COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12920

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12920

Space in Hand and Mind: Gesture and Spatial Frames of Reference in Bilingual Mexico

Tyler Marghetis,[a,b] Melanie McComsey,[c] Kensy Cooperrider[d]

[a]*Cognitive & Information Sciences, University of California Merced*
[b]*Santa Fe Institute*
[c]*Department of Anthropology, University of California San Diego*
[d]*Department of Psychology, University of Chicago*

Received 14 April 2020; received in revised form 23 September 2020; accepted 13 October 2020

Abstract

Speakers of many languages prefer allocentric frames of reference (FoRs) when talking about small-scale space, using words like "east" or "downhill." Ethnographic work has suggested that this preference is also reflected in how such speakers gesture. Here, we investigate this possibility with a field experiment in Juchitán, Mexico. In Juchitán, a preferentially allocentric language (Isthmus Zapotec) coexists with a preferentially egocentric one (Spanish). Using a novel task, we elicited spontaneous co-speech gestures about small-scale motion events (e.g., toppling blocks) in Zapotec-dominant speakers and in balanced Zapotec-Spanish bilinguals. Consistent with prior claims, speakers' spontaneous gestures reliably reflected either an egocentric or allocentric FoR. The use of the egocentric FoR was predicted—not by speakers' dominant language or the language they used in the task—but by mastery of words for "right" and "left," as well as by properties of the event they were describing. Additionally, use of the egocentric FoR in gesture predicted its use in a separate nonlinguistic memory task, suggesting a cohesive cognitive style. Our results show that the use of spatial FoRs in gesture is pervasive, systematic, and shaped by several factors. Spatial gestures, like other forms of spatial conceptualization, are thus best understood within broader ecologies of communication and cognition.

Keywords: Gesture; Space; Frames of reference; Cognitive diversity; Cognitive ecology

1. Introduction

Thinking and talking about location and movement involves adopting a particular frame of reference (FoR). Imagine you had experienced the following. *You are at a*

Correspondence should be sent to Tyler Marghetis, Cognitive & Information Sciences, University of California Merced, Merced, CA. E-mail: tyler.marghetis@gmail.com

Tyler Marghetis, Melanie McComsey, and Kensy Cooperrider contributed equally to this work.

wedding, and it's time to cut the cake. A 5-foot-tall cake is wheeled out before a crowd of amazed guests. As the cart comes to a halt, the cake totters, sways, and then topples. When later recounting this event, you could describe how the cake "toppled to the right," had it toppled in that direction from your vantage point. If you described it in this way, you would be construing the scene using an *egocentric* FoR, one in which spatial relations are determined relative to one's body. Alternatively, you could have construed the same scene in terms of cardinal directions (e.g., "toppled to the south") or geographic features (e.g., "toward the mountain"), thus using an *allocentric* FoR, in which relations are determined relative to the environment. When we call up and communicate a memory like this, we may not reflect on the fact that we are also settling on a particular coordinate system. But we are. Such coordinate systems are foundational to spatial thinking, forming part of the invisible infrastructure of cognition (Levinson, 2003).

Describing a cake as "toppling to the south" may seem like an unusual choice to many readers, but in some cultures it would be perfectly natural. Speakers of many languages preferentially describe even small-scale spatial relations using allocentric coordinates such as cardinal directions or topographic features (e.g., "the fork is east/downhill of the plate;" Levinson, 2003; Levinson & Wilkins, 2006; Majid, Bowerman, Kita, Haun, & Levinson, 2004). Allocentric-favoring communities have now been documented worldwide. These include speakers of Hai||om in Namibia (Haun, Rapold, Janzen, & Levinson, 2011); Yupno in Papua New Guinea (Cooperrider, Slotta, & Núñez, 2017); Isthmus Zapotec in Mexico (Moore, 2018; Pérez Báez, 2011); Arrernte, Guugu Yimithirr, and Warwa in Australia (Majid et al., 2004); and many others (Levinson & Wilkins, 2006). Thus, as foundational as spatial FoRs are to everyday communication and cognition, how they are used varies from one community to the next.

Early reports of cross-cultural variation in FoR preferences claimed that speakers in allocentric-favoring groups do not just *talk* about space allocentrically, but also *gesture* allocentrically (Haviland, 1993; Levinson, 2003; Majid et al., 2004). What would this look like? Returning to the cake scenario, rather than describe the cake as toppling "to the right," you might say simply that the cake "toppled," and supply directional information in gesture by sweeping a hand to the right (if using an egocentric FoR) or to the south (if using an allocentric FoR). Ethnographic work has furnished vivid examples of this phenomenon. Haviland (1993) described how a Guugu Yimithirr storyteller reproduced the correct cardinal orientations of events in his gestures (e.g., a boat capsizing) across two retellings that occurred 2 years apart and in different settings. Similar cases of apparently allocentric gesturing have also been observed among Tzeltal speakers (Levinson, 2003). Naturalistic observations like these have been marshaled to support a view in which FoR preferences are not just a matter of superficial linguistic conventions but of deep-seated cognitive styles (Levinson, 2003; Majid et al., 2004). Yet, despite the interest and importance of these early qualitative observations, little work since has examined spatial FoRs in gesture more systematically. This is the broad goal of the present study.

At least three different ideas can be discerned in early reports of FoR use in gesture. The first is that speakers' gestures reliably reflect a particular FoR. Prior work has already demonstrated this phenomenon compellingly for large-scale space.[1] When speakers of

Yucatec Maya were asked explicitly to describe where one familiar landmark was in relation to another (i.e., a gas station in relation to a store), they used cardinally oriented gestures to convey the relationship (Le Guen, 2011). A study with another community in rural Mexico, using the same style of task, also reported high rates of allocentric gesture (Calderón, De Pascale, & Adamou, 2019). Whether members of other groups would gesture any differently is unknown, as even speakers of preferentially egocentric groups use cardinal directions for describing large-scale space (Levinson, 2003). It would thus be more striking to demonstrate the use of an allocentric FoR when gesturing about *small-scale* or "table-top" space, a context in which speakers of preferentially egocentric languages avoid allocentric descriptions. More striking still would be a demonstration of consistent FoR use in gesture when spatial relationships are communicatively backgrounded, rather than asked about explicitly. A fruitful testbed for a more striking demonstration along these lines is small-scale events that involve motion. People often gesture to convey the direction of motion in recently seen events (Alibali, 2005; Kita & Özyürek, 2003; McCullough, 2005). But they do not often label the direction of motion verbally (e.g., "to the right"; Kita & Özyürek, 2003), presumably because that information is communicatively backgrounded. Here, we ask: Would people reliably use a particular FoR in gesture, either egocentric or allocentric,[2] when describing spatial relations? And would they do so even for relations that are small scale and communicatively backgrounded?

A second idea found in early reports is that FoR preferences in gesture are shaped by language (Majid et al., 2004). There are several possible versions of this proposal. FoR use in gesture could be shaped by how space is described in one's *dominant* language (Haviland, 1993; Levinson, 2003); by the language one is using *in the moment* (e.g., "thinking-for-speaking," Slobin, 1996); or by the long-term effects of mastering specific *lexical items* that refer to spatial relations such as like "left" and "right" (e.g., Gentner, Özyürek, Gurcanli, & Goldin-Meadow, 2013). In monolingual populations, these three versions are hard to distinguish, but they may dissociate in bilingual populations. For example, some bilinguals remain dominant in one of their languages, while others—balanced bilinguals—flexibly use different languages in different settings (e.g., Grosjean, 2010). Of course, nonlinguistic factors might also matter for the choice of FoR in gesture (Li & Abarbanell, 2019; Shapero, 2017). Indeed, Haviland (1993) noted that not all Guugu Yimithirr gestures were oriented allocentrically, raising the question of which specific nonlinguistic factors might drive the use of one FoR over another. Here, we focus on nonlinguistic factors that relate to how an event is experienced. For example, people may be more likely to adopt an egocentric perspective on an event if they are an active participant rather than a passive observer, or if the event involves motion along their sagittal (away–toward) axis, which is more strongly asymmetric than the lateral (left–right) axis (Clark, 1973; Shusterman & Li, 2016; Tversky, 2011). If people do use a particular FoR in gesture, which factors—linguistic or nonlinguistic—predict the FoR they will use?

