Gestural Relativity of Spatial Cognition: Speakers' co-speech gestures shape listeners' spatial frame of reference

Shervin Nosrati (nosrati.shervin@gmail.com)
Tyler Marghetis (tyler.marghetis@gmail.com)

Cognitive & Information Sciences, University of California, Merced
5200 Lake Rd, Merced, CA 95343

Abstract

To think about objects' locations, people adopt a spatial frame of reference anchored either to their own body (egocentric; e.g., left vs. right) or to something external (allocentric; e.g., cardinal directions). Within cultures, people habitually rely on the same frame of reference, manifested in language, gesture, and memory. How are these norms transmitted? One account, linguistic relativity, argues they are transmitted through language. Here we explore a complementary route: gesture. In a between-subjects experiment (N = 70), we manipulated the spatial frame of reference used in gesture to describe table-top locations. As predicted, participants reliably adopted this frame of reference in a subsequent spatial search task, even after the speaker stopped gesturing. This suggests that a speaker's gesture has the capacity to reshape listeners' spatial reasoning. We argue that this offers a mechanism for "gestural relativity," which we consider in light of a larger cognitive-ecological perspective on spatial cognition.

Keywords: space, gesture, frames of reference, gestural relativity, linguistic relativity

Introduction

How else could we think about objects if not in space? To think about objects' locations, people adopt a spatial frame of reference. We can thus conceive of spatial relationships in multiple ways. We can think about spatial relationships *egocentrically*, anchoring them to ourselves — as in when one thinks about their coffee *to the left of the computer.* The same spatial layout could be conceived of *allocentrically,* anchored to external features such as cardinal directions (eg. "The coffee is *west of* the computer") or local topography (eg. "The coffee is *mountainward of* the computer"). While people in many industrialized cultures default to an egocentric encoding of small-scale space, other communities default to an allocentric encoding, sometimes lacking words for egocentric relations entirely (Brown & Levinson, 1993; Levinson, 1996). While people can flexibly shift between frames of reference, communities have default preferences that reflect the habitual adoption of one frame of reference.

There is considerable global variation in these default preferences in spatial thinking, with people in different communities defaulting to different frames of reference in reasoning and memory (Levinson, 2003; Majid et al, 2004; Marghetis et al, 2024). This variation in spatial thinking is correlated with variation in spatial language. Communities that habitually describe spatial relations egocentrically tend also to adopt an egocentric frame of reference in spatial thinking (e.g., reasoning, memory), while those that describe spatial relations allocentrically tend also to think allocentrically. This cross-cultural correlation between language and thought has been taken as evidence of linguistic relativity, a causal influence of language use on thought (Levinson, 1996; Levinson, 2003; Majid et al, 2004). On this account, a community's preference for a particular spatial frame of reference in cognition is transmitted through the habitual use of, and repeated exposure to, that frame of reference in language (Levinson, 1996; Majid et al, 2004). When members of a community reliably describe spatial relations in a particular way, for instance, each spatial description may act as a cue for listeners that prompts them to think in a similar way. While language does not *determine* thought, it may influence and shape it.

But correlation does not imply causation, and the linguistic relativity account faces a number of challenges. In some communities, for instance, people habitually adopt one frame of reference in cognition but another in speech. For instance, Yucatec Mayan women and Ngigua-Spanish bilinguals reliably adopt an allocentric frame of reference in spatial reasoning but an egocentric frame of reference in language (Le Guen, 2011; Calderón et al., 2019). This suggests that linguistic relativity, at minimum, cannot be the whole story. Norms of spatial cognition might sometimes be transmitted through language, but other routes of transmission must exist.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

