than those on neither-aligned configurations.

The impact of the configural cue on response dynamics, in Experiment 2, was selective and limited to other-centric responders. This could be because the continually changing orientation of the configuration may have not been a salient enough cue, and its relevance to the task may have been obvious only to those attending to the partner’s position. Other-centric responders, who experienced an overall decrement in processing

relative to ego-centric responders, showed a reduction of this decrement on other-aligned configurations and an increase of this decrement on neither-aligned configurations (relative to ego-aligned ones).

However, we hesitate to make the claim that other-centric responders experienced a relative facilitation in performance on other-aligned configurations, given the findings of the follow-up study to Experiment 2 (in Appendix C). The explicit alignment judgments of a new set of participants in that follow-up study suggested that what we considered to be “other-aligned” configurations in Experiment 2 may have been perceived as “ego-aligned” configurations by listeners in Experiment 2, and vice versa. If that is the case, other-centric participants may have in fact experienced a facilitation when responding on ego-aligned configurations (relative to other-aligned configurations), although the relative cost of processing neither-aligned configurations persists. We acknowledge that there may be an inconsistency between what we, the researchers, took to be an other-aligned cue, and what the participants in Experiment 2 may have taken to be response relevant. There may be considerable variability in how such stimuli are perceived, which offers ideas for future research on mapping participant expectations about configural features. To better model how configural cues guide response dynamics, it may be valuable to map participants’ explicit judgments about configural cues onto their responses. We leave that as an open question that naturally follows this work.

In either case, our findings do underscore that, relative to Experiment 1b—where the configural cue’s orientation was consistently paired with the speaker’s changing position—the inconsistent pairing of task features in Experiment 2 (the orientation of the

configuration and the speaker position) made listeners more susceptible to the influence of the configural cue when they were other-centric.

The finding that the alignment of the configural cue had a limited impact on the listeners' response dynamics—either no effect in Experiments 1a and 1b, or a selective effect in Experiment 2—could be due to the fact that, in all experiments, the speaker's position changed across trials. This could have made other-centric responding relatively difficult in all experiments, since interpreting spatial instructions from changing headings is taxing on cognitive resources (e.g., Avraamides & Carlson, 2003; Ryskin et al., 2014; Ryskin, Wang, & Brown-Schmidt, 2016). Across all experiments, other-centric responders were slower, made longer or more divergent trajectories that included more directional shifts than egocentric responders. This finding can be seen as congruent with evidence that the egocentric perspective causes interference in early processing when taking another's perspective into account (e.g., Duran, Dale, & Kreuz, 2011; Keysar, Barr, & Horton, 1998; Keysar, Barr, Balin, & Paek, 1998). Nevertheless, listeners invested the effort to adopt the task partner's perspective: despite its associated cognitive cost, the other-centric response strategy was predominant in all experiments—and even more so when a configural cue was introduced.

Beyond the selective impact of the configural cues on the listeners' response dynamics, there was also an unexpected and dramatic impact of these cues on the interpretation of sagittal spatial terms. In Experiment 2, we observed a dramatic shift in how listeners interpreted *front* and *back*: on 92% of these sagittal control trials, listeners used the reverse mapping than the one anticipating, by assigning the folder at the “front” to be the one farther from the person whose perspective they were adopting rather than

the one closer. As discussed in section 7.6., this could be because of the nature of the configural cue (i.e., a triangular configuration with an axis of symmetry): the configuration could be apprehended as having an intrinsic “front” or “top”, which on person-aligned critical trials was always farther away from that person’s position. An object’s “face” or “front”—including the location of the eyes, its functionality, and its direction of motion—is often thought to be a very strong cue for assigning direction (Jackendoff, 1996). In Experiment 2, despite objects being inanimate, the “front” of their configuration appeared to serve as a strong cue for assigning direction, shaping how sagittal spatial terms were interpreted. In Experiment 1b, although there wasn’t such a dramatic shift in how sagittal terms were interpreted, we did observe that mixed responders also used the same reversed mapping on nearly half of the sagittal control trials. This could be because, based on the intrinsic axes of the objects, these listeners construed the 2-object configurations as having an intrinsic “front” or “top” (with the “front” folder being farther away from the location of the person whose perspective they adopted).

Relative to sagittal terms, lateral terms incurred a greater cognitive cost in all experiments. This was driven by other-centric responders: compared to egocentric responders, they took longer to respond on lateral than on sagittal instructions, made trajectories that included more directional shifts, and were longer or more divergent from the optimal path to the folder. These findings are compatible with proposals that mapping *left* and *right* to appropriate regions of space is slower because it requires defining the *front-back* dimension first, and additionally because the *left-right* axis is highly symmetric (relative to a person’s body), containing no salient cues to differentiate *left*

from *right* (e.g., Avraamides & Sofroniou, 2006; Franklin & Tversky, 1990). Egocentric responders were less affected by the type of instruction and other task features, as these were irrelevant to their response strategy. Given their fixed perspective, the mappings of spatial terms (*left, right, front, back*) onto space remained stable across trials for egocentric responders. This stable mapping could have made less relevant the other cues that varied on displays, including the speaker's position (examined in the statistical models for Experiments 1a and 1b).

Beyond the different response profiles of egocentric and other-centric responders, the fact that individuals responding to identical instructions had such dramatically different interpretations of sagittal spatial terms in different contexts underscores the ambiguity of spatial terms. In dialogue settings where feedback can be exchanged contingently, language users explicitly negotiate which perspective or spatial schema to use (e.g., Galati, Panagiotou, Tenbrink, & Avraamides, 2017; Schober, 2009; Garrod & Anderson, 1987). But here, in the absence of contingent feedback from the simulated partner, language users had to resolve the ambiguity by making a strategic perspective selection and adhering, in most cases, to a consistent perspective strategy.

Resolving such ambiguity is often thought to occur through the selection of reference frames (Carlson, 1999; Logan & Sadler, 1996). The reported experiments could have accommodated several candidate reference frames that differed in terms of the orientation, direction, and origin of their axes, which defined space. These could have included, among other options: a relative reference frame with the listener (ego) as its origin and their fixed facing direction determining the orientation of the axes; a relative reference frame with the speaker (other) as the origin and their variable facing direction



determining the orientation of the axes; an intrinsic reference frame with the center of the table as the origin, and with the intrinsic axes of the folders determining the directionality of axes of the reference frame; in Experiment 2, an intrinsic reference frame with the midpoint of the base of the triangular configuration as the origin and the top vertex determining the directionality of the axes. While we had initially formed our predictions under the assumption that only the first two options would be at play in our design, the listeners' interpretation of sagittal spatial terms suggests that additional reference frames were likely at play, in whole or in part. Indeed, some work suggests that components of different reference frames (e.g., selected axes and their endpoints) can be independently accessed, and that non-selected reference frames are not entirely inhibited (Carlson & van Deman, 2008).

We propose that the co-activation and competition of reference frames—or more broadly the mappings of linguistic terms onto space—differ depending on how social and configural cues co-occur. For example, they may differ depending on whether cues are aligned consistently (as in Experiments 1a and 1b) or inconsistently (as in Experiment 2). As we saw, the patterns of co-occurrence of these cues can result in radically different interpretations of spatial terms. This view is consistent with a cognitive dynamics account where selection involves integrating multiple sources of probabilistic information, with probabilities set by in-the-moment and long-scale demands, cues, and environmental constraints. The selection outcome is not a simple linear progression of stages, but a non-monotonic and nonlinear combination of information (Duran, Dale, & Galati, 2016). Since the task partner's perspective was always changing across trials here, future research should explore how perspective selection and its unfolding dynamics are

influenced by the consistency of the partner's perspective and its coincidence with configural features. For example, in a study using a similar paradigm, in which the speaker's perspective was held constant across trials (at 90°), while the *listener's* depicted perspective around the tabletop changed across trials and was thus dissociated from the participant's sensorimotor perspective, the egocentric perspective no longer exhibited a processing advantage (Galati, Diavastou, & Avraamides, 2018).

In light of these findings, we argue that it is not always productive to construe perspective-taking as involving a “baseline” or “default” state, as this places an undue theoretical emphasis on early and rapid processes. Our findings underscore that the perspective that often characterizes behavior best, under a given constellation of task features, is not necessarily the first or quickest cognitive state to be reached. Instead, what best characterizes behavior is the pattern that emerges beyond the initial milliseconds of processing, under a longer timescale, and as a strategy in the task. In this sense, the “default” should be regarded as the most stable perspective that emerges as the target of processing over a longer timescale (see also, Dale, Galati, Alviar, Contreras Kallens, Ramirez-Aristizabal, Tabatabaeian, & Vinson, in revision).