A final idea from early reports is that a speaker's preferred FoR in gesture reflects a cohesive preference or "cognitive style" that is evident in their spatial thinking and communicating more broadly. Some evidence for this proposal has come from array-reconstruction tasks. In these tasks, a participant is familiarized with a static array, usually of toy figurines; the array is then removed and the participant is asked to reconstruct it in another location, rotating in the process (e.g., Levinson, 2003; Li & Gleitman, 2002; Pederson et al., 1998). FoR preferences on such nonverbal tasks mirror preferences in verbal description: People in communities that talk about space allocentrically tend to reconstruct arrays allocentrically, and people in communities that talk about space egocentrically tend to reconstruct arrays egocentrically. If each person has a cognitive style that coheres across different facets of their spatial thinking and communicating, then the FoR they use in gesture should similarly align with the FoR they use on array-reconstruction tasks. Of course, there are reasons the two forms of behavior may not align. Spontaneous gesture is communicative, while array-reconstruction is not. Moreover, gesture is often characterized as "unwitting" (McNeill, 1992) and automatic (Hostetter & Alibali, 2008)—thus not particularly strategic. In contrast, array-reconstruction tasks have been criticized for inviting strategic responding, as participants are given the ambiguous instruction to make the array the "same" (Li, Abarbanell, Gleitman, & Papafragou, 2011). Does FoR use in gesture reflect a cohesive cognitive style also evident in noncommunicative spatial reasoning?

The present study sought to examine FoRs in hand and mind more systematically—and, in particular, to address the three ideas discernible in earlier ethnographic reports. To address the first idea—that people reliably use a particular FoR in gesture—we developed a novel task, "Toppling Blocks." The task was designed to elicit gestures about small-scale events involving directional motion, a context in which FoR information is communicatively backgrounded. To address the second idea—that linguistic factors predict FoR use in gesture—we identified a bilingual field site in which preferences for one or the other FoR were expected to vary. The study was conducted in the city of Juchitán de Zaragoza (pop. 98,043, as of 2015), in Oaxaca, Mexico, a community in which Isthmus Zapotec and Spanish, languages from different families (Otomanguean vs. Romance), are widely used. Speakers of Isthmus Zapotec have been shown to rely preferentially on cardinal direction terms, even for describing small-scale space (Moore, 2018; Pérez Báez, 2011; e.g., *guxiú que la, nuuni neza guiá de xiga que*, "the knife is to the north of the bowl"). Speakers of Mexican Spanish, like American speakers of English, have been shown to favor egocentric descriptions of small-scale space (Bohnemeyer et al., 2015; e.g., *el cuchillo está a la derecha de la jícara*, "the knife is to the right of the bowl"). To assess different versions of the idea that language shapes FoR use, we did three things: first, recruited participants from this bilingual community who differed in their habitual use of and proficiency in Spanish (Zapotec Dominants and Balanced Bilinguals); second, manipulated the language in which the task was conducted (with bilinguals tested twice in separate sessions); and, third, assessed individual differences in the mastery of terms denoting egocentric and allocentric relations. We also examined nonlinguistic factors by manipulating how participants experienced the event. Finally, to address the third idea—that gesture reflects a cohesive cognitive style—we tested the same participants on an array-reconstruction task.

2. Methods

As part of a multiyear ethnography of local spatial practices (McComsey, 2015), experimental data were collected during 48 sessions throughout 2012 and 2013. Where indicated, additional details are available in the Supplementary Materials on the Open Science Framework (https://osf.io/agkvh/).

2.1. Participants

Participants were adult residents of Juchitán de Zaragoza, a majority-indigenous, multilingual municipality that has attracted much recent research into bilingualism, culture, and cognition (De Korne, 2016; McComsey, 2015; Moore, 2018). All were native speakers of Isthmus Zapotec (hereafter "Zapotec" for simplicity). Half relied on Zapotec almost exclusively in daily life and had limited proficiency in Spanish (Zapotec Dominant; n = 16); the rest relied on both Spanish and Zapotec in daily life and were thus highly proficient in both (Balanced Bilingual; n = 16). These two language groups were formed during initial participant recruitment based on a holistic, multipart assessment of language proficiency. This assessment included several parts: experimenter evaluation of prompted Spanish narratives (e.g., "Tell me about something good that happened today"); self-reports of language proficiency (e.g., "Which is your preferred language?") and language attitudes (e.g., "Is Zapotec beautiful?"); and self-reports of language use across different settings (e.g., at church) and with different interlocutors (e.g., children). The two language groups formed on the basis of this assessment did not differ significantly in demographic factors such as gender (p = .22, Fisher's exact test) or age (t_{30} = 1.22, p = .23). Sample sizes were determined on the basis of the availability of Zapotec Dominants at the field site. See the Supplementary Methods and Analyses document for full participant demographics, and for further information about the assessment of language dominance.

2.2. Design

Zapotec Dominant participants completed assessments of lexical competence, spontaneous gesture, and spatial reconstruction, in a single session conducted entirely in Zapotec. Balanced Bilingual participants completed these assessments twice—once in Spanish, once in Zapotec—in sessions separated by at least 7 weeks, with session order counterbalanced. Zapotec sessions were administered by a native speaker of Zapotec; Spanish sessions, by one of the authors (M.M.) who has native-like fluency in Spanish; both lived in the community. Sessions were conducted in the semi-outdoor patios that are a traditional feature of local homes and a common site for social interaction.

2.3. Materials and procedure

2.3.1. Lexical competence

We assessed participants' comprehension of terms denoting egocentric spatial relations (left-of, right-of) and allocentric spatial relations (north-of, south-of, east-of, west-of), which we refer to as "lexical competence." Participants were seated at a table, facing in a cardinal direction. Two reference objects (inverted buckets) were each surrounded by (but not touching) four target objects (toy animals or wooden blocks). One bucket was "near," immediately in front of the participant on the table. The other bucket was "far," placed 3 m away on the floor. On the 12 critical trials, the experimenter described a target object's location relative to a reference object using an allocentric (e.g., "to the north of the far bucket") or egocentric description (e.g., "to the left of the near bucket"), and participants had to identify the target. All six spatial terms were tested once in the initial setup, and then tested again after the entire setup was rotated 180° so the participant was facing in the opposite direction. In the initial setup, two cardinal direction terms were used for near objects and two used for far objects; after rotation, cardinal terms that had been used for near objects were used for far objects, and vice versa. These critical questions were interspersed with five filler questions about color terms and two questions about direct reference to body parts ("Show me your left/right hand").

We tested spatial terms from the language used for the session. For egocentric relations in Spanish, we tested the terms *lado izquierdo* ("left side") and *lado derecho* ("right side"). The variety of Isthmus Zapotec spoken in Juchitán has a native term for "left," *biga'*, and borrows the Spanish term for "right" as *derechu* (McComsey, 2015, p. 146). For allocentric relations, in Spanish we tested the cardinal terms *norte* ("north"), *sur* ("south"), *este* ("east"), and *oeste* ("west"), and in Zapotec we tested *guiá'* ("north"), *guete'* ("south"), *ladu rindani gubidxa* ("east," lit. "the side the sun is born"), and *ladu riaazi gubidxa* ("west," lit. "the side the sun sets"). Among Balanced Bilingual participants, lexical competence was significantly correlated across sessions (ego: Kendall's $\tau = 0.50$, $p = .03$; allo: $\tau = 0.46$, $p = .03$), and we measured their overall competence with these lexical concepts using mean accuracy across sessions.[3]

2.3.2. Spontaneous gestures representing small-scale space ("Toppling Blocks")

To elicit spontaneous gestures during descriptions of small-scale space, we developed a task in which participants observe a complex motion event involving wooden blocks of varying shapes and sizes—for example, sequentially toppling "dominos"—and then, after a delay and rotation, describe the event (Fig. 1).

The setup consisted of a square table and, on the other side of an occluding sheet, two chairs facing away from the sheet and table. Each trial began with participants viewing a complex motion event in an array of wooden blocks (cubes, cylinders, etc.), constructed in the center of the table. Events were designed to involve one salient direction of motion (e.g., a cylinder rolling along a path). See OSF Supplementary Materials for images (Fig. S1 in the Supplementary Methods and Analyses) and videos of all stimuli. The blocks were covered immediately after the motion event. After a 30 s delay, participants moved to the other side of the occluding sheet and sat with their back to the table, rotating 90° in the process. They were asked to describe "what you saw" by an experimenter who was seated to their right at 45°. Gesture was not mentioned.