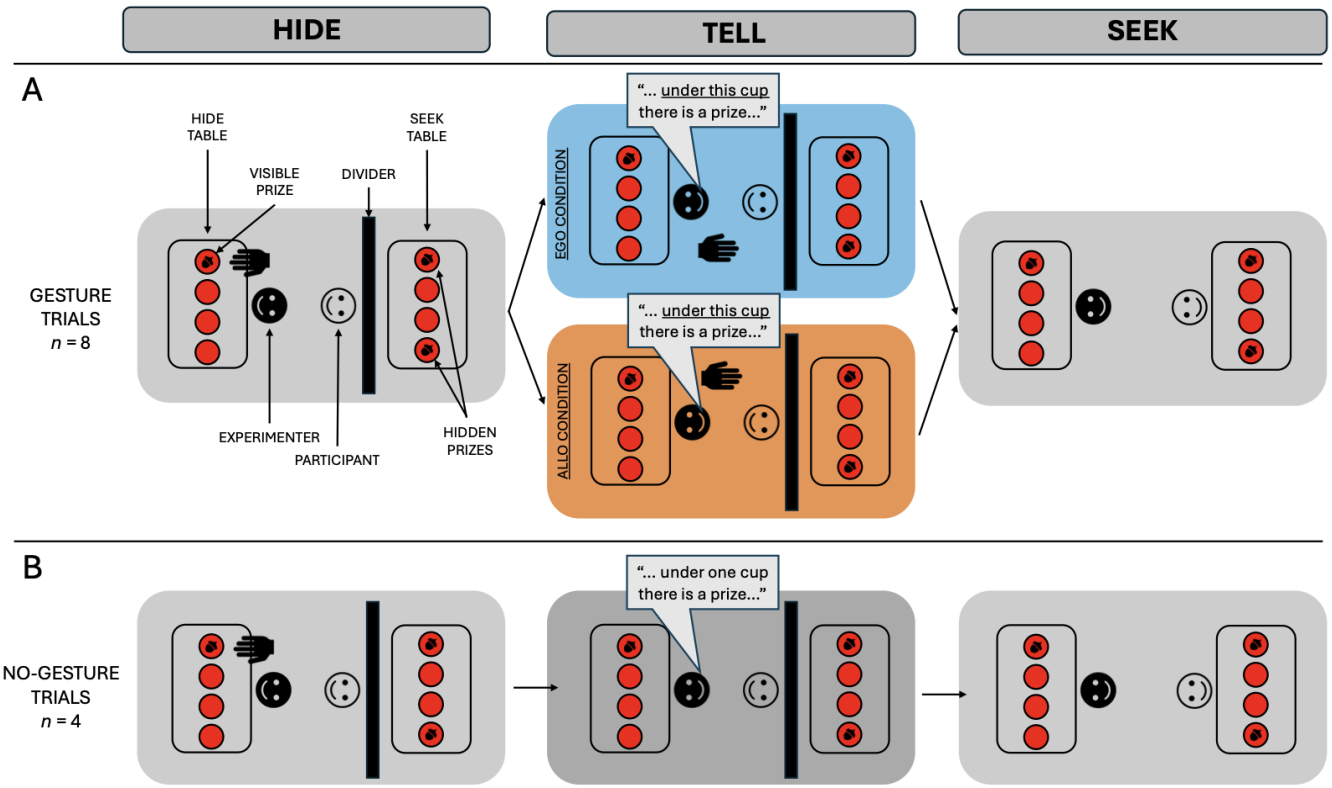

Fig. 1: Procedure of the spatial search task. (A) Procedure for Gesture trials. While describing the location of the prize they had hidden in the initial HIDE phase, the experimenter produced a gesture (represented by the hands in the TELL section) that adopted either an egocentric (blue) or an allocentric (orange) frame of reference. The participant then rotated 180º to search for a matching prize in the "same" location. (B) Procedure for No-Gesture trials. The No-Gesture trials were identical to Gesture trials except the experimenter never gestured.

How else might communities convey their default preference for one spatial frame of reference over others? Spatial frames of reference are not only encoded in language (Levinson, 1996) but also in gesture (Haviland, 1993; Marghetis et al, 2020). The spatial frame of reference habitually adopted in gesture has been speculated to correlate with the frame of reference used in language, memory, and reasoning (Levinson et al., 2002). This raises the possibility that co-speech gesture may transmit a community's default preferences for a particular spatial frame of reference (Levinson, 2002; Le Guen, 2011). Since spatial encoding in co-speech gesture covaries with spatial descriptions in speech (Haviland, 1993; Kita, S., & Özyürek; Marghetis et al, 2020), this would also explain the cross-linguistic correlation between the frame of reference habitually adopted in language and in cognition. Further, it would explain instances where frames of reference in language and gesture diverge. Calderón et al. (2019), for instance, suggest that "direct pointing gestures suffice to maintain geocentric linguistic conceptualizations even after the loss of the geocentric vocabulary." Crucially, in the case of Yucatec Mayans, Le Guen (2011) speculated that Yucatec Mayan women may acquire a preference for allocentric encoding in cognition through exposure to co-speech gesture. If the adoption of a particular frame of reference can be transmitted through co-speech gesture, then community norms in spatial cognition might reflect a form of *gestural relativity*: repeated exposure to frame-of-reference-encoding gestures could develop stable, shared preferences for a particular frame of reference.