As we have shown, in this sense, the listeners' “default” or most stable perspective choice is inextricably tied to task constraints, including configural constraints. This means that any “default” or “baseline” we uncover is specific to those task features. For example, assuming the existence of a default perspective state, one might ask whether the shift in perspective preference observed across Experiments 1a and 1b is due to participants becoming more other-centric in Experiment 1b or less egocentric in Experiment 1a. An investigation that removes configural cues (i.e., by using objects

that do not carry any orientation cues) could shed light to this question. Nevertheless, we would argue that such an undertaking would only reveal a baseline perspective that is tethered to the remaining relevant features of the task (importantly, the social attribution that the task partner was a simulated partner). Indeed, in the previously mentioned study by Galati, Diavastou, and Avraamides (2018), the directional cues of the objects were eliminated (only CDs were used as objects), and consistently with Duran et al. (2011), the listeners' preferred perspective in each experiment was predicted by attributions about the partner (e.g. believing that the speaker was the experimenter vs. another participant) or other task features related to task difficulty (e.g., having the listener's perspective change across trials, rather than the speakers). Rather than uncovering a "baseline" or "default" perspective in a narrow task context, we are interested precisely in how changes in task contexts elicit shifts in the predominant, most stable perspective.

This approach, with its focus on perspective shifts under different cue constellations, can offer important insights for real world behavior. For example, one principle emerging from the context-sensitive shifts in perspective preference that we have observed is that reinforcing the other-centric perspective with configural cues (e.g., the orientation of objects) can promote its adoption.

We acknowledge that the experimental paradigm used in here may be reductive with respect to the complexities of real life communication (e.g., involving simple top-down configurations, unidirectional communication, and a simulated partner). However, it also affords a controlled environment in which to make an initial foray into the dynamics of perspective-taking. Despite the limitations of the task, our findings are in many respects consistent with other experimental work that has examined perspective-

taking in more naturalistic and interactive settings. For instance, Galati and colleagues (Galati & Avraamides, 2013, 2015; Galati et al., 2017) found evidence consistent with the idea that people can apprehend and use configural cues in tasks in which conversational partners are embedded in the same real-world environment. In that work, partners apprehended the axis of symmetry of a spatial configuration they were jointly reconstructing, and adapted the preferred perspective from which they remembered that array, as well as their spatial descriptions, according to the alignment of this cue with the egocentric and other-centric perspective.

The convergence of configural cues and social perspectives appears to influence perspective-taking behavior across different kinds of settings—both computer-mediated and with co-present partners—with implications about how these cues might play out in more ecologically valid settings. In such settings, we should acknowledge, multiple social and environmental cues beyond the ones we have examined could be at play. For example, the proximity of objects to the social partner has been shown to play a role in how requests are processed: in a context where a request for an object can refer to two plausible candidate objects, listeners are more likely to fixate their gaze on the object farther from the speaker, since it would be infelicitous for the speaker to request the object that is within reach (Hanna & Tanenhaus, 2004). Perceptually available information—whether it concerns salient features of the environment (e.g., object orientation) or socially relevant features (e.g., partner’s location, proximity to objects, their gaze, and so on)—guides language processing, highlighting the close interaction of language and vision (Vinson, Dale, Tabatabaeian & Duran, 2015)

The present work has served as an initial step in clarifying that interaction by addressing how environmental and social cues interact during language processing—a question that has been underexplored. In future work, this undertaking can be scaled up to more naturalistic settings, permitting the systematic parameterization of environmental and social cues, and can take into account individual differences that may modulate perspective-taking (e.g., including age, Long, Horton, Rohde, & Sorace, 2018, and executive functioning, Brown-Schmidt, 2009).

In sum, our findings suggest that a salient configural cue can increase the likelihood of other-centric responding and can have a significant impact on how linguistic spatial terms are interpreted. Moreover, other-centric responding is more sensitive to the influence of configural cues. We saw that other-centric responders stabilize on their other-centric perspective strategy more readily when that perspective is reinforced by a configural cue (Experiment 1b), and show signatures of the influence of the configural cue in their mouse movements (Experiment 2).

Our work underscores that the way we design task interfaces—and how we represent agents and relevant objects within them—can have a great impact on perspective-taking, and by extension on coordination and performance. This may be especially true in tasks that involve coordinating with others under conditions that are noisy, that involve unidirectional communication channels, temporal lags, and other limitations in shared affordances. Understanding which task features support the selection of non-egocentric perspectives will therefore be critical to optimizing behavior in a host of real-world situations involving perspective-taking, from following a drone's path in a viewfinder to interacting with others in real and in virtual spaces.

**Authors' note**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 705037 to A.G. We are thankful to Moreno Coco for useful discussions about statistical modeling. We are also thankful to Michele Burigo and two anonymous reviewers for their constructive comments.

## References

- Avraamides, M. N., & Sofroniou, S. G. (2006). Spatial frameworks in imagined navigation. *Psychonomic Bulletin & Review*, 13, 510-515.
- 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.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46(2), 391-418.
- 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(3), 255-278.
- Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 95-129.
- 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.
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1, 274-291.
- Bretz, F., Hothorn, T., & Westfall, P. (2010). *Multiple Comparisons Using R*. CRC Press: Boca Raton.

- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin and Review*, 16, 893-900.
- Burigo, M. & Sacchi, S. (2013). Object orientation affects spatial language comprehension. *Cognitive Science*, 37, 1471–1492.
- Burigo, M., Coventry, K., Cangelosi, A., & Lynott, D. (2016). Spatial language and converseness. *The Quarterly Journal of Experimental Psychology*, 69, 2319-2337.
- Carlson, L. A. (1999). Selecting a reference frame. *Spatial Cognition and Computation*, 1, 365-379.
- Carlson, L. A., & Van Deman, S. R. (2008). Inhibition within a reference frame during the interpretation of spatial language. *Cognition*, 106, 384-407.
- Carlson-Radvansky, L. A. & Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language*, 33, 646- 671.
- Carlson-Radvansky, L. A. & Logan, G. D. (1997). The influence of reference frame selection on spatial template construction. *Journal of Memory and Language*, 37, 411-437.
- 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.
- Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., & Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Experimental Brain Research*, 207(1-2), 95-103.
- Dale, R., Galati, A., Contreras Kallens, P., Alviar, C., Ramirez-Aristizabal, A., Vinson, D., & Tabatabaeian, M. (in revision). Perspective-taking is complex and dynamic.



- [http://co-mind.org/rdmaterials/php.cv/pdfs/article/dale\\_galati\\_et\\_al\\_perspective.pdf](http://co-mind.org/rdmaterials/php.cv/pdfs/article/dale_galati_et_al_perspective.pdf)
- Duran, N. D. (2011). Uncovering the hidden cognitive processes and underlying dynamics of deception (Dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3485883).
- Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, 121, 22-40.
- Duran, N. D., & Dale, R. (2014). Perspective-taking in dialogue as self-organization under social constraints. *New Ideas in Psychology*, 32, 131-146.
- Duran, N., Dale, R., & Galati, A. (2016). Toward Integrative Dynamic Models for Adaptive Perspective Taking. *Topics in Cognitive Science*. 8(4), 761-779.
- Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 63.
- Galati, A., & Avraamides, M. N. (2015). Social and representational cues jointly influence spatial perspective-taking. *Cognitive Science*. 39, 739-765.
- Galati, A., Diavastou, A., & Avraamides, M. N. (2018). Signatures of cognitive difficulty in perspective-taking: Is the egocentric perspective always the easiest to adopt?. *Language, Cognition, and Neuroscience*, 33, 467–493. DOI: 10.1080/23273798.2017.1384029
- 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.

- Galati, A., Panagiotou, E., Tenbrink, T., & Avraamides, M. N. (2017). Dynamic strategy selection in collaborative spatial tasks. *Discourse Processes*. 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.
- Gibson, J. (1979). *The ecological approach to visual perception*. Houghton-Mifflin: Boston.
- Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4), 596-615.
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28, 105–115.
- 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.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91-117.
- Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M. Peterson, L. Nadel & M. Garrett, eds., *Language and Space* (pp. 1-30). Cambridge, MA: MIT Press.
- Johnson, P. (2014). Extension of Nakagawa & Schielzeth's R<sup>2</sup> GLMM to all random slopes models. *Methods in Ecology and Evolution*, 5, 944–946.