Each participant completed 12 trials, consisting of the same 12 motion events. The axis of motion (away–toward, left–right) was counterbalanced between subjects; direction of

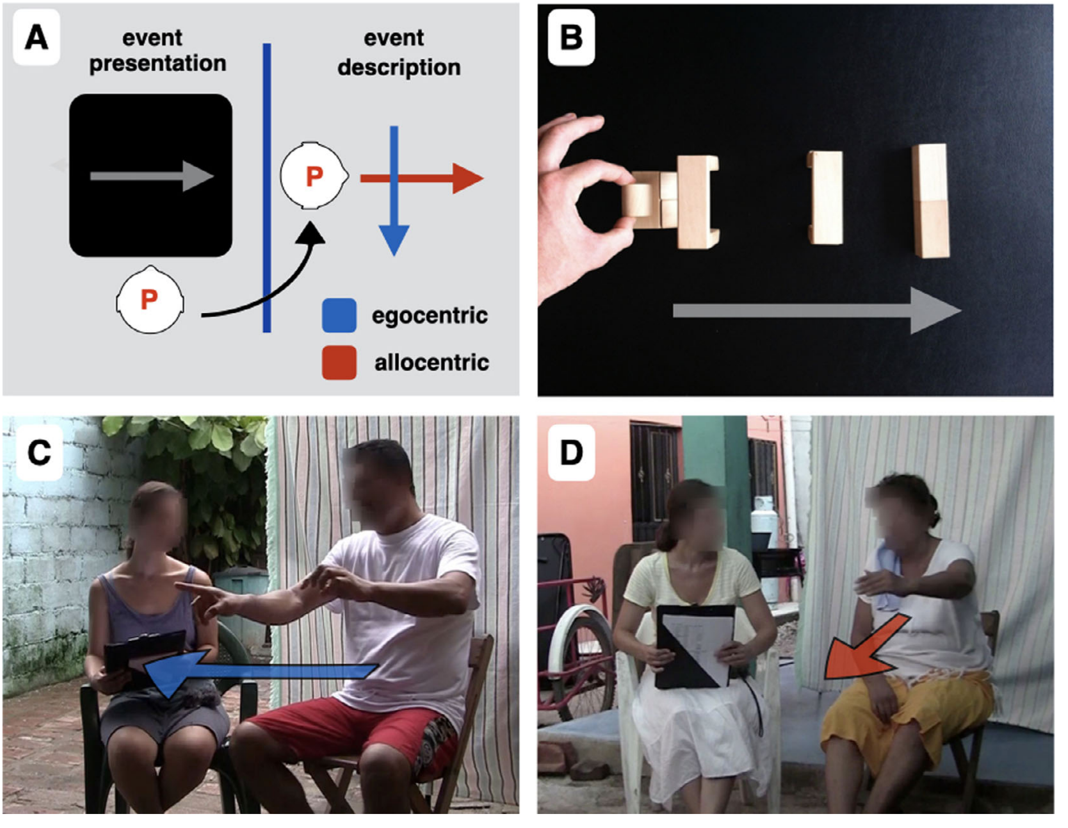

Fig. 1. Assessing spontaneous gesture with the "Toppling Blocks" task. (A) After watching a motion event, participants stepped around an opaque barrier, rotating 90° in the process, and then described what they had seen to an experimenter. For a rightward-presented event, a person gesturing egocentrically would depict the motion as unfolding rightward, across their body (blue arrow); a person gesturing allocentrically would depict the motion as unfolding forward, away from their body (red arrow). (B) One of the motion events, seen from above. Each event involved different configurations of blocks but always featured motion in one of four directions: rightward, leftward, toward the participant, or away from the participant. (C, D) Examples of gestures using an egocentric FoR (C) and an allocentric FoR (D), produced while describing the same rightward motion event in Spanish. These two speakers used different Spanish motion verbs to describe the event (*tumbar, derrumbar*; both translate as "to topple"), but neither used FoR language—such as "rightward" or "eastward"—to specify the event's direction. Note that the allocentric strategy involves gesturing away from the body to describe an event that had originally unfolded rightward from the participant's perspective, which may seem unusual to speakers of egocentric-favoring languages. The direction of motion was usually conveyed only in gesture. Video clips of the examples are in the OSF Supplementary Materials.

motion along that axis alternated between trials, within subjects. To familiarize participants with the task, the experimenter initiated the movement for the first half of trials, after which participants initiated the movement themselves.

Speech was transcribed from video. Irregularities (e.g., no clear video) made 3% of trials unanalyzable, and participants did not produce codable gestures on 4% of trials. For the remaining trials, gestures representing the primary motion of the event were coded for direction (out of eight possible directions: away, away-rightward, rightward, toward, toward-rightward, etc.) and for whether they were co-produced with FoR language (e.g., "left," "west"). Coders were blind to the axis and direction of the presented motion event. The gesture direction for a trial was determined from the predominant direction of motion gestures produced on that trial. The FoR used in gesture for a trial was determined from this predominant gesture direction relative to motion event direction (Fig. 1). For further details, see the coding manual in the OSF Supplementary Materials.

Overall, participants spontaneously produced motion gestures at high rates (N = 1,400, M = 2.4 gestures/trial), and on nearly every trial where a participant gestured, those gestures exhibited a predominant direction (95%). Coding reliability for gesture direction was high. A second coder analyzed three randomly selected trials from each session (i.e., 25% of the data). Directions assigned by the two coders were within 45° of each other on 93% of gestures; predominant directionality was judged the same on 80% of trials.

2.3.3. Array-reconstruction task ("Animals-in-a-Field")

To assess FoR preferences in noncommunicative spatial reasoning, we used a variant of the classic Animals-in-a-Row task (Pederson et al., 1998), which we made incrementally more complex to better tap FoR preferences (Levinson, Kita, Haun, & Rasch, 2002). In such tasks, participants memorize a spatial array of toy animals and then, after a delay and rotation, reconstruct the array. Participants' reconstructions reveal the FoR used: an egocentric FoR when animals were reconstructed relative to the participant's body (e.g., pig to the left); an allocentric FoR, when they were reconstructed according to the environment (e.g., pig to the north).

The setup consisted of two identical square tables, arranged parallel to the house, separated by an occluding sheet. The experimenter began each trial by placing three toys (two identical and one different) on one table to create an equilateral triangle that pointed toward the participant (see OSF Supplementary Materials for stimuli). Toys were laterally symmetric; each represented a familiar animal (rooster, sheep, cow, pig). Participants were instructed to "remember how they are." After a 30-s delay, they moved around the sheet to the other table, rotating 90° in the process. Participants were handed a bowl containing two of each animal and asked to "make it again, the same." Each participant completed six trials. See the Supplementary Methods and Analyses for still images of all items (Fig. S2) and of the task being completed by participants at the field site (Fig. S3).

The orientation of a reconstructed triangle was determined from overhead images by a naïve coder. Responses were coded as egocentric or allocentric if shape, orientation, and between-toy relations were all consistent with that FoR.

2.4. Data and statistical analyses

To investigate the factors that predict the use of a particular FoR in gesture, we used Bayesian multilevel generalized linear models of trial-level responses. We modeled each trial as a binomial outcome, fit separately for each FoR (i.e., one model of whether or not an egocentric FoR was used on each trial; another model of whether or not an allocentric FoR was used). We included predictors for factors related both to language and to the participants' experience of the motion event. Linguistic predictors were the participants' overall language dominance (i.e., Zapotec Dominant or Balanced Bilingual); the language they used in the session (i.e., Zapotec or Spanish); competence with egocentric terms; and competence with allocentric terms. Nonlinguistic predictors were the axis of motion (i.e., away–toward or left–right); and the participants' physical involvement with the event (i.e., whether they merely watched or physically initiated the motion on that trial). We also included a predictor for trial number (centered on the first trial). Dichotomous predictors were contrast coded (i.e., -0.5 vs. $+0.5$), and continuous predictors were rescaled to be on the same range (i.e., $[-0.5, +0.5]$). Priors for parameters were weakly informative: $N(0,100)$.

To investigate how participants' overall FoR preferences emerged from the dynamics of how, from trial to trial, they used and switched between FoRs, we modeled the time series of FoR use in gesture within each session as a Markov process (i.e., the FoR used on a particular trial is a function of the FoR used on the preceding trial). Specifically, we used a multilevel Markov model of trial-by-trial FoR use in gesture, clustered at the Session level, with two possible states: using an egocentric FoR or using an allocentric FoR. This allowed us to estimate the probability of sticking with a particular FoR after using it, and the probability of transitioning to using the other FoR. When analyzing the factors associated with FoR use, we also included as covariates the linguistic and nonlinguistic predictors described earlier.

Bayesian multilevel generalized linear models were fit using the brms package (Bürkner, 2017), and Markov models were fit using the MSM package (Jackson, 2011), in the R statistical software environment (R Core Team, 2013). See OSF Supplementary Materials for data and analysis scripts.

3. Results

3.1. Do spontaneous gestures reliably reflect a spatial FoR?

Only a small minority of motion gestures were accompanied by language that specified a particular FoR (e.g., "right-of" or "north-of"; 6.4% of trials). Despite the absence of spatial information in participants' speech, their gestures revealed systematic, stable preferences for allocentric and egocentric FoRs (Fig. 2A). Overall, across all trials, gesture motion was usually consistent with either an egocentric or allocentric FoR (72.9%), far more often than expected by chance ($p < .001$, binomial test, compared to 25% due to random gesturing alone).