Here, to test the viability of gestural relativity of spatial frames of reference, we investigate whether co-speech gesture can shape observers' spatial frame of reference and thus serve as a pathway for transmission. In a between-subjects experiment, we manipulated the spatial frame of reference adopted in co-speech gesture to describe a spatial location; participants then conducted a spatial search which could be accomplished either allocentrically or egocentrically. We hypothesised that observed co-speech gesture would shape the spatial frame of reference that participants adopted. For gestural relativity to be a viable mechanism for the propagation and perpetuation of community conventions governing spatial reasoning, co-speech gesture must be capable of shaping observers' spatial reasoning. More precisely, observing a speaker encode spatial relationships egocentrically in co-speech gesture should cause observers to adopt an egocentric frame

of reference in their own reasoning, while observing a speaker encode them allocentrically in co-speech gesture should cause observers to adopt an allocentric frame of reference.

Methods

Participants completed a rotation-based spatial search task (adapted from Haun et al, 2006, and Pitt et al., 2021). Crucially, we incorporated a gesture manipulation (Fig. 1). Participants' search behavior revealed the spatial frame of reference they used to encode the task.

Participants

Participants were college students at an American research university (N = 70; gender: 44 female, 26 male: M = 19 years, range [18, 29]). The majority of participants reported English as their primary language (n = 60), and the majority of those that did not reported Spanish as their primary language (n = 6). Nearly a quarter of participants reported proficiency in Spanish (22%). They participated in return for partial course credit. Informed consent was obtained prior to the start of the study. We excluded 4 participants who had previous knowledge of the study or due to mistakes in experiment implementation.

Setup and Materials

The experiment took place in an indoor room with two tables on opposite sides of the room (the Hide and Seek tables; Fig. 1). On each table was an identical row of four red cups, placed upside down. A tall grey divider on wheels hid the Seek table from the participants' view while the experimenter set up each trial; at the start of each trial, it was wheeled to the side to reveal the Seek table.

Procedures

Participants completed a rotation-based spatial search task. On each trial (N = 12), they were asked to observe an experimenter hide a small ball under one of four cups arranged in a row (Fig. 1, HIDE column). The experimenter then turned to the participant and said, "On the table behind me, there are four cups." On the initial *Gesture* trials (n = 8; Fig. 1A), the experimenter continued, "And under this cup, there is a prize," accompanied by a gesture that adopted either an egocentric or allocentric spatial framing to represent the ball's location (Fig. 1, TELL column). The gestural frame of reference was manipulated between-groups and counterbalanced. The experimenter then asked the participant to turn 180º to find the prize on the table behind them "in the same location." The participant then turned to the Seek table and searched under one of the cups for a hidden ball (Fig. 1, SEEK column). Before the start of each trial, the experimenter hid balls under two of the four cups on the Seek table, which corresponded to the egocentric and the allocentric search locations. Thus, so long as participants searched under a cup that would correspond to either an allocentric or an egocentric interpretation of "the same location," they would find a prize.

The Gesture trials were followed by No Gesture trials (n = 4; Fig. 1B). These were identical to the Gesture trials except the experimenter never gestured. These trials allowed us to test whether the effects of the gesture manipulation would persist even after the experimenter stopped gesturing. If the experimenter's gestures caused participants to adopt a particular frame of reference, then they should continue to use that frame of reference even when no longer directly primed by gesture.

Analyses

If participants searched locations at random, they would choose the egocentric location 25% of the time. However, participants seldom searched under cups that did not correspond to either an allocentric or an egocentric interpretation (n = 11 of all 840 trials). If participants only searched the egocentric or allocentric locations, they would select the egocentric interpretation half the time (50%). As a result, egocentric and allocentric responses were nearly complimentary; if participants did not search egocentrically, they searched allocentrically, and vice versa.