- 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.
- 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, 46-50.
- Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory & Language*, 56, 436-455.
- Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity* (Vol. 5). 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.
- Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M.A. Peterson, L. Nadel and M. Garrett (eds.), *Language and Space*. Cambridge, MA: MIT Press, pp. 493–529.
- Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition*, 170, 25-30.
- Lupyan, G. (2017). The paradox of the universal triangle: concepts, language, and prototypes. *The Quarterly Journal of Experimental Psychology*, 70(3), 389-412.

- 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.
- 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.
- 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.
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press.
- Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 162-170.
- 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.
- Nardi, D., Newcombe, N. S., & Shipley, T. F. (2011). The world is not flat: Can people reorient using slope? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 354-367.
- Newman-Norlund, S. E., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A. C., De

- Ruiter, J. P., Hagoort, P., et al. (2009). Recipient design in tacit communication. *Cognition*, 111, 46–54.
- Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, 46, 688-704.
- Quinlan, P. T., & Humphreys, G. W. (1993). Perceptual frames of reference and two-dimensional shape recognition: Further examination of internal axes. *Perception*, 22, 1343-1364.
- Richard, L., & Waller, D. (2013). Toward a definition of intrinsic axes: The effect of orthogonality and symmetry on the preferred direction of spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/a0032995
- 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.
- 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.
- Samson, D., Apperly, I. A., Braithwaite, J. J., & Andrews, B. A. (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.
- Schober, M.F. (1993). Spatial perspective-taking in conversation. *Cognition*, 47, 1-24.

- 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.
- Schultheis, H., & Carlson, L. A. (2017). Mechanisms of reference frame selection in spatial term use: computational and empirical studies. *Cognitive Science*, 41(2), 276-325.
- Sekuler, A. B., & Swimmer, M. B. (2000). Interactions between symmetry and elongation in determining reference frames for object perception. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 54, 42.
- Shelton, A.L., & McNamara, T.P. (2001). Visual memories from nonvisual experiences. *Psychological Science*, 12, 343-347.
- Shelton, A. L., & McNamara, T. P. (2004). Spatial memory and perspective taking. *Memory and Cognition*, 32, 416-426.
- Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, 1, 260-273.
- Taylor, H. A., & Tversky, B. (1992). Descriptions and depictions of environments, *Memory and Cognition*, 20, 483-496.
- Tenbrink, T. (2007). *Space, time, and the use of language: An investigation of relationships*. Berlin: Mouton de Gruyter.
- Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407-433.

- Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, 110, 124-129.
- Tversky, B., Lee, P., & Mainwaring, S. D. (1999). Why do speakers mix perspectives? *Spatial Cognition and Computation*, 1, 399-412.
- Vinson, D. W., Dale, R., Tabatabaeian, M. & Duran, N. D. (2015). Seeing and believing: social influences on language processing. In R. Mishra, N. Srinivasan & F. Huettig (Eds.), *Attentional and Visual Processes in Language* (pp. 197-213). Springer.
- Weisberg, S. M., & Newcombe, N. S. (2014). A slippery directional slope: Individual differences in using slope as a directional cue. *Memory & Cognition*, 42, 648-661.
- Werner, S., & Schmidt, K. (1999). Environmental reference systems for large-scale spaces. *Spatial Cognition and Computation*, 1, 447-473.
- Wu, S., Barr, D. J., Gann, T. M., & Keysar, B. (2013). How culture influences perspective taking: differences in correction, not integration. *Frontiers in human neuroscience*, 7.

## Figure Captions

*Figure 1.* Examples of trials from Experiments 1a and 1b. The participant's perspective was always indicated by the word "You" at 0°, and the participant received verbal instructions to choose a folder from an imagined task partner, the speaker. In these examples, the speaker's perspective is at 90°, indicated by the word "Partner". These configurations would result in a critical trial when paired with the instruction "Give me the folder on the right" and in a control trial with the instruction "Give me the folder at the front." The configuration in Figure 1a is from Experiment 1a, where the folders were always aligned with the listener (ego-aligned), whereas the one in Figure 1b is from Experiment 1b, where the folders were always aligned with the speaker (other-aligned).

*Figure 2.* Proportion of egocentric and other-centric object selections on critical trials in Experiments 1a, 1b, and 2. Error bars indicate standard errors of the mean.

*Figure 3.* Mean proportion of egocentric responses across trial order for other-centric responders in Experiment 1a (top panel) and 2b (bottom panel).

*Figure 4.* Examples of trials from Experiment 2. The configuration in Figure 4a illustrates an ego-aligned trial, where the intrinsic axis of the triangular configuration is aligned with the listener's perspective, which is indicated by the word "You" at 0°. The speaker is depicted by the words "Partner" at 90°. This configuration constitutes a critical trial when accompanied by the instruction "Give me the folder that's in the front". It illustrates a control trial when accompanied by the instruction "Give me the folder on the



right”. Figure 4b illustrates an other-aligned configuration, with the triangular configuration aligned with the speaker’s perspective at  $90^\circ$ . Figure 4c illustrates a neither-aligned configuration, with the speaker at  $90^\circ$ ; the axis of symmetry of the right angle triangle formed by the objects is not aligned with either partner.

## Appendix A

Results of control trials for Experiments 1a-1b, and Experiment 2.

### Control trials of Experiments 1a and 1b.

We built omnibus statistical models using the same contrast coding scheme as for critical trials, except that for speaker Position the scheme involved a contrast between the 0° and 90° offsets (0° = -0.5, 90° = 0.5).

**Errors.** The overall variance captured by the omnibus model for errors (with both fixed and random effects) was 62%. The orientation of the folders did not influence the number of errors (i.e., selections of the folder that did not coincide with both the egocentric and other-centric choice) that listeners made across the two experiments ( $\beta = -.43$ ,  $SE = .44$ ,  $z = -.97$ ,  $p = .33$ ). As shown in Table 2, mixed responders made more errors than egocentric responders ( $\beta = -2.57$ ,  $SE = .64$ ,  $z = -4.00$ ,  $p < .001$ ), who in turn made more errors than other-centric responders ( $\beta = 1.45$ ,  $SE = .55$ ,  $z = 2.65$ ,  $p < .01$ ) on control trials.

As discussed in the main text, what was striking was that in both experiments mixed responders made significantly more errors on sagittal than lateral trials ( $\beta = 1.45$ ,  $SE = .23$ ,  $z = 6.38$ ,  $p < .001$ ). As shown in Table 2, mixed responders in Experiment 1b chose the incorrect folder on 46% of those sagittal trials, whereas egocentric and other-centric responders rarely made incorrect selections. In Experiment 1a, mixed responders also committed more errors on sagittal trials (21%) relative to the other responders. As we discussed in the main text, the high errors committed by mixed responders on sagittal control trials suggest that some of these individuals used the reverse mapping of the terms *front-back* onto space. Despite this numerical pattern, the relevant interaction term

(involving the interaction of the mixed vs. egocentric, Experiment 1b vs. 1a, and lateral vs. sagittal contrasts) was not significant ( $p = .13$ ).

**Response times.** The overall variance captured by the omnibus model for response times was 55%. The orientation of the folders did not predict the listeners' response times ( $\beta = -.03$ ,  $SE = .05$ ,  $t = -.63$ ,  $p = .53$ ), despite listeners being numerically faster on control trials in Experiment 1a than 1b. As with critical trials, across both experiments, egocentric responders were faster to respond ( $M = 1191$ ,  $SD = 683$  ms) than other-centric responders ( $M = 1425$ ,  $SD = 707$  ms;  $\beta = .19$ ,  $SE = .06$ ,  $t = 2.98$ ,  $p < .01$ ), but not reliably faster than mixed responders ( $M = 1372$ ,  $SD = 872$  ms;  $\beta = .01$ ,  $SE = .08$ ,  $t = .15$ ,  $p = .88$ ).

Echoing the pattern of errors described above, as shown in Table 3, mixed responders took numerically longer to respond in Experiment 1b ( $M = 1467$ ,  $SD = 1021$  ms) than in Experiment 1a ( $M = 1273$ ,  $SD = 670$  ms), relative to egocentric responders, although the relevant interaction term (involving the mixed vs. egocentric and Experiment 1b vs. 1a contrasts) did not predict response times ( $\beta = .16$ ,  $SE = .16$ ,  $t = .98$ ,  $p = .33$ ).

Listeners were slower when the speaker was at  $90^\circ$  than at  $0^\circ$  ( $\beta = .12$ ,  $SE = .01$ ,  $t = 7.96$ ,  $p < .001$ ). Although the type of instruction did not predict response times on its own ( $p = .89$ ), its interaction with the folder orientation did ( $\beta = -.07$ ,  $SE = .03$ ,  $t = -2.26$ ,  $p = .02$ ), with the difference between lateral and sagittal terms being greater in Experiment 1a than 1b.