At an individual level, we observed strong preferences for a single FoR. For each individual, we defined their predominant FoR as the one—either egocentric or allocentric—that they used most often. On average, individuals adopted their predominant FoR on nearly two-thirds of trials (M = 64.2%, SEM = 4.73%). For most people (66%), this

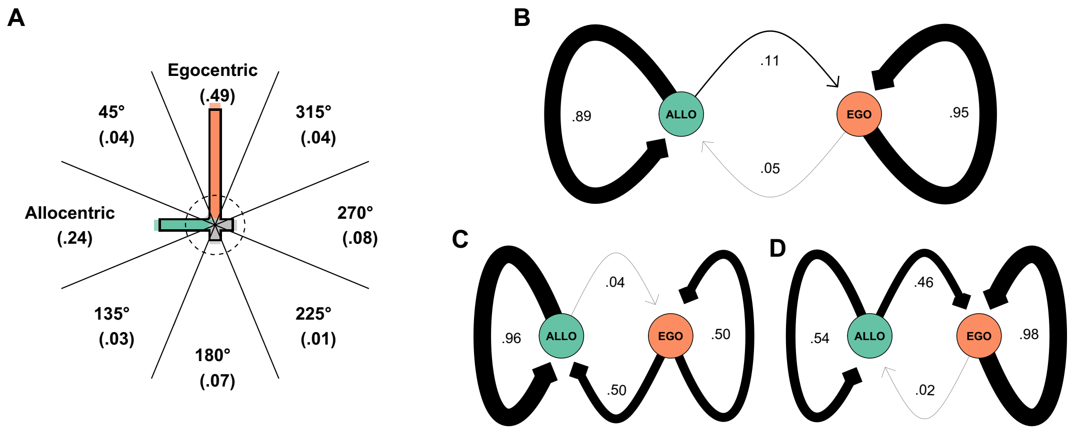

Fig. 2. Speakers systematically used either an egocentric or allocentric frame of reference (FoR) in gesture. (A) Distribution of gesture directions. Each octant represents one of eight possible gesture stroke directions on each trial, as viewed from above a speaker, oriented so that the top octant always represents the egocentric direction for a given trial. Since speakers turned 90° between viewing the event and describing it, the egocentric direction was always 90° clockwise from the allocentric direction. Within each octant, the proportion of trials for which the gesture was in that direction is represented by the radius of the black rectangle (also in parentheses). The lighter outer rectangle indicates the standard error of the proportion. The dashed circle indicates the distribution of gesture directions expected by chance alone. The majority (73%) of gestures were either egocentric or allocentric, with few gestures in any other direction. (B) Transition probabilities between allocentric and egocentric gesturing (estimated from a multilevel Markov model of trial-to-trial gesturing). Transition probabilities are indicated by line thickness and accompanying numbers. Both egocentric and allocentric FoRs were strong attractors in the trial-to-trial dynamics of gesture. (C) Transition probabilities for sessions with more allocentric than egocentric gestures. The egocentric FoR was an unstable strategy, with a 50% probability of switching to the allocentric FoR, while the allocentric FoR was a stable attractor. (D) Transition probabilities for sessions with more egocentric than allocentric gestures. The egocentric FoR was a stable attractor, while the allocentric FoR was unstable.

predominant FoR accounted for the majority of their trials. This was true despite the fact that, for a particular trial, each FoR was consistent with only one of the eight coded gesture directions. Individuals' gestures, therefore, revealed stable preferences for using either an egocentric or an allocentric FoR.

These preferences were highly stable over time, both within and between sessions. Within each session, the proportion of egocentric gestures in the first half of the session was highly correlated with the proportion in the second half ($r = .77$, $p < .001$), as it was for allocentric gestures ($r = .89$, $p < .001$). Recall that Balanced Bilinguals completed all tasks twice, in sessions separated by at least 7 weeks. In this group, the use of egocentric gestures was highly correlated between sessions ($r = .74$, $p = .001$), as it was for allocentric gestures ($r = .90$, $p < .001$). Thus, individuals' preferences for using an allocentric or egocentric FoR in gesture were highly stable over time, both short term and long term.

This pattern of preferences was also reflected in the trial-to-trial dynamics of FoR use. A multilevel Markov model of FoR use in gesture found that, overall, both the allocentric and the egocentric FoR were strong attractors (Fig. 2B). After using a particular FoR, participants were highly likely to stick with that same FoR on the next trial (probability of sticking with the same FoR: egocentric, .95; allocentric, .89) and thus highly unlikely to switch to the other FoR (probability of switching: egocentric to allocentric, .05; allocentric to egocentric, .11).

This overall pattern, however, collapses the dynamics of two distinct populations. Among participants who preferred the allocentric FoR (Fig. 2C), the allocentric FoR was highly stable (probability of sticking with the allocentric: .96), but the egocentric FoR was unstable and half the time they switched to the allocentric FoR (probability of switching from egocentric to allocentric: .50). Among participants who preferred the egocentric FoR (Fig. 2D), these dynamics reversed: The egocentric FoR was highly stable (probability of sticking with the egocentric: .98), but after using an allocentric FoR they switched half the time to the egocentric FoR (probability of switching from allocentric to egocentric: .46). Thus, while individuals exhibited mixed use of the two FoRs, they also exhibited strong preferences that generated two very different dynamics of FoR use in gesture.

In sum, most trials involved gesture that was interpretable as using either an egocentric or allocentric FoR; individuals had preferences for using one specific FoR in gesture, preferences which were often stable over many weeks; and these overall preferences emerged from the trial-to-trial dynamics of FoR use in gesture.

3.2. *What factors predict the use of a particular FoR in gesture?*

We first report results for linguistic factors, starting with their effects on the egocentric FoR (Table 1; Fig. 3A). Participants exhibited naturally occurring variability in lexical competence (Balanced Bilinguals: range = [.29, 1.00], $M = .79$, $SD = .22$; Zapotec Dominants: range = [.38, 1.00], $M = .77$, $SD = .17$). Moreover, performance on the lexical competence task varied independently from overall language dominance and from the particular language being used in the session (ps > .33). We were thus able to estimate independently the effects of language dominance, session language, and lexical competence.

Neither a participant's dominant language nor the language used during the task was associated with their use of an egocentric FoR (Table 1; Fig. 3A). In particular, Spanish-Zapotec Balanced Bilinguals were no more likely than Zapotec Dominants to use an egocentric FoR in gesture ($b = 0.06 \pm 1.40$ SD, 95% credible interval [−2.72, 2.81]). Likewise, completing the task in Spanish rather than in Zapotec had no reliable effect on the use of an egocentric FoR in gesture ($b = 0.85 \pm 0.72$ SD, 95% credible interval [−0.54, 2.34]). On average, Balanced Bilinguals used an egocentric FoR in gesture equally often whether they were speaking Spanish or Zapotec ($M_{\text{Spanish}} = 3.7$ trials vs. $M_{\text{Zapotec}} = 3.6$ trials), and Zapotec Dominants did not use the egocentric FoR reliably less often ($M = 2.9$ trials).

The only linguistic measure that was associated significantly with the use of an egocentric FoR was each participant's competence with egocentric terms ($b = 4.74 \pm 2.43$

Table 1
Predictor mean estimates, standard deviations, and 95% credible intervals from two Bayesian logistic mixed-effects models, predicting the use of an egocentric (left) or an allocentric (right) frame of reference (FoR) in gesture

	Egocentric FoR		Allocentric FoR	
Predictor	Estimate (*SD*)	95% CIs	Estimate (*SD*)	95% CIs
Intercept	−1.66 (1.03)	[−3.84, 0.25].	−2.20 (1.58)	[−5.76, 0.53]
Trial	−0.09 (0.08)	[−0.26, 0.07]	−0.23 (0.13)	[−0.49, 0.01].
Dominant Language	0.06 (1.40)	[−2.72, 2.81]	−1.49 (1.90)	[−5.56, 1.97]
Session Language	0.85 (0.72)	[−0.54, 2.34]	0.17 (1.16)	[−2.09, 2.58]
Egocentric Vocabulary	4.74 (2.43)	[0.30, 9.82]*	−2.91 (3.07)	[−9.47, 2.91]
Allocentric Vocabulary	1.23 (1.89)	[−2.59, 5.01]	−0.99 (2.75)	[−6.91, 4.18]
Axis	2.99 (1.27)	[0.62, 5.63]*	−5.35 (2.11)	[−10.23, −1.86]**
Initiated (vs. Watched)	1.01 (0.56)	[−0.06, 2.14].	0.98 (1.08)	[−1.33, 2.95]
Total observations	559		559	
F_1	0.82		0.89	

Note. Trial was centered to start at 0 and increase by 1 for each subsequent trial. Binary predictors were contrast coded (i.e., +0.5 vs. −0.5), so the parameter estimate indicates the effect of going from one level to the other: Dominant Language (Balanced Bilingual vs. Zapotec), Session Language (Spanish vs. Zapotec), Axis (Sagittal vs. Lateral), and whether the motion event was actively initiated (vs. passively watched). Continuous predictors (i.e., competence with Egocentric Vocab and with Allocentric Vocab) were rescaled to the same range, [−0.5, +0.5], so parameter estimates indicate the effect of going from worst to best competence. Credible intervals (CIs) that do not include zero are indicated by double asterisks (**) for 99% CIs, a single asterisk (*) for 95% CIs, and a period (.) for 90% CIs. Each model's predictive accuracy is captured by the F_1 score (the harmonic mean of precision and recall).