We thus analyze the effect of the experimenter's gesture on participants' spatial reasoning with a multilevel Bayesian logistic model. The model predicted trial-by-trial search strategy (egocentric, yes or no), with fixed predictors for gesture condition (allocentric vs. egocentric), trial type (gesture vs. no-gesture), and their interaction, along with random slopes by participants.

Results

Systematic and internally consistent spatial frames of reference

Overall, participants adopted either an allocentric or an egocentric search strategy on nearly all trials (Fig 2.). Since there were four cups under which they could search, but only two corresponded to egocentric or allocentric solutions, random responding would only search the allocentric or egocentric target locations half the time. Instead, participants searched in the allocentric or egocentric location on 99% of trials. Participants were slightly more likely to adopt an egocentric frame of reference (M = .55, ± .07 SEM) than an allocentric frame of reference (M = .44, ± .07 SEM). Participants were also highly systematic and internally consistent in their preference for spatial frames of reference: 87% of participants used their preferred spatial frame on every trial or every trial but one.

Observing gesture shaped participants' spatial frames of reference

We next investigated whether the gesture used by the experimenter to describe the target location shaped the participants' search behavior. Participants' search strategy was systematically shaped by the experimenter's gestures

(Fig. 3). As predicted, when the target location was described using an egocentric gesture, participants mostly searched in the egocentric location (M = .70, ± .07 SEM)). Conversely, when the target location was described using an allocentric gesture, participants mostly searched in the allocentric location (M = .58, ± .09 SEM).

The effect of gesture was confirmed by a multilevel Bayesian logistic model. Observing an egocentric gesture made participants significantly more likely to search egocentrically (b = 8.93, 95% Bayesian Credible Interval = [3.63, 15.86]). The 95% credible interval of the effect of gesture does not include 0, indicating a high probability that gesture shaped subsequent search.

Stable spatial frames of reference without gestural cueing

The effect of gesture persisted even after the experimenter stopped gesturing (i.e., the final n = 4 trials). Although the effect of gesture remained credibly greater than zero even after the experimenter stopped gesturing, the size of effect did shrink (b = -3.01, 95% Bayesian Credible Interval = [-4.81, -1.31]) (Fig. 4B). Thus, the effect of co-speech gesture was greatest on search behavior immediately after observing the gesture, but nevertheless persisted even after the gestures stopped (Fig. 4.). Participants in the egocentric condition were most likely to search in the egocentric location (M = .65, ± .08 SEM), while participants in the allocentric condition were most likely to search in the allocentric location (M = .56 ± .08 SEM). This suggests that participants adopted a particular spatial frame of reference for the task as a whole, not just on a trial-by-trial basis. We elaborate on this in the Discussion below.

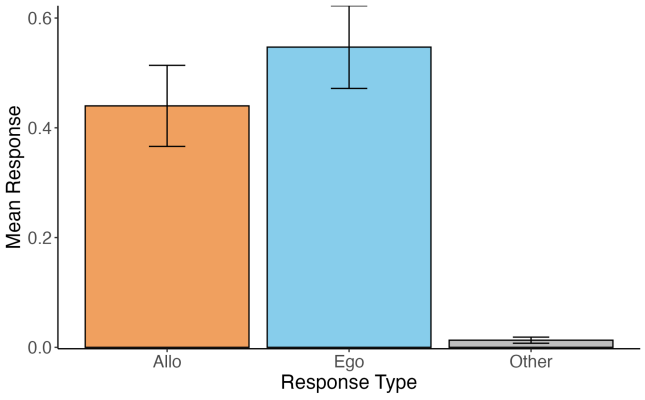

Fig. 2: Participants reliably adopted either allocentric or egocentric strategies. Responses that did not correspond to either frame of reference were uncommon.