As with critical trials, the response times of other-centric responders were influenced by the type of instruction more so than egocentric responders ( $\beta = .21$ ,  $SE =$

.04,  $t = 5.60$ ,  $p < .001$ ). Mixed responders were also somewhat influenced by instruction type but in the opposite direction ( $\beta = -.09$ ,  $SE = .05$ ,  $t = -1.90$ ,  $p = .06$ ), being slower on sagittal rather than lateral instructions. This could be indicative of the competition in reference frames that some mixed responders experience when interpreting sagittal terms.

**Total distance.** The overall variance captured by the omnibus model for total distance was 28%. The orientation of the folders did not significantly predict the mouse trajectories' total distance on control trials ( $\beta = -.02$ ,  $SE = .02$ ,  $t = -.65$ ,  $p = .51$ ). As suggested by the means reported in Table 3, similar to response times, egocentric responders made more efficient trajectories than other-centric responders ( $\beta = .09$ ,  $SE = .03$ ,  $t = 2.78$ ,  $p < .01$ ), but not more so than mixed responders ( $\beta = .03$ ,  $SE = .04$ ,  $t = .78$ ,  $p = .43$ ).

Relative to egocentric responders, other-centric responders made longer trajectories on lateral than sagittal instructions ( $\beta = .13$ ,  $SE = .03$ ,  $t = 4.65$ ,  $p < .001$ ) and when the speaker was at a  $90^\circ$  vs. at  $0^\circ$  offset ( $\beta = .10$ ,  $SE = .02$ ,  $t = 4.09$ ,  $p < .001$ ). This contextualizes the significant effect of the speaker position contrast ( $90^\circ$  vs. at  $0^\circ$ :  $\beta = .04$ ,  $SE = .01$ ,  $t = 4.22$ ,  $p < .001$ ), and its interaction with the instruction type contrast ( $\beta = .04$ ,  $SE = .02$ ,  $t = 2.16$ ,  $p = .03$ ): the decrement of performance on lateral instructions (a 7.6 pixel difference) emerged when the speaker was at  $90^\circ$  (at  $0^\circ$  there was a 2 pixel decrement on sagittal instructions).

**Directional shifts.** The overall variance captured by the omnibus model for total distance was 35%. The orientation of the folders did not reliably predict the number of directional shifts on the control trials of the two experiments (Experiment 1a:  $M = 1.64$ ,  $SD = 1.29$  and Experiment 1b:  $M = 1.77$ ,  $SD = 1.42$ ;  $\beta = .17$ ,  $SE = .12$ ,  $t = 1.38$ ,  $p = .17$ ).

In contrast to the response time and total distance measures, other-centric responders did not make reliably more directional shifts than egocentric responders on control trials ( $\beta = .01$ ,  $SE = .14$ ,  $t = .05$ ,  $p = .96$ ). Mixed responders did not differ significantly from egocentric responders either ( $\beta = .33$ ,  $SE = .19$ ,  $t = 1.77$ ,  $p = .08$ ).

Compared to egocentric responders, other-centric responders made more directional shifts on lateral than sagittal (or more accurately, fewer shifts on sagittal than lateral) instructions ( $\beta = .32$ ,  $SE = .11$ ,  $t = 2.75$ ,  $p < .01$ ). This pattern was more pronounced in Experiment 1b, with other-aligned folders, than in 1a, with ego-aligned folders, as reflected in a significant interaction of the relevant contrasts (involving Experiment, instruction type, and perspective preference: other-centric vs. egocentric:  $\beta = .60$ ,  $SE = .23$ ,  $t = 2.61$ ,  $p < .01$ ). This pattern was due to the fact that, whereas other-centric responders made comparable numbers of shifts on sagittal and lateral instructions in each experiment (in Experiment 1a: 1.71 shifts on average on both instruction types; and in Experiment 1b: 1.73 shifts on sagittal and 1.76 on lateral instructions), egocentric responders made more shifts on *sagittal* instructions, and this difference was greater in Experiment 1b (sagittal: 1.70 vs. lateral: 1.50) than in 1a (sagittal: 1.58 vs. lateral: 1.50). That is, in Experiment 1b, where the configural cue was in conflict with the egocentric responders' perspective, egocentric responders exhibited a relative difficulty in interpreting sagittal terms. The egocentric responders' behavior qualifies the finding that, contrary to the previous measures, on control trials listeners were overall less efficient (making more directional shifts) on sagittal trials than lateral trials ( $\beta = -.11$ ,  $SE = .05$ ,  $t = -2.31$ ,  $p = .02$ ).

## **Control trials of Experiment 2**

Again, we followed the same model building approach as for critical trials. However, to simplify the resulting omnibus models, we considered ego-reversed control trials (in which the apex of the triangular configuration pointed toward the listener) to be ego-aligned and, similarly, other-reversed control trials to be other-aligned. We did so because these pairs of trials (ego- and ego-reversed; other- and other-reversed) the axis of symmetry of the configuration was the same. This resulted in four types of configurations for control trials: ego-aligned, other-aligned, both-aligned, and neither-aligned.

Configuration alignment was coded in terms of planned contrasts that compared responses on ego-aligned vs. other-aligned configurations (ego-aligned = -0.5, other-aligned = 0.5), ego-aligned vs. neither-aligned configurations (ego-aligned = -0.5, neither-aligned = 0.5), and ego-aligned vs. both-aligned configurations (ego-aligned = -0.5, both-aligned = 0.5). The contrast coding for the remaining factors (perspective preference and type of instruction) was the same as for ambiguous trials (see section 6).

**Errors.** As we discussed in the main text, the present analyses involve the recoding of the terms *front* and *back*, since the initial assignment of the sagittal spatial terms resulted in extremely high errors of control trials. After this recoding, we established that listeners chose the incorrect folder on only 6% of the trials ( $SD = 24\%$ ). Error rates were higher for mixed responders, who selected the incorrect folder on 21% ( $SD = 41\%$ ) of control trials compared to the 4% ( $SD = 20\%$ ) error rate of both egocentric and other-centric responders. Indeed, relative to egocentric responders, mixed responders made significantly more errors ( $\beta = -2.96$ ,  $SE = .54$ ,  $z = -5.50$ ,  $p < .001$ )

It appears that, even after this recoding, some mixed responders used the mapping that was predominant in Experiments 1a and 1b. Mixed responders made 26% errors

( $SD= 44\%$ ) on sagittal instructions. They also made relatively high errors ( $M= 16\%$ ,  $SD= 37\%$ ) on lateral instructions as well, suggesting that at least some mixed responders may have exhibited confusion on lateral terms, making genuine errors, or else may have been responding haphazardly to both types of instructions (see Appendix B for more details about the mixed responders' behavior). Given the errors of mixed responders, there were overall more errors on sagittal than lateral instructions ( $\beta = .76$ ,  $SE = .21$ ,  $z= 3.61$ ,  $p < .001$ ). The overall variance captured by this omnibus model for errors was 44%.

**Response times.** The overall variance captured by the omnibus model for response times was 28%. Surprisingly, listeners took numerically the longest to respond on both-aligned configurations ( $M= 1785$ ,  $SD= 961$  ms) compared to the other configuration types (ego-aligned ones:  $M= 1759$ ,  $SD= 1077$  ms; other-aligned:  $M= 1726$ ,  $SD= 913$  ms; neither-aligned:  $M= 1533$ ,  $SD= 801$  ms). Response times on ego-aligned configurations differed significantly from all other types (ego-aligned vs. both aligned:  $\beta = .11$ ,  $SE = .03$ ,  $t= 3.90$ ,  $p < .001$ ; ego-aligned vs. other-aligned:  $\beta = .09$ ,  $SE = .03$ ,  $t= 2.70$ ,  $p < .01$ ; ego-aligned vs. neither aligned:  $\beta = -.14$ ,  $SE = .03$ ,  $t= -4.10$ ,  $p < .001$ ).

As with critical trials, egocentric responders were overall faster, taking on average 1470 ( $SD= 843$ ) ms to respond compared to the 1807 ( $SD= 965$ ) ms of other-centric responders and 1560 ( $SD= 939$ ) ms of mixed responders. The contrast between egocentric and other-centric responders significantly predicted response times ( $\beta = .25$ ,  $SE = .05$ ,  $t= 4.90$ ,  $p < .001$ ).