SD, 95% credible interval [0.30, 9.82]; Table 1; Fig. 3A). All other things being equal, the odds of using an egocentric FoR in gesture were more than 100 times greater among participants with the very best competence with egocentric terms, compared to those with the very worst competence with egocentric terms (i.e., $e^{4.74} = 114$). Indeed, participants whose competence with egocentric terms was in the top third used an egocentric FoR more than twice as often as participants in the bottom third ($M = 50.3\%$ of trials vs. $M = 19.1\%$). By contrast, the use of an allocentric FoR in gesture was not predicted by any of the linguistic measures (Table 1), including lexical competence with allocentric terms ($b = -0.99 \pm 2.75$ *SD*, 95% credible interval [−6.91, 4.18]).

We next report results for the effect of nonlinguistic features on FoR use in gesture (Table 1; Fig. 3B). Was the FoR used in gesture influenced by whether the participant passively observed or actively initiated the motion, or by the axis along which the observed motion had occurred? Whether participants actively initiated (vs. passively observed) the motion event did not have a reliable effect on the particular FoR used in gesture when describing the event. We were intrigued to find, however, that initiating the event was associated with a numerical increase in the probability of using *both* FoRs (egocentric: $b = 1.01 \pm 0.56$ *SD*; allocentric: $b = 0.98 \pm 1.08$ *SD*). While these effects were small and did not differ reliably from 0, the consistent numerical trend suggested

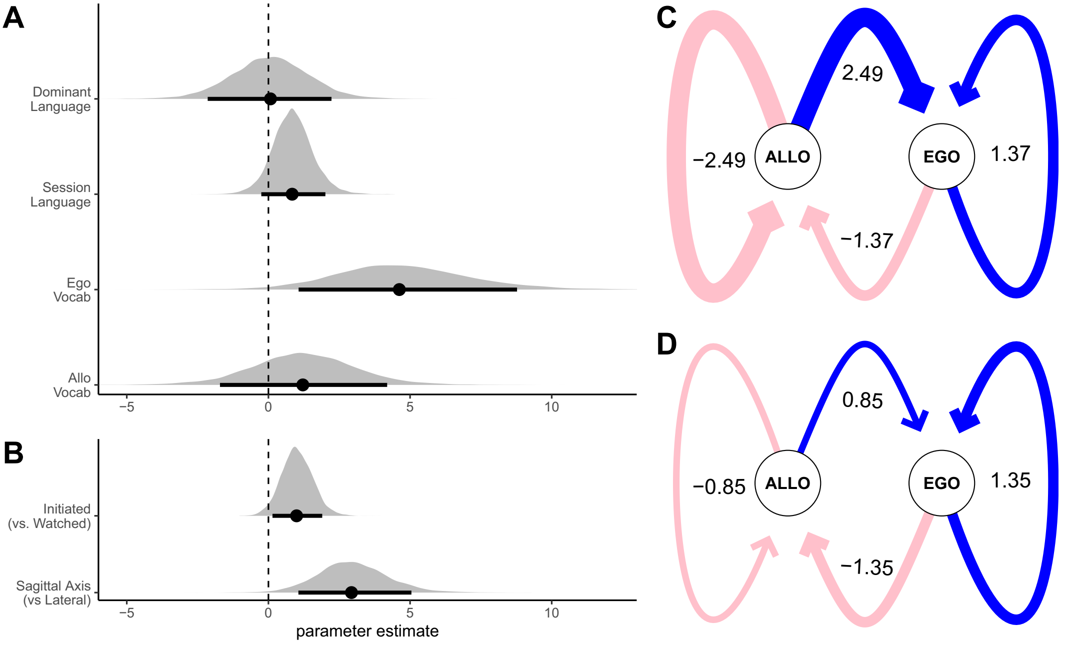

Fig. 3. Predicting the use of an egocentric frame of reference (FoR) in gesture. (A) Posterior estimates of the relations between language-related predictors and the use of an egocentric FoR in gesture, from a Bayesian logistic mixed-effects model. Participants with better competence with egocentric vocabulary were more likely to use the egocentric FoR in gesture. All other linguistic factors had numerically small effects, with 95% credible intervals that included 0. Points indicate mean parameter estimates, black error lines indicate 89% credible intervals, and gray distributions show the entire posterior distribution. (B) Posterior estimates for aspects of the observed event, from the same model as in Panel A. Motion events along the sagittal axis were more likely to be described using egocentric gestures. (C) Effect of egocentric vocabulary on the trial-to-trial dynamics of FoR use in gesture, as captured by a multilevel Markov model. Participants with better competence with egocentric vocabulary were more likely to switch from an allocentric to an egocentric FoR, and more likely to stick with the egocentric FoR once they used it. Log odds ratios (best vs. worst egocentric vocabulary) are indicated by the numbers, along with line thickness (magnitude) and color (sign; blue = better egocentric vocabulary competence associated with *higher* transition probability; pink = better competence associated with *lower* transition probability). (D) Effect of axis of motion on the trial-to-trial dynamics of FoR use in gesture. When participants described motions along the sagittal axis, they were more likely to switch from the allocentric to the egocentric FoR, and more likely to stick with the egocentric FoR once they used it. Log odds ratios (sagittal vs. lateral axis) are indicated by the numbers, line thickness (magnitude), and color (sign; blue = sagittal axis associated with *higher* transition probability; pink = sagittal axis associated with *lower* transition probability).

that physical involvement might increase the general propensity to gesture about the motion event. This was confirmed by an exploratory analysis: Using the same model but instead predicting whether *any* consistent FoR was produced on a trial, whether ego- or allocentric, we estimated that the odds of producing an FoR-consistent gesture more than tripled after initiating (vs. watching) the initial motion event (b = 1.32 ± 0.49, 95% credible interval [0.38, 2.31]). No other predictors were credibly different from zero.

By contrast, the axis along which the motion occurred did have a reliable effect on the FoR used in gesture. Compared to motion events along the lateral (left–right) axis, events along the sagittal (away–toward) axis were more likely to be depicted using an egocentric FoR (b = 2.99, 95% credible interval [0.62, 5.63]) and less likely to be depicted using an allocentric FoR (b = −5.35, 95% CI [−10.23, −1.86]). While participants who saw motion events along the lateral axis were just as likely to use either FoR (M_{ego} = 0.33 vs. M_{allo} = 0.32, t_{30} = 0.1, p = .93), participants who saw motion events along the sagittal axis were much more likely to use egocentric gestures (M_{ego} = 0.51 vs. M_{allo} = 0.10, t_{30} = 4.1, p < .01).

Overall, therefore, individuals with greater competence with egocentric terms were more likely to use an egocentric FoR in gesture, and events presented along the sagittal axis were more likely to be depicted in gesture using an egocentric FoR. These results were also reflected in the dynamics of trial-to-trial FoR use (Fig. 3C,D). Participants with the best (vs. worst) competence with egocentric terms were more likely to switch from using an allocentric FoR to using an egocentric FoR, and more likely to stick with an egocentric FoR once they had used it (Fig. 3C). This was also true of participants who had seen motion along the sagittal (vs. lateral) axis—they were more likely to transition from using an allocentric FoR to using an egocentric FoR, and more likely to stick with an egocentric FoR once they had used it (Fig. 3D). Aspects of individual lexical competence and of the event being described, therefore, reliably predicted the dynamics of how FoRs were used in gesture.

3.3. Do gesture and noncommunicative reasoning reflect a cohesive cognitive style?

To investigate whether people's gestures reflect a cohesive cognitive style for conceptualizing space, we compared participants' FoR use in gesture with their FoR use on the noncommunicative spatial reconstruction task. If both reflect people's cognitive styles, then FoR use in these tasks should be correlated. As predicted, within each session, the number of egocentric responses across the two tasks was significantly correlated (Pearson's correlation, r = .65, p < .001, bootstrapped 95% confidence interval [.40, .81]). For allocentric responses, however, the correlation was also positive but did not differ significantly from zero (r = .29, p = .10, bootstrapped 95% confidence interval [−.11, .65]).

This pattern was confirmed by a multilevel Bayesian logistic mixed-effects model of trial-by-trial use of an egocentric FoR on the spatial reconstruction task. The proportion of egocentric trials on the gesture task predicted the use of an egocentric FoR on the array-reconstruction task (b = 8.87 ± 5.31 *SD*, 95% credible interval [0.2, 21.3]), while other factors such as language dominance and the language being used in the session did not (both $|b|$ < 1.1 and 95% CIs included zero).