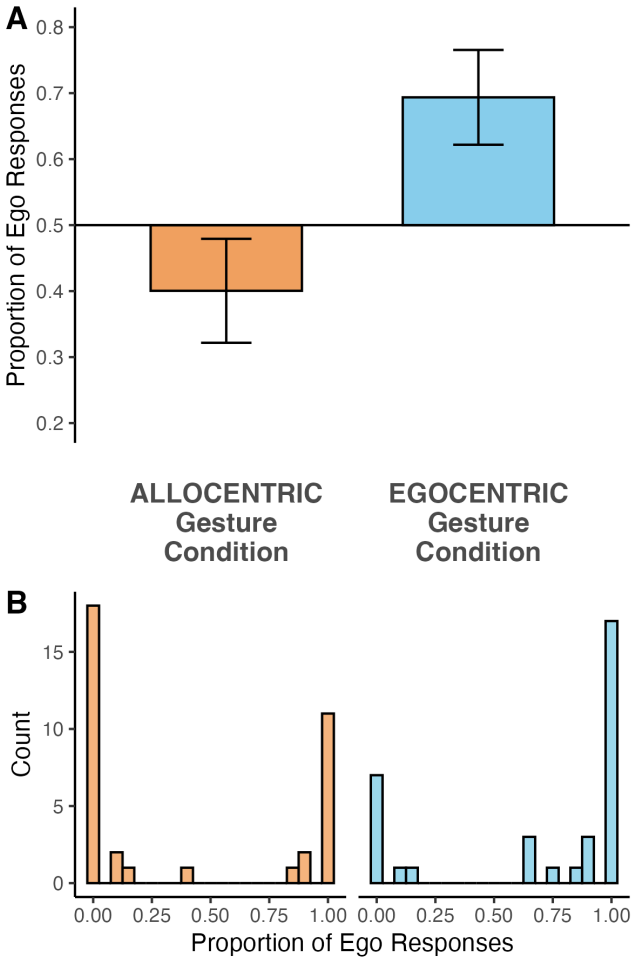

Fig. 3: Co-speech gesture influenced the frame of reference adopted by observers. (A) Proportion of egocentric responses for both the allocentric and egocentric gesture conditions. The horizontal line indicates the baseline of randomly selecting the egocentric or allocentric location. (B) Individual differences in egocentric search, split by gesture condition. Participants in the allocentric condition (orange, left) were most likely to respond allocentrically on all trials. Participants in the egocentric condition (blue, right) were most likely to respond egocentrically on all trials.