Overall, listeners were slower to respond to lateral instructions ( $M= 1816$ ,  $SD= 998$  ms) than sagittal ones ( $M= 1618$ ,  $SD= 892$  ms;  $\beta = .06$ ,  $SE = .02$ ,  $t= 3.10$ ,  $p < .01$ ). There was a significant interaction between perspective preference (other-centric vs.

egocentric) and type of instruction (lateral vs. sagittal) ( $\beta = .16$ ,  $SE = .05$ ,  $t = 3.40$ ,  $p < .001$ ), as the difference of the two instruction types was greater for other-centric responders (255 ms) than egocentric responders (77 ms). There was also a significant interaction between perspective preference (other-centric vs. egocentric) and type of configuration (ego-aligned vs. other-aligned) ( $\beta = -.17$ ,  $SE = .08$ ,  $t = -2.20$ ,  $p = .03$ ), as other-centric responders were faster on other-aligned configurations than ego-aligned ones (by 219.5 ms) to a greater extent than egocentric responders (by 19.60 ms).

**Divergence from optimal distance.** The overall variance captured by the omnibus model for divergence from optimal distance was 20%. On control trials, listeners produced trajectories that diverged from optimal distance on by 92 ( $SD = 101$ ) pixels on ego-aligned configurations, 93 ( $SD = 134$ ) pixels on both-aligned configurations, 131 ( $SD = 101$ ) pixels on other-aligned configurations, and 119 ( $SD = 128$ ) pixels on neither-aligned ones. The divergence of trajectories on ego-aligned configurations differed significantly from the rest (ego-aligned vs. both aligned:  $\beta = -.22$ ,  $SE = .06$ ,  $t = -4.10$ ,  $p < .001$ ; ego-aligned vs. other-aligned:  $\beta = .37$ ,  $SE = .07$ ,  $t = 5.48$ ,  $p < .001$ ; ego-aligned vs. neither-aligned:  $\beta = .14$ ,  $SE = .07$ ,  $t = 2.00$ ,  $p = .045$ ).

Consistent with response times, egocentric responders made trajectories that were less divergent from optimal distance ( $M = 88$ ,  $SD = 105$  pixels) than mixed ( $M = 109$ ,  $SD = 112$  pixels) and other-centric responders ( $M = 110$ ,  $SD = 118$  pixels), but neither of these differences were significant ( $p = .14$  and  $p = .09$ , respectively).

Consistent with response times, listeners made more divergent trajectories on lateral ( $M = 111$ ,  $SD = 116$  pixels) than sagittal ( $M = 100$ ,  $SD = 114$  pixels) instructions ( $\beta = .14$ ,  $SE = .04$ ,  $t = 3.67$ ,  $p < .001$ ). There was also a significant interaction between



instruction type (lateral vs. sagittal) and configuration type (neither vs. ego) ( $\beta = .52$ ,  $SE = .14$ ,  $t = 3.76$ ,  $p < .001$ ). This is qualified by the fact that, for lateral instructions, listeners were much less divergent on ego-aligned than on neither aligned configurations (ego-aligned:  $M = 85$ ,  $SD = 94$  pixels; neither-aligned:  $M = 143$ ,  $SD = 141$  pixels), whereas this wasn't the case for sagittal instructions (with the reverse pattern, in fact, being true: ego-aligned:  $M = 99$ ,  $SD = 106$  pixels; neither-aligned:  $M = 96$ ,  $SD = 109$  pixels).

**Directional shifts.** The overall variance captured by the omnibus model for directional shifts was 20%. Consistent with response times, listeners made more directional shifts on both-aligned configurations ( $M = 2.20$ ,  $SD = 1.64$ ) than the other types (ego-aligned:  $M = 1.86$ ,  $SD = 1.53$ ; other-aligned:  $M = 1.98$ ,  $SD = 1.29$ ; neither-aligned:  $M = 1.94$ ,  $SD = 1.46$ ). Only the contrast between both-aligned and ego-aligned configurations was significant in predicting directional shifts ( $\beta = .37$ ,  $SE = .10$ ,  $t = 3.75$ ,  $p < .001$ ). It's unclear why both-aligned trials appeared to incur a cognitive cost on control trials.

Again, egocentric responders made the most efficient trajectories, with fewer directional shifts ( $M = 1.72$ ,  $SD = 1.36$ ) than other-centric ( $M = 2.14$ ,  $SD = 1.56$ ) and mixed responders ( $M = 1.93$ ,  $SD = 1.45$ ); the difference between egocentric and other-centric responders was significant ( $\beta = .40$ ,  $SE = .16$ ,  $t = 2.51$ ,  $p = .01$ ).

Unlike response times and the divergence from optimal distance, the type of instruction did not predict directional shifts ( $p = .90$ ; lateral instructions:  $M = 2.05$ ,  $SD = 1.56$ ; sagittal instructions:  $M = 2.02$ ,  $SD = 1.49$ ). There was nevertheless an interaction between instruction type and configuration type (ego-aligned vs. neither-aligned), following the same pattern as divergence from optimal distance ( $\beta = .52$ ,  $SE = .22$ ,  $t = 2.31$ ,  $p = .02$ ): for lateral instructions, participants made fewer directional shifts on ego-

aligned than neither aligned configurations (ego-aligned:  $M= 1.73$ ,  $SD= 1.55$ ; neither-aligned:  $M= 2.06$ ;  $SD= 1.50$ ), whereas the reverse pattern was true for sagittal instructions (ego-aligned:  $M= 1.99$ ,  $SD= 1.50$ ; neither-aligned:  $M= 1.82$ ,  $SD= 1.42$ ). No other parameters in the model predicted the number of directional shifts.

## Appendix B

### Exploration of mixed responders' behavior and supplementary analyses that exclude mixed responders

#### Exploring the behavior of mixed responders

As we have noted in the main text, when reclassifying participants as egocentric, other-centric, or mixed based on their performance on lateral terms only, we observed shifts in category membership. We examine these shifts in category membership and explore the mixed responders' response patterns in more detail here.

**Experiments 1a and 1b.** It is notable that, whereas the number of mixed responders in Experiment 1b decreased when only lateral terms are considered, it increased in Experiment 1a. This is because in Experiment 1a, there were 14 other-centric responders (out of the 43 other-centric responders based on our original classification) who made an egocentric choice on 30% or more of lateral trials (range: 30-40%). These individuals were originally classified as other-centric due to the smaller proportion of egocentric choices they had made on sagittal trials, which had resulted in a mean of egocentric responses below 30%. When only lateral responses were considered, these individuals were now classified as mixed responders. At the same time, based on responses on lateral trials, 9 of the 17 mixed responders in Experiment 1a were reclassified as egocentric (and none as other-centric). This resulted in a total of 22 mixed responders (the 8 remaining, plus 14 new).

In contrast, in Experiment 1b, there was only one other-centric responder who made over 30% egocentric choices on lateral critical trials, thus switching group membership (to mixed) when only lateral trials were considered. Of the original 18 mixed

responders in Experiment 1b, 7 were reclassified as egocentric and 3 as other-centric, resulting in a total of 9 (8 remaining plus 1 new) mixed responders when only lateral instructions were considered.

These classification shifts suggest that, in Experiment 1a, where folders were ego-aligned, interpreting the terms *left-right* other-centrally may have been more challenging than in Experiment 1b, if we take the high proportions of egocentric choices (30-40%) that the 14 other-centric responders made in Experiment 1a to reflect errors. In contrast, in Experiment 1b where the folders were other-aligned, other-centric responders may have found lateral terms less confusable when making other-centric judgments: only one other-centric responder changed status to mixed based on their performance on lateral trials.

When seeking to identify those “true” mixed responders, who fell in the “mixed” response strategy range (i.e., 30-70% of egocentric choices) on *both* lateral and sagittal trials, we observed that they accounted for only a third of mixed responders (5 out of 17 mixed responders in Experiment 1a and 6 out of 18 mixed responders of Experiment 1b). To gain insight into what these mixed responders were doing, we coded their responses across trials to determine whether they were switching strategies or responding randomly. A “stretch of a strategy” was operationalized as observations of 3 or more consecutive critical trials of the same response type (egocentric or other-centric). This qualitative analysis revealed that, in Experiment 1a, only one mixed responder was responding randomly (no strategy stretches), 2 responders made a single strategy switch (from ego- to other-centric), 1 responder made multiple switches in strategy, and 1 responder adopted an other-centric strategy with occasional ego-centric interjections. In Experiment

1b, only one mixed responder responded randomly, 3 made multiple switches, and 2 made a single switch (again, from ego-centric to other-centric).