The preference for an egocentric FoR, therefore, was shared across different forms of spatial cognition, suggesting that an individual's spatial thinking and communicating reflects a cohesive style.

3.4. Spatial gestures without accompanying spatial language

While few motion gestures in the Toppling Blocks task were accompanied by language that specified a particular FoR (e.g., "right" or "north"; 6.25% of trials), we also re-conducted the analyses above with those gestures excluded. Results were nearly identical. Even after removing all trials in which participants produced FoR-specific language, their gestures systematically used either an egocentric or allocentric FoR, with very few gestures in other directions; analyses of between-trial transition probabilities revealed that egocentric and allocentric gestures were stable attractors (i.e., once people produced an egocentric or allocentric gesture, they typically stuck with that FoR; see Fig. S4 in the Supplementary Methods and Analyses, which is parallel to Fig. 1 in the main text). Moreover, even after removing trials with FoR-specific language, the same factors predicted the use of an egocentric FoR in gesture: better competence with egocentric vocabulary, and the axis along which the motion was initially observed (see Fig. S5, which is parallel to Fig. 3 in the main text, and Table S2, which is parallel to Table 1 in the main text).

4. Discussion

Spatial coordinate systems are foundational to communication and cognition, but preferences for a particular coordinate system vary across people and communities. More than two decades ago, ethnographers used naturalistic observations to suggest that such preferences are reflected not only in spoken language but also in spontaneous gesture. Yet, in the time since, the possibility has not been investigated systematically. This was the broad aim of the present work. We found that after people observed a complex, small-scale motion event, they spontaneously depicted this motion with their hands, reliably preserving how the event unfolded either relative to their bodies (egocentrically) or relative to the world (allocentrically). Moreover, despite trial-to-trial variability, people were remarkably stable in their preferences for one FoR or the other. We also found that people's FoR use was predicted by both linguistic and nonlinguistic factors. On the linguistic side, it was predicted by competence with relevant spatial vocabulary; on the nonlinguistic side, it was predicted by the axis along which the motion was experienced (sagittal vs. lateral). Finally, we found that participants' FoR use in gesture aligned with their FoR use in a noncommunicative array-reconstruction task, particularly for those using the egocentric FoR, suggesting that individuals have a cohesive cognitive style. Overall, the present study establishes the pervasiveness and systematicity of FoR use in spontaneous gesture, while also clarifying how gesture relates to other aspects of language, thought, and experience.

4.1. Spatial frames of reference in gesture, language, and thought

The present findings go beyond prior observations about FoR use in gesture in several ways. Much of the early pioneering work was qualitative and observational (Haviland, 1993; Levinson, 2003); it offered naturalistic examples of apparent FoR use in gesture but also left open critical questions about the robustness and limits of the phenomenon. More recently, a couple of controlled elicitation studies have analyzed the gestures speakers produce when explicitly asked to describe the spatial relationship between familiar, large-scale landmarks (Calderón et al., 2019; Le Guen, 2011). Here, we sought to build on these contributions by eliciting descriptions of small-scale events, using a procedure in which we expected directional information to be backgrounded. As expected, participants rarely verbalized the direction of motion (e.g., "to the right" or "to the north") in their descriptions. And yet directional information was nonetheless conveyed implicitly in the hands, replicating prior observations about how speech and gesture dissociate during motion-event descriptions (Kita & Özyürek, 2003; McCullough, 2005), and consistent with emerging results using similar methods (Núñez, Celik, & Nakagawa, 2019). Critically, most gestures were consistent with either an egocentric or allocentric FoR. Another important aspect of these findings is that gestures conveying directional information are pervasive. Gesture was not mentioned in our instructions or otherwise prompted, and yet participants consistently conveyed the direction of motion with their hands, producing gestures with a predominant direction on 95% of trials. On this task, at least, the amount of FoR information in gesture dwarfs the amount of FoR information in speech.

Another key finding of the present study is that FoR preferences in gesture were predicted by individual differences in competence with spatial lexical items. Of the three linguistic factors—language dominance, language used during the session, and lexical competence—only lexical competence proved predictive.[4] Specifically, we found that those with better mastery of "left" and "right" were more likely to gesture egocentrically. An interesting aspect of this finding, however, is that we found a stronger predictive relationship between *egocentric* vocabulary and egocentric gesturing than between allocentric vocabulary and allocentric gesturing. One possible explanation for this pattern is an evolved bias for allocentric encoding (Haun, Rapold, Call, Janzen, & Levinson, 2006). On this account, allocentric encoding is a kind of natural default in humans and great apes, while egocentric encoding is a more recent add-on or "cultural override" (Haun et al., 2006) that requires linguistic and cultural scaffolding. In the present case, such scaffolding may include the acquisition and mastery of lexical items for "right" and "left," as well as literacy and related practices. Recent developmental work fits this "cultural override" picture, showing, for instance, that young English-speaking children interpret new spatial words as having *allocentric* meanings, despite belonging to a community of speakers that overwhelmingly favors the egocentric FoR (Shusterman & Li, 2016; but see also Li & Abarbanell, 2019). Thus, egocentric and allocentric FoR vocabulary may function differently: Mastering egocentric words may crystallize and promote a strategy that is initially dispreferred, whereas mastering allocentric words may serve merely to put a label on a prepotent preference. A prediction that follows from this account—and one consistent with recent findings (Calderón et al., 2019; Le Guen, 2011)—is that allocentric behavior may occur with or without support from allocentric language, while egocentric behavior may be found only where linguistic and cultural support is in place.

We also tested whether nonlinguistic factors would shape FoR choice, focusing in particular on aspects of how an event is experienced. We found little evidence for an effect of initiating a motion event rather than merely observing it on FoR choice (though it appears to increase use of *either* FoR). But we did observe a strong effect of the axis of motion: Participants who saw sagittal events—unfolding toward or away from their bodies—were much more likely to represent those events egocentrically than participants who saw lateral events—unfolding leftward or rightward. This pattern suggests an important boundary condition on FoR use in gesture and perhaps on memory more generally. Indeed, several earlier studies—in different populations, using different noncommunicative tasks—have reported congruent patterns (Brown & Levinson, 1993; Li & Abarbanell, 2019; Shapero, 2017). Why would this be the case? As several observers have noted, the body's sagittal axis is more salient than its lateral axis by virtue of being more strongly asymmetric (Clark, 1973; Tversky, 2011). Evidence from language acquisition supports this idea: Children learn words for "front" and "back" long before they learn terms for "left" and "right," even in communities where "left" and "right" are privileged by adults (reviewed in Shusterman & Li, 2016). A related possibility is that motion away or toward the body invites a so-called deictic or body-anchored interpretation, both when making sense of an ambiguous scenario (Li & Abarbanell, 2019) or, as in the current study, when describing a past experience. More generally, our results build on prior work showing that aspects of an event can shape how it is subsequently depicted in gesture (e.g., from the viewpoint of an outside observer or that of an embedded character; Parrill, 2010). Though the present study focused on factors related to the experience of an event, future work might also examine how factors related to the communication of the event (e.g., sitting configuration) could affect FoR choice.

A final issue we examined was whether FoR preferences in gesture relate to FoR preferences in other forms of spatial behavior. In addition to the "Toppling Blocks" task, we tested people on an array-reconstruction task of the type widely used in prior work. While the relationship was weak for the allocentric FoR, use of the egocentric FoR was strongly correlated across the two tasks. This correlation was present despite the many ways in which the two tasks differed—indeed, were designed to differ so that participants would not see them as the same. For example, the "Toppling Blocks" task elicits dynamic representations of motion that are produced spontaneously in a communicative context; the "Animals-in-a-Field" task elicits reconstructions of static relationships in a noncommunicative setting. Moreover, whereas spontaneous gestures are often characterized as unwitting and implicit (McNeill, 1992), array-reconstructions are often characterized—and criticized—as strategic (Li et al., 2011). Despite such differences, we found a relationship between preferences on these two tasks, particularly for the egocentric FoR. These results are consistent with prior claims that individual differences in preferred FoR reflect cohesive cognitive styles of spatial thought and behavior in individuals and communities of speakers (Levinson, 2003; Majid et al., 2004).

4.2. Correlation, causation, and natural variability in the study of language and thought

In the traditional cross-cultural methodology that has generated much of the evidence for relations between language and thought, communities that differ in how they speak are shown to also differ in how they think (e.g., Dutch and Namibian children; Haun et al., 2011). As many commentators have pointed out (e.g., Li & Gleitman, 2002), this confounds linguistic differences with countless other nonlinguistic differences between the communities (e.g., rural vs. urban, degree of economic development, etc.).