In sum: (i) in Experiments 1a and 1b some participants were categorized as mixed because they mapped sagittal trials differently than the rest; many of these responders were in fact egocentric, given their shift in category membership when only lateral trials were considered, (ii) when focusing on lateral trials only, some participants were categorized as mixed because, they seemed to find the terms *left-right* confusable (as indicated by their shift in category membership from other-centric to mixed), (iii) only 1/3 of mixed responders were “truly mixed” on both sagittal and lateral trials; among those, most made one or more strategy switches in the course of the experiment, (iv) only a very small number of participants (approximately 1%) appeared to be responding randomly.

**Experiment 2.** Following the same exploratory approach, we also reclassified responders in Experiment 2 based their object choices on lateral trials. When only lateral trials were considered, there were 125 other-centric, 43 egocentric, and 16 mixed responders. Of the 21 original mixed responders, 10 were reclassified as egocentric, suggesting that these were individuals who used consistently the reverse mapping than the rest on sagittal trials: these individuals made overwhelmingly egocentric responses on lateral trials and large proportions of “other-centric” (or reverse-mapped responses) on sagittal trials, resulting in a mixed classification when all trials were considered. In addition, one of the original mixed responders was reclassified as other-centric, and 10 remained as mixed. Of these 10 mixed responders, 7 were “truly mixed”, with egocentric responses within the 30-70% range on both sagittal and lateral trials. (The remaining

shifts in category membership involved 5 original other-centric responders and one original egocentric responder being reclassified as mixed, giving rise to the reported distribution of reclassified responders.)

As with Experiment 1, the mixed responders who were “truly mixed” on both sagittal and lateral trials exhibited a range of response patterns during the experiment: some made a single switch in strategy (2 from egocentric to other-centric, and 1 from other-centric to egocentric), two made multiple switches, and two did not exhibit any clear pattern in their responses (presumably responding randomly).

### **Models without mixed responders for Experiments 1a and 1b**

In light of the observations above, and in particular the point that when all trials were considered, mixed responders included individuals who had used the reverse mapping of sagittal trials than the rest, we examined models that excluded mixed responders. For each dependent measure, we examined separate omnibus models upon excluding the 17 mixed responders of Experiment 1a and the 18 mixed responders of Experiment 1b (with a total of  $N = 153$  participants remaining).

**Proportion of egocentric responses.** We built an omnibus model to examine how the orientation of the objects, the speaker’s position, instruction type, and their interactions predicted the proportion of egocentric responses that listeners made. The contrast coding scheme for the objects’ orientation (i.e. Experiment), the speaker’s position, and instruction type was the same as in the models reported in the main text (see section 3.2). The overall variance captured by this model was 84%.

The orientation of the objects, the instruction type, and the speaker’s position were each predictive of the proportion of egocentric responses in the omnibus model:

listeners made more egocentric responses in Experiment 1a than 1b ( $\beta = -2.59$ ,  $SE = .71$ ,  $z = -3.66$ ,  $p < .001$ ), when responding to lateral than sagittal instructions ( $\beta = 1.13$ ,  $SE = .14$ ,  $z = 7.85$ ,  $p < .001$ ),

Importantly, contrary to the model in the main text that included mixed responders, the interaction between folder orientation and the type of instruction was no longer predictive of perspective choice ( $\beta = -.34$ ,  $SE = .28$ ,  $z = -1.20$ ,  $p = .22$ ). As we discuss in section 4.1.1., the mixed responders' mapping of the sagittal terms seems to have been driving that interaction. The present model, without mixed responders, suggests that listeners were more likely to make an egocentric choice when the folders were oriented with themselves (than with the speaker), and when the instructions were lateral (than sagittal) independently of folder orientation.

Interestingly, without mixed responders, listeners were slightly more egocentric when the speaker was at  $90^\circ$  ( $M = 42.20\%$ ;  $SD = 49.40\%$ ) rather than at  $180^\circ$  ( $M = 41.45\%$ ;  $SD = 49.27\%$   $\beta = -.35$ ,  $SE = .14$ ,  $z = -2.45$ ,  $p = .01$ ).

**Response time, divergence from optimal distance, and directional shifts.** The omnibus models for these mouse-tracking measures also included perspective preference and its interaction with the other factors (object orientation, speaker position, and instruction type) as predictors. With the exclusion of mixed responders, perspective preference was now coded as a single planned contrast comparing the performance of egocentric and other-centric responders (egocentric = -0.5, other-centric = 0.5). The overall variance captured by each of the three omnibus models—for response times, total distance, and directional shifts—was 52%, 31%, and 29%, respectively.

Consistent with the results of the models that included all responders, the orientation of the folders was still not predictive of response times ( $p = .43$ ), the total distance of their trajectories ( $p = .06$ ), or the number of directional shifts ( $p = .59$ ).

Other-centric responders showed a decrement in performance relative to egocentric responders in all measures: they were slower to respond ( $\beta = .25$ ,  $SE = .05$ ,  $t = 4.74$ ,  $p < .001$ ), and made longer trajectories ( $\beta = .10$ ,  $SE = .03$ ,  $t = 3.66$ ,  $p < .001$ ), with more directional shifts ( $\beta = .38$ ,  $SE = .13$ ,  $t = 2.95$ ,  $p < .01$ ).

When the speaker at  $180^\circ$  speakers took longer to respond ( $\beta = .08$ ,  $SE = .01$ ,  $t = 5.77$ ,  $p < .001$ ) and made longer trajectories ( $\beta = .04$ ,  $SE = .01$ ,  $t = 4.14$ ,  $p < .001$ ), compared to when the speaker was at  $90^\circ$ . Listeners also took longer to respond ( $\beta = .08$ ,  $SE = .02$ ,  $t = 4.23$ ,  $p < .001$ ) and made longer trajectories ( $\beta = .04$ ,  $SE = .01$ ,  $t = 2.82$ ,  $p < .01$ ) on lateral instructions than sagittal ones.

Consistent with the main results, other-centric responders were influenced to a greater degree than egocentric responders by the speaker's position (for response times:  $\beta = .13$ ,  $SE = .02$ ,  $t = 4.61$ ,  $p < .001$ ; for total distance:  $\beta = .09$ ,  $SE = .02$ ,  $t = 4.16$ ,  $p < .001$ ; for directional shifts:  $\beta = .35$ ,  $SE = .11$ ,  $t = 3.26$ ,  $p < .01$ ) and by type of instruction (for response times:  $\beta = .39$ ,  $SE = .04$ ,  $t = 10.39$ ,  $p < .001$ ; for total distance:  $\beta = .09$ ,  $SE = .03$ ,  $t = 3.62$ ,  $p < .001$ ; for directional shifts:  $\beta = .49$ ,  $SE = .11$ ,  $t = 4.20$ ,  $p < .001$ ).

Again, consistent with the main results, the interaction of the three contrasts concerning perspective preference, folder orientation (Experiment), and instruction type was significant in predicting response times ( $\beta = .16$ ,  $SE = .07$ ,  $t = 2.13$ ,  $p = .03$ ), suggesting that other-centric responders were more sensitive to the type of instruction in Experiment 1a than 1b.



## Models without mixed responders for Experiments 2

We excluded the 21 mixed responders of Experiment 2 (with  $N = 163$  participants remaining), and examined separate omnibus models for each dependent measure.

**Proportion of egocentric responses.** For this model, the contrast coding scheme (concerning configuration type and instruction type) was the same as in the egocentric choice model reported in the main text.

Consistent with the results based on all responders, configuration type did not reliably influence their folder choice: listeners were no more likely to make an egocentric choice on ego-aligned configurations than on other-aligned configurations ( $p = .26$ ), or neither-aligned ones ( $p = .06$ ). They selected the egocentric choice on 24% ( $SD = 43\%$ ) of ego-aligned configurations, 24% ( $SD = 43\%$ ) of other-aligned configurations, and 26% ( $SD = 44\%$ ) of neither-aligned ones.

Consistent with the results in the main text, listeners were more likely to make an egocentric response on lateral ( $M = .27$ ,  $SD = .44$ ) than sagittal trials ( $M = .22$ ,  $SD = .42$ ;  $\beta = .65$ ,  $SE = .17$ ,  $z = 3.95$ ,  $p < .001$ ). Again, as with the main results, the difference between the two instruction types was greater on neither-aligned configurations (a 7% difference) compared to ego-aligned ones (a 3% difference), resulting in a significant interaction of those terms ( $\beta = 1.12$ ,  $SE = .34$ ,  $z = 3.23$ ,  $p = .001$ ). Similarly, the difference of the two instruction types was smaller on other-aligned configurations (a 0.7% difference) relative to ego-aligned ones ( $\beta = -.94$ ,  $SE = .39$ ,  $z = -2.43$ ,  $p = .02$ ). The overall variance captured by this omnibus model was 81%.