Here, by contrast, we leveraged both naturally occurring variability and controlled experiment within a single, close-knit community. Within the community, we observed variability in both language dominance and lexical competence. We were thus able to isolate linguistic differences from many other cultural or environmental differences. Of course, this does not allow us to identify the precise causal relation between lexical competence and spatial gesture. Egocentric and allocentric terms may participate in different spatial practices, for instance, and our assessment of lexical competence may be an indirect measure of facility with these practices.[5] We suspect, however, that the relation between lexical competence and gesture is a causal one. Learning to use egocentric terms like "left" and "right" may facilitate access to, or entrench, the egocentric FoR, leading to an increased use of egocentric gesture. This interpretation is in line with other work suggesting that acquiring words may shape reasoning, both in the domain of space (Gentner, 2003; Gentner et al., 2013; Pyers, Shusterman, Senghas, Spelke, & Emmorey, 2010; Shusterman & Spelke, 2005) and beyond (e.g., Frank, Everett, Fedorenko, & Gibson, 2008; Winawer et al., 2007). The spatial FoR used in speech and gesture, however, are likely shaped by a variety of factors, as discussed later, and any monocausal story would erase the richness of how we reason and communicate about space.

At the same time, we directly manipulated factors such as the language used during the task and how the motion event was experienced. As a result, our finding that egocentric gestures were used more often to describe sagittal motion is direct evidence for a causal effect of one aspect of a speaker's experience—namely, the axis of motion—on the spatial FoR used to describe that event.

4.3. A cognitive ecology of space

We argue that the present results invite us to understand FoR in gesture as part of a broader cognitive ecology. By a cognitive ecology, following Hutchins (2010), we mean the larger web of mutually dependent elements relevant to cognition—in this case, spatial cognition. Such an ecology comprises diverse elements, including linguistic regularities, such as grammatical rules, lexical resources, and discourse norms (e.g., Lucy, 2016); gesture (e.g., Le Guen, 2011), perhaps especially in its "foreground" versions (Cooperrider, 2017); other shared cultural practices, including literacy (e.g., Danziger & Pederson, 1998; Meakins, Jones, & Algy, 2016), subsistence modes (e.g., Shapero, 2017), navigational strategies and artifacts (e.g., Hutchins & Hinton, 1984; MacEachren, 1986); and features of the physical environment, both natural (Palmer, 2015; Palmer, Lum, Schlossberg, & Gaby, 2017) and built (Cooperrider, Slotta, et al., 2017; Majid et al., 2004).

Beyond underscoring the complexity of spatial communication and cognition, an ecological framework invites several important shifts of emphasis. A first shift is that, within a cognitive ecology, mutual dependence is the rule rather than the exception. The framework thus moves our focus away from simple influences of language on thought and toward the complex ways in which different elements in the ecology shape or "regiment" others (Cooperrider, Marghetis, & Núñez, 2017). Habitual adoption of a particular FoR in gesture, for instance, might play a role in regimenting how the broader community thinks and talks about space (Le Guen, 2011; Levinson, 2003), even while gesture itself is regimented by aspects of language (such as individuals' mastery of egocentric terms, or the very availability of those terms within a language). Spatial language, in turn, might be shaped by other elements of the broader cognitive ecology. While our focus here has been on spatial gesture—and its relation to other aspects of language and thought—future work may well uncover a much larger web of mutual dependence.

A second shift is that some parts of a cognitive ecology may play especially powerful roles, much like "keystone species" in natural ecosystems (e.g., Mills, Soulé, & Doak, 1993). Indeed, in the cognitive ecology of space, gesture may often play such a role. There is indirect evidence, for instance, that spatial gesture can transmit norms of FoR use in the absence of spatial language (e.g., Le Guen, 2011). In general, however, the importance of each element is not fixed in a rigid hierarchy but may depend critically on the larger sociocultural context in which it is embedded. Formal classroom education, for instance, might privilege the spatial resources available in speech, while other modes of cultural transmission may rely more on embodied resources like gesture. An ecological perspective thus invites us to investigate how sociocultural diversity might be associated with different causal relations among speech, gesture, and other facets of spatial cognition.

A third shift is that an ecological framework invites us to think about how, on different timescales, cognition can exhibit both change and stability (Hutchins, 1995)—much as natural ecologies can undergo "regime shifts" between alternative stable configurations (Scheffer & Carpenter, 2003). Individuals in our study exhibited trial-to-trial variability; but they also had a strong tendency to settle into a preferred FoR, and those who were tested in two sessions were strikingly consistent across time. This stability may be a consequence of how different elements within the ecology complement and reinforce each other. But cognitive ecologies can also reconfigure (Hutchins, 1995, Ch. 8), sometimes suddenly, such as when communities shift from one preferred spatial framework to another (e.g., Hendricks, Bergen, & Marghetis, 2018; Meakins & Algy, 2016; Meakins et al., 2016). Such cognitive regime shifts may be associated with major sociocultural transitions, such as the demographic shifts currently underway in Juchitán (McComsey, 2015) and elsewhere in Mexico (Calderón et al., 2019). The ecological framework just sketched stands to illuminate not only the domain of space, but other domains as well, such as time (Hendricks et al., 2018), number (Cooperrider, Marghetis, et al., 2017), and even higher mathematics (Goldstone, Marghetis, Weitnauer, Ottmar, & Landy, 2017; Marghetis, Landy, & Goldstone, 2016).

5. Conclusion

When people speak and gesture about events, their hands convey rich spatial information, including information about action, perspective, and manner of motion (Alibali, 2005;

Parrill, 2010; Sekine, 2009), as well as the coordinate system they are using to conceptualize this information. In certain contexts, such as the one studied here, people may in fact be more likely to convey FoR information with their hands than with their words. Gesture may thus be considered a prominent, even privileged, vehicle for conveying FoR. This fact has several implications. For one, it suggests that gesture provides a naturalistic research tool for examining cognitive representations of space, including how they vary across individuals, contexts, and communities. More broadly, it invites us to move beyond an understanding of spatial FoR as being foremost about language—that is, as primarily reflected in, and shaped by, language. Instead, it urges a view in which language—but also hand, mind, and much else besides—operate within larger ecologies of space.

Acknowledgments

All authors share primary authorship of this article. All authors contributed substantially to study conception and design, acquisition of data, analysis and interpretation of data, as well as drafting and revising the manuscript. All authors spent time at the field site, which was established by M.M. M.M. led the field research, K.C. led the coding of the gesture, and T.M. led the statistical analyses. This work was supported by the American Philosophical Society (K.C.), the Wenner-Gren Foundation (M.M.), the National Science Foundation #1061053 (M.M.), Jacobs Research Funds (M.M.), and the Elizabeth Bates Graduate Research Award (M.M.). For research assistance, we thank Alec Gasperian, Felicia Jin, Reyna López López, Luis Bernardo Quesada Nieto, and Ana Luis Ruiz Regalado. For critical suggestions and discussion, we thank B. Bergen, S. Coulson, J. Haviland, E. Hutchins, R. Núñez, A. Majid, and two anonymous reviewers.

Open Research badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/agkvh/.

Notes

1. This kind of allocentric reasoning is different from the ability to dead-reckon the location of distant landmarks. Members of some cultural groups can point with staggering precision to distant, invisible landmarks (Levinson, 2003; Lewis, 1976). Our focus here is a separate issue: the preservation of a spatial coordinate system as one moves around an environment, as in the example of someone gesturing about the cake toppling to the south.

2. Researchers in the cross-linguistic study of spatial cognition often distinguish two subtypes of allocentric FoR: the *absolute*, based on abstract coordinates such as cardinal directions; and the *intrinsic*, where one object is located based on the asymmetries of a reference point, such as the front or back of a building (e.g., Danziger, 2010; Levinson, 2003). We collapse this distinction because, in the nonlinguistic tasks used here as well as in naturalistic spatial gestures, the subtype involved is often ambiguous. For example, when a speaker, under rotation, reproduces the allocentric coordinates of an event in gesture, it can be impossible to tell whether they are conceptualizing that event in terms of cardinal directions (absolute FoR) or in terms of a nearby landmark or aspect of the experimental setup (intrinsic FoR). Accompanying speech, if used, may disambiguate the subtype, but the gesture on its own does not.
3. Later we investigate whether lexical competence predicts FoR use in gesture. We repeated those analyses but with lexical competence calculated from only trials within the same session as the gesture trials (i.e., using Balanced Bilingual's lexical competence in Spanish to predict their gestures in Spanish, and their lexical competence in Zapotec to predict gestures in Zapotec). Results were qualitatively identical and numerically similar. See Supplementary Table S3.
4. While it may seem surprising that we did not see effects of the other linguistic factors, recent work on bilingualism in gesture has found that monolinguals and bilinguals from the same population do not necessarily differ in their gestural styles (e.g., Calderón, et al., 2019) and that bilinguals do not necessarily switch gestural patterns when switching languages (e.g., So, 2010). Results from this literature have been mixed, however (Gullberg, 2012).
5. See McComsey (2015, especially Ch. 3), for a discussion of spatial language use among Spanish-Zapotec bilinguals, both at this field site and in the surrounding region.