**Response time, divergence from optimal distance, and directional shifts.** For these measures, the omnibus models also included perspective preference and

its interaction with the other factors (configuration type and instruction type). Perspective preference was coded as a planned contrast comparing the performance of egocentric and other-centric responders (egocentric = -0.5, other-centric = 0.5). The overall variance captured by each of the three omnibus models—for response times, divergence from optimal distance, and directional shifts—was 37%, 23%, and 24%, respectively.

Consistent with the results of the models that included all responders, listeners took longer to respond ( $\beta = .06$ ,  $SE = .02$ ,  $t = 3.20$ ,  $p = .004$ ) and made more divergent trajectories ( $\beta = .11$ ,  $SE = .04$ ,  $t = 2.92$ ,  $p = .004$ ) on neither-aligned configurations than ego-aligned configurations. These two types of configurations did not differ in terms of directional shifts, as with the main results. Also, ego-aligned and other-aligned configurations did not differ in terms of any measure, consistent with the main results.

Again, other-centric responders experienced a greater processing cost than egocentric responders, which was reflected in their response times ( $\beta = .23$ ,  $SE = .04$ ,  $t = 5.30$ ,  $p < .001$ ), the distance of their trajectories ( $\beta = .36$ ,  $SE = .08$ ,  $t = 4.65$ ,  $p < .001$ ), and the number of directional shifts they made ( $\beta = .49$ ,  $SE = .13$ ,  $t = 3.67$ ,  $p < .001$ ).

And again, listeners experienced a greater processing cost when responding on lateral instructions than sagittal ones, as reflected in their response times ( $\beta = .11$ ,  $SE = .02$ ,  $t = 6.8$ ,  $p = .002$ ) and in the divergence of their trajectories from optimal distance ( $\beta = .09$ ,  $SE = .03$ ,  $t = 3.00$ ,  $p = .003$ ). As with the main results, directional shifts did not differ across the two instruction types, although other-centric responders (relative to egocentric ones) did make more directional shifts on lateral than sagittal instructions ( $\beta = .41$ ,  $SE = .12$ ,  $t = 3.61$ ,  $p < .001$ ). The increased sensitivity of other-centric responders to instruction type was evidenced in their response times as well ( $\beta = .17$ ,  $SE = .03$ ,  $t =$

5.20,  $p < .001$ ), but not in the divergence of their trajectories ( $p = .11$ ), in line with the main results (see Table 5).

Again, consistent with the main results, the two-way interaction between perspective preference and configuration type (neither-aligned vs. ego-aligned) was predictive of the mouse-tracking measures, suggesting that the decrement in performance for other-centric responders was more pronounced on neither-aligned than ego-aligned configurations (for response times:  $\beta = .16$ ,  $SE = .03$ ,  $t = 4.70$ ,  $p = .01$ ; for distance:  $\beta = .20$ ,  $SE = .08$ ,  $t = 2.54$ ,  $p = .01$ ; for directional shifts:  $\beta = .57$ ,  $SE = .13$ ,  $t = 4.37$ ,  $p < .001$ ). Moreover, the three-way interaction with instruction type suggested that this decrement held predominately for lateral instructions (for response times:  $\beta = .33$ ,  $SE = .07$ ,  $t = 4.70$ ,  $p < .001$ ; for distance:  $\beta = .53$ ,  $SE = .16$ ,  $t = 3.42$ ,  $p < .001$ ; for directional shifts:  $\beta = .54$ ,  $SE = .26$ ,  $t = 2.07$ ,  $p = .03$ ).

Finally, the results in our main text suggested that the decrement in the performance of other-centric responders (relative to egocentric responders) was less pronounced on other-aligned configurations than ego-aligned ones. In the analyses here, without mixed responders, this two-way interaction significantly predicted response times ( $\beta = -.13$ ,  $SE = .04$ ,  $t = -3.40$ ,  $p < .001$ ) and directional shifts ( $\beta = -.51$ ,  $SE = .15$ ,  $t = -3.49$ ,  $p < .001$ ), but not divergence from optimal distance ( $p = .08$ ).

Overall, our interpretation of the results reported in the main text does not change substantially when mixed responders are excluded. The findings without mixed responders, here, converge with those based on all responders.

## Appendix C

Follow-up experiments conducted to examine how configural cues were perceived

### Follow-up to Experiment 1a and 1b

We conducted a follow-up to Experiments 1a and 1b to establish whether individuals perceived the alignment of the folders with each task partner consistently with our classification of folders as ego-aligned and other-aligned. This undertaking was motivated by the fact, in our experiments, the elongated axis of the folder was taken to reflect its width (i.e., it was aligned with the folder's left/right dimension), whereas typically the elongated internal axis of objects is a cue for assigning the objects' "upright" orientation (e.g., Quinlan & Humphreys, 1993; Sekuler & Swimmer, 2000). We therefore aimed to establish whether, despite the orientation of the elongated axis of the objects in our materials, the alignment of the objects relative to the two task partners was apprehended as intended (i.e., as ego-aligned in Experiment 1a and as other-aligned in 1b).

**Participants and Materials.** 104 participants were recruited from Amazon Mechanical Turk. The task was available to participants who were over 18 years old and located in the United States. They were paid \$.60 for approximately 4 minutes of effort. Participants were presented with all 20 unique items from Experiments 1a and 1b corresponding to the 20 unique visual configurations of the two experiments, taken together. This set included 8 ego-aligned configurations (according to our classification) from Experiment 1a, 8 other-aligned configurations from Experiment 1b, and 4 both-aligned configurations, which appeared in both experiments and were a subset of the control trials. For each set of ego-aligned and other-aligned configurations, 2 depicted

the partner at 90° and 2 at 270° (with the objects in diagonal arrangement), and 4 depicted the partner at 180° (2 of these with the objects in a diagonal arrangement, 1 in a horizontal and 1 in a vertical arrangement). The 4 both-aligned configurations all depicted the partner at 0° (2 had objects in a diagonal arrangement, 1 with a horizontal, and 1 with a vertical arrangement). These unique configurations can be viewed at our OSF and GitHub repositories for the project. (Note that the 40 trials of Experiments 1a and 1b were generated by pairing these configurations with the different types of instructions.)

**Procedure.** In the instructions, participants were informed that they would see screenshots in which they and a partner would be depicted around a table “with the words YOU and PARTNER”. For each screenshot participants were asked to make a judgment about the orientation of the objects on the table by responding to the question: “Are the folders aligned with: You, The Partner, Both You and the Partner, Neither You and the Partner”. The four response options were presented as a radio button list. All 20 question items were vertically arranged on the same page, and were presented in a randomized order for each participant.

**Results and Discussion.** An accurate response was taken to be a classification congruent with our assignment of the configuration as ego-aligned, other-aligned, or both-aligned. We aggregated participants’ classifications by item, producing a data set of 20 cases that were independently and randomly sampled across participants.

The mean proportion of accurate responses was .47 ( $SD = .11$ ; range: .36–.83 ), which was significantly higher than chance (.25);  $t(19) = 9.07, p < .001$ . Participants were

most accurate in classifying both-aligned configurations ( $M = .62$ ,  $SD = .13$ ) followed by other-aligned ( $M = .47$ ,  $SD = .07$ ) and ego-aligned ones ( $M = .40$ ,  $SD = .03$ ).

Errors in classifying the ego-aligned and other-aligned configurations stemmed from interference from the “both-aligned” and “neither-aligned” options. On ego-aligned configurations, participants were more likely to make a classification error by making a “both-aligned” or “neither-aligned” selection than an “other-aligned” selection ( $t(7) = 2.70$ ,  $p = .03$ , and  $t(7) = 4.79$ ,  $p = .002$ , respectively). Similarly, on other-aligned configurations, participants were more likely to make a classification error by making a “both-aligned” or “neither-aligned” selection than an “ego-aligned” selection ( $t(7) = 2.70$ ,  $p = .03$ , and  $t(7) = 2.67$ ,  $p = .03$ , respectively).

To further explore what predicted participants’ classifications, we used an ordinary least squares (OLS) regression approach. The classification of a configuration by participants as “other-aligned” was significantly predicted by the configuration actually being other-aligned (by our classification) compared to it being ego-aligned ( $\beta = -34.87$ ,  $SE = 3.36$ ,  $t = -10.39$ ,  $p < .001$ ) or both-aligned ( $\beta = -41.75$ ,  $SE = 4.11$ ,  $t = -10.15$ ,  $p < .001$ ). Similarly, the classification of a configuration as “ego-aligned” was significantly predicted by the configuration actually being ego-aligned than being other-aligned ( $\beta = -36.75$ ,  $SE = 2.94$ ,  $t = -12.51$ ,  $p < .001$ ) or both-aligned ( $\beta = -30.00$ ,  $SE = 2.40$ ,  $t = -12.51$ ,  $p < .0001$ ). The classification of a configuration by participants as “both-aligned” was significantly predicted by the configuration actually being both-aligned relative to ego-aligned ( $\beta = -36.00$ ,  $SE = 6.96$ ,  $t = -5.17$ ,  $p < .001$ ) or and other-aligned ( $\beta = -40.75$ ,  $SE = 6.96$ ,  $t = -5.85$ ,  $p < .001$ ). And finally, the classification of folders as “neither-aligned”—a category not applicable to this set of materials—was not

significantly predicted by folders being ego-aligned or other-aligned relative to both-aligned (both  $p = .10$ ).

These results suggest that the folders used in Experiments 1a and 1b provided a directional cue consistent with our classification of their alignment relative to task partners. Participants in this follow-up study were relatively accurate (more so than chance) at classifying the configurations from Experiment 1a and 1b. Their errors stemmed from the tendency to classify ego-aligned and other-aligned configurations as “both-aligned” or “neither-aligned”. Importantly, the interference of the competing person-aligned option (i.e., of the “other-aligned” option during the classification of an ego-aligned item, and vice-versa) was smaller than the interference of the other two options.

### **Follow-up to Experiment 2**

We conducted a follow-up to Experiments 2 to examine whether individuals perceive the alignment of the 3-object configurations consistently with our classification of configurations as ego-aligned, other-aligned, and neither-aligned. Although the mouse trajectories of other-centric responders in Experiment 2 differentiated between the three types of configurations that were the focus of critical trials (i.e., showing that other-centric responders experienced an increased processing cost on neither-aligned configurations and a reduced processing cost on other-aligned configurations), we considered the possibility that the three objects may not have been perceived as a triangle that afforded a directional cue in the way we had intended.

**Participants and Materials.** 95 participants were recruited from Amazon Mechanical Turk. The task was available to participants who were over 18 years old and located in the United States. They were paid \$.60 for approximately 4 minutes of effort. Participants were presented with all 23 unique items from Experiment 2.

This set included 6 ego-aligned configurations (2 of which were ego-reversed, with the apex of the triangle pointing towards “You” , rather than away from “You”), 6 other-aligned configurations (2 of which were other-reversed, with the apex of the triangle pointing toward the “Partner”, rather than away from the “Partner”), 9 neither-aligned configurations, and 2 both-aligned configurations. For each set of ego-aligned and other-aligned configurations, 3 depicted the partner at 90° and 3 at 270°. Of the 9 neither-aligned configurations, 3 depicted the partner at 90°, 3 at 270°, and 3 at 180°. The 2 both-aligned configurations depicted the partner at 0°. These configurations can also be found at our OSF and GitHub repositories.

**Procedure.** In the instructions, participants were informed that they would see screenshots in which they and a partner would be depicted around a table “with the words YOU and PARTNER”. For each screenshot, they were asked to make a judgment about how the objects were oriented on the table by responding to the question: “Is the arrangement of the three objects aligned with: You, The Partner, Both You and the Partner, Neither You and the Partner”. The four response options were presented as a radio button list. All 23 question items were vertically arranged on the same page, and were presented in a randomized order for each participant.

**Results and Discussion.** An accurate response was taken to be a classification consistent with our assignment of configurations as ego-aligned, other-aligned, neither-



aligned, or both-aligned. Participants' responses were aggregated by item, producing a dataset of 23 cases that were independently and randomly sampled across participants.

The mean proportion of accurate classifications was only .24 ( $SD = .14$  range: .09–.56), and it was not significantly different than chance (.25);  $t(22) = -.20, p = .85$ . Participants were numerically most accurate in classifying both-aligned configurations ( $M = .56, SD = .001$ ) followed by neither-aligned ( $M = .25, SD = .10$ ), other-aligned ( $M = .22, SD = .11$ ) and ego-aligned configurations ( $M = .15, SD = .08$ ).

An OLS regression approach provided some insight into what contributed to this low classification accuracy. Making “ego-aligned” and “other-aligned” classifications involved interference from other-aligned and ego-aligned configurations, respectively. The classification of a configuration as “ego-aligned” was significantly predicted by the configuration actually being other-aligned over ego-aligned ( $\beta = 21.67, SE = 9.60, t = 2.21, p = .04$ ), and not by the other two types of configurations (ego- vs. both-aligned:  $p = .36$ ; ego vs. neither-aligned:  $p = .44$ ). Similarly, the classification of a configuration by participants as “other-aligned” was significantly predicted by the configuration actually being ego-aligned over being ego-aligned ( $\beta = 23.33, SE = 10.49, t = 2.22, p = .04$ ), but not by the other two types of configurations (ego- vs. both-aligned:  $p = .36$ ; ego vs. neither-aligned:  $p = .44$ ). The classification of configurations as “neither-aligned” was significantly predicted by configurations actually being neither-aligned relative to being ego-aligned ( $\beta = -11.28, SE = 4.13, t = -2.73, p = .01$ ), though not relative to partner-aligned configurations ( $\beta = -6.94, SE = 4.13, t = -1.68, p = .11$ ). Moreover, the classification of configurations as “neither-aligned” was predicted by both-aligned configurations (relative to neither-aligned configurations,  $\beta = 14.06, SE = 6.12, t$

= 2.30;  $p = .03$ ), suggesting that there was interference from both-aligned configurations during “neither aligned” classifications. Finally, as suggested by the relatively higher accuracy of classifying both-aligned configurations, the classification of configurations as “both-aligned” was significantly predicted by configurations actually being both-aligned relative to them being ego-aligned ( $\beta = -29.33$ ,  $SE = 13.08$ ,  $t = -2.24$ ,  $p = .04$ ), other-aligned ( $\beta = -31.17$ ,  $SE = 13.08$ ,  $t = -2.38$ ;  $p = .03$ ) or neither-aligned ( $\beta = -21.94$ ,  $SE = 12.52$ ,  $t = -1.75$ ;  $p = .10$ ).

These results—and in particular the finding that “ego-aligned” and other-aligned” judgments indicated interference from the competing type of configuration—may suggest that participants perceived ego-aligned configurations as other-aligned, and vice-versa. This is possible if participants were using the orientation of the *base* of the triangle (as opposed to its perpendicular vertex) to make a judgment about the triangle’s alignment.

To explore this possibility, we recoded ego-aligned configurations as other-aligned, and vice-versa, and recomputed classification accuracy. The mean proportion of accurate responses was now .37 ( $SD = .20$  range: .13–.67), and it was significantly higher than chance (.25);  $t(22) = 2.88$ ,  $p = .009$ . Participants were now 47% accurate ( $SD = 24\%$ ) at classifying other-aligned configurations and 38% accurate ( $SD = 20\%$ ) at classifying ego-aligned ones—an increase of 25% and 23%, respectively, from our original classification scheme. This interpretation—that participants were making judgments by considering the base of the triangle—also contextualizes the previously reported finding that both-aligned configurations were frequently classified as “neither-aligned”, since in the 2 unique both-aligned configurations of the materials the base of

the triangle was horizontal and perpendicular to (i.e., not aligned with) the perspective of the task partners located at 0°.

Thus, participants in this follow-up experiment seem to use the base of the triangle, rather than its vertex, to make explicit judgments about the configuration's alignment. As we discuss in the main text, if listeners in Experiment 2 also perceived configural alignment in terms of the base of the triangle, this has implications for how we interpret the performance of other-centric responders: it would mean that the decrement of other-centric responders (relative to egocentric responders) was reduced on ego-aligned compared to other-aligned configurations (rather than the reverse).

Despite this issue, the results of this follow-up experiment suggest that people do appear to perceive these 3-object configurations as triangles. We have also considered and have ruled out the possibility that participants perceived the three objects as different shapes, such as an L-shape<sup>6</sup>.

---

<sup>6</sup> To examine whether participants perceived configurations as L-shapes, each of the 23 items was reclassified according to the orientation of an L-shape: specifically, according to the alignment of the vertical axis of the L relative to the perspective of the two task partners. Under this classification scheme, most configurations were neither-aligned ( $N=16$ , as in Figures 4a and 4b), 3 were ego-aligned, 3 were other-aligned (as in Figure 4c), and only 1 was both-aligned. The participants' accuracy under this classification scheme was low ( $M = .22$ ,  $SD = .15$ ), and not different from performance at chance,  $t(22) = -1.14$ ,  $p = .27$ .