References

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition & Computation*, *5*(4), 307–331.

Bohnemeyer, J., Donelson, K. T., Moore, R. E., Benedicto, E., Eggleston, A., O'Meara, C. K., Pérez Báez, G., Capistrán Garza, A., Hernández Green, N., de Jesús Selene Hernández Gómez, M., Herrera Castro, S., Palancar, E., Polian, G., & Romero Méndez, R. (2015). The contact diffusion of linguistic practices: Reference frames in Mesoamerica. *Language Dynamics and Change*, *5*(2), 169–201.

Brown, P., & Levinson, S. C. (1993). Linguistic and nonlinguistic coding of spatial arrays: Explorations in Mayan cognition. Working Paper 24. Nijmegen, Netherlands: Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics.

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Calderón, E., De Pascale, S., & Adamou, E. (2019). How to speak "geocentric" in an "egocentric" language: A multimodal study among Ngigua-Spanish bilinguals and Spanish monolinguals in a rural community of Mexico. *Language Sciences*, *74*, 24–46.

Cooperrider, K. (2017). Foreground gesture, background gesture. *Gesture*, *16*(2), 176–202.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 28–63). New York: Academic Press.

Cooperrider, K., Marghetis, T., & Núñez, R. (2017). Where does the ordered line come from? Evidence from a culture of Papua New Guinea. *Psychological Science*, *28*(5), 599–608. https://doi.org/10.1177/0956797617691548

Cooperrider, K., Slotta, J., & Núñez, R. (2017). Uphill and downhill in a flat world: The conceptual topography of the Yupno house. *Cognitive Science*, *41*, 768–799. https://doi.org/10.1111/cogs.12357

Danziger, E. (2010). Deixis, gesture, and cognition in spatial frame of reference typology. *Studies in Language*, *34*(1), 167–185. https://doi.org/10.1075/sl.34.1.16dan

Danziger, E., & Pederson, E. (1998). Through the looking glass: Literacy, writing systems and mirror-image discrimination. *Written Language & Literacy*, *1*(2), 153–169.

De Korne, H. (2016). Imagining convivial multilingualism: Practices, ideologies and strategies in Diidxazá/Isthmus Zapotec indigenous language education [Unpublished doctoral dissertation].University of Pennsylvania.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*, 819–824.

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.

Gentner, D., Ozyürek, A., Gürcanli, O., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition*, *127*, 318–330.

Goldstone, R. L., Marghetis, T., Weitnauer, E., Ottmar, E. R., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science*, *26*, 434–441.

Grosjean, F. (2010). *Bilingual*, Cambridge, MA: Harvard University Press.

Gullberg, M. (2012). Bilingualism and gesture. In T. Bhatia & W. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 417–437). Malden, MA: Wiley-Blackwell.

Haun, D., Rapold, C. J., Call, J., Janzen, G., & Levinson, S. C. (2006). Cognitive cladistics and cultural override in Hominid spatial cognition. *Proceedings of the National Academy of Sciences*, *103*, 17568–17573.

Haun, D., Rapold, C. J., Janzen, G., & Levinson, S. C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition*, *119*, 70–80.

Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology*, *3*, 3–45.

Hendricks, R., Bergen, B. K., & Marghetis, T. (2018). Do metaphors move from mind to mouth? Evidence from new metaphors for time. *Cognitive Science*, *42*, 2950–2975.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hutchins, E. (1995). *Cognition in the Wild*, Cambridge, MA: MIT Press.

Hutchins, E. (2010). Cognitive ecology. *Topics in Cognitive Science*, *2*, 705–715.

Hutchins, E., & Hinton, G. E. (1984). Why the islands move. *Perception*, *13*(5), 629–632.

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, *38*, 1–29.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32.

Le Guen, O. (2011). Speech and gesture in spatial language and cognition among the Yucatec Mayas. *Cognitive Science*, *35*, 905–938.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*, Cambridge, UK: Cambridge University Press.

Levinson, S. C., Kita, S., Haun, D. B. M., & Rasch, B. H. (2002). Returning the tables: Language affects spatial reasoning. *Cognition*, *84*, 155–188.

Levinson, S. C., & Wilkins, D. (2006). *Grammars of space*, Cambridge, UK: Cambridge University Press.

Lewis, D. (1976). Observations on route finding and spatial orientation among the aboriginal peoples of the Western desert region of central Australia. *Oceania*, *46*(4), 249–282.

Li, P., & Abarbanell, L. (2019). Alternative spin on phylogenetically inherited spatial reference frames. *Cognition*, *191*, 103983. https://doi.org/10.1016/j.cognition.2019.05.020

Li, P., Abarbanell, L., Gleitman, L., & Papafragou, A. (2011). Spatial reasoning in Tenejapan Mayans. *Cognition*, *120*, 33–55.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*, 265–294.

Lucy, J. A. (2016). Recent advances in the study of linguistic relativity in historical context: A critical assessment. *Language Learning*, *66*, 487–515.

MacEachren, A. M. (1986). A linear view of the world: Strip maps as a unique form of cartographic representation. *The American Cartographer*, *13*(1), 7–25.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*, 108–114.

Marghetis, T., Landy, D., & Goldstone, R. L. (2016). Mastering algebra retrains the visual system to perceive hierarchical structure in equations. *Cognitive Research: Principles and Implications*, *1*, 25.

McComsey, M. (2015). Bilingual spaces: Approaches to linguistic relativity in bilingual Mexico [Unpublished doctoral dissertation]. University of California San Diego.

McCullough, K. (2005). Using gestures during speaking: Self-generating indexical fields [Unpublished doctoral dissertation]. University of Chicago.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Meakins, F., & Algy, C. (2016). Deadly reckoning: Changes in Gurindji children's knowledge of cardinals. *Australian Journal of Linguistics*, *36*(4), 479–501.

Meakins, F., Jones, C., & Algy, C. (2016). Bilingualism, language shift and the corresponding expansion of spatial cognitive systems. *Language Sciences*, *54*, 1–13.

Mills, L. S., Soulé, M. E., & Doak, D. F. (1993). The keystone-species concept in ecology and conservation. *BioScience*, *43*(4), 219–224.

Moore, R. (2018). Spatial language and cognition in Isthmus Zapotec [Publication No. 10823881] (Doctoral dissertation, State University of New York at Buffalo). Proquest Dissertations Publishing.

Núñez, R., Celik, K., & Nakagawa, N. (2019). Absolute frames of reference in bilingual speakers of endangered Ryukyuan languages: An assessment via a novel gesture elicitation paradigm. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual conference of the Cognitive Science Society* (pp. 890–896). Montreal: Cognitive Science Society.

Palmer, B. (2015). Topography in language: Absolute frame of reference and the topographic correspondence hypothesis. In R. de Busser & R. J. LaPolla (Eds.), *Language structure and environment: Social, cultural, and natural factors* (pp. 177–226). Amsterdam: John Benjamins.

Palmer, B., Lum, J., Schlossberg, J., & Gaby, A. (2017). How does the environment shape spatial language? *Evidence for Sociotopography. Linguistic Typology*, *21*(3), 457–491.

Parrill, F. (2010). Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, *25*(5), 650–668.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S. C., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language*, *74*, 557–589.

Pérez Báez, G. (2011). Spatial frames of reference preferences in Juchitán Zapotec. *Language Sciences*, *33*, 943–960.

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences*, *107*, 12116–12120.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Scheffer, M., & Carpenter, S. R. (2003). Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends in Ecology & Evolution*, *18*(12), 648–656.

Sekine, K. (2009). Changes in frame of reference use across the preschool years: A longitudinal study of the gestures and speech produced during route descriptions. *Language and Cognitive Processes*, *24*(2), 218–238.

Shapero, J. A. (2017). Does environmental experience shape spatial cognition? Frames of reference among Ancash Quechua speakers (Peru). *Cognitive Science*, *41*(5), 1274–1298.

Shusterman, A., & Li, P. (2016). Frames of reference in spatial language acquisition. *Cognitive Psychology*, *88*, 115–161.

Shusterman, A., & Spelke, E. S. (2005). Language and the development of spatial reasoning. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 89–108). Oxford: Oxford University Press.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.

So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese-English bilinguals. *Language and Cognitive Processes*, *25*(10), 1335–1353.

Tversky, B. (2011). Spatial thought, social thought. In T. W. Schubert & A. Maass (Eds.), *Spatial dimensions of social thought* (pp. 17–38). Berlin: Mouton de Gruyter.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.