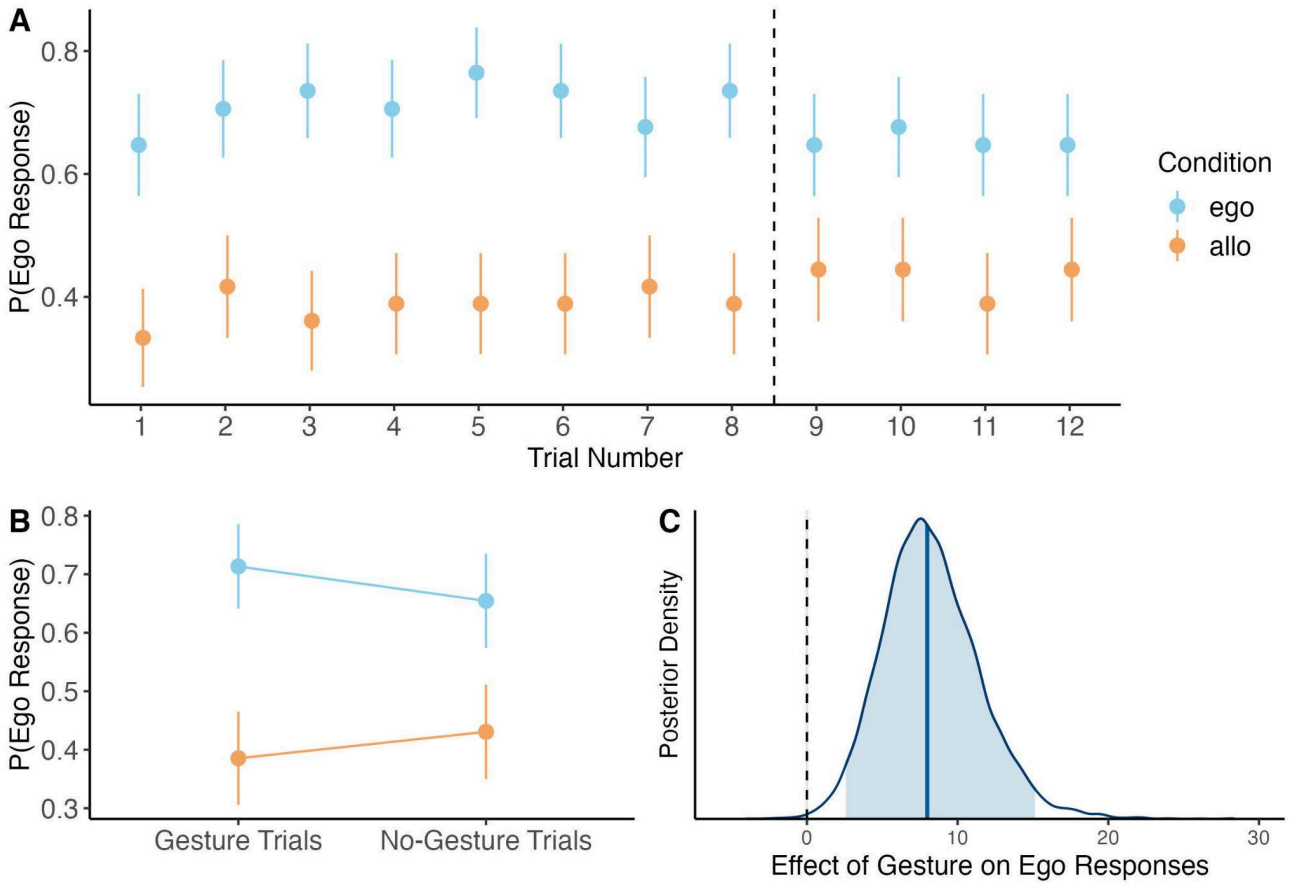

Fig. 4: Gesture had a robust and lasting effect on observers' spatial cognition. (A) Mean egocentric responses across all 12 trials for both gesture conditions. The dotted line divides the Gesture trials from the No-Gesture trials. (B) Collapsing across trials, but grouping by whether the experimenter gestured on the trial. The effect of the frame of reference adopted in gesture in the initial trials carried over to trials that did not include gesture. (C) Posterior density of the effect gesture on observers' frame of reference, specifically during the trials that included gesture.

Discussion

We set out to investigate whether a speaker's co-speech gesture could shape the spatial frame of reference adopted by the speaker's interlocutors. As predicted, observers reliably adopted the spatial frame of reference adopted in a speaker's gestures. When target locations were encoded in gesture egocentrically, participants searched in the corresponding egocentric locations; when target locations were encoded in gesture allocentrically, participants searched in the corresponding allocentric locations. Co-speech gesture is thus a candidate pathway for the transmission of spatial frames of reference.

Gestural relativity explains dissociations of spatial language and spatial thought

These findings offer an explanation of the dissociation of spatial frames of reference encoded in language and thought in Yucatec Mayan and Ngigua-Spanish bilingual communities (Le Guen, 2011; Calderón et al., 2019). Members of these communities may learn to use one spatial frame of reference in spoken language but develop a preference for another in gesture. It may be that preferences for a particular spatial frame of reference in memory might be transmitted within a community via gesture. One way to account for this effect is to include the effects of gesture into the repertoire of communicative resources that shape cognition in non-communicative domains. By 'shape' here, we do not mean that gesture fully determines spatial thinking, but rather that gesture is causally influencing spatial thinking in lasting ways. Spatial cognition may be shaped by multiple communicative resources existing in a mutually constrained semiotic network. Each of the semiotic resources in the network — resources that include language, speech, and cultural artifacts — are at once independent of each other but also function together to signal and give shape to preferences in spatial framing. On this account, gesture and language are complementary mechanisms for shaping spatial cognition.

The capacity for gesture to (re)shape spatial cognition suggests a reformulation of the Whorfian Hypothesis, as

suggested by Calderón et al. (2019), that expands the influence of language to include other multimodal aspects of communication. This simply parallels a preexisting criterion for linguistic relativity: That habitual patterns of communication can affect cognition in non-linguistic domains. If encoding spatial frames of reference in communication within the gestural modality can shape how we reason about space, then we believe this qualifies as what we call *gestural relativity*.

What remains to be shown is that these effects of a speaker's gesture on their interlocutors' spatial cognition are lasting. Here we have shown that such effects persist even after the speaker has stopped gesturing. If gestural relativity is in fact one of the mechanisms by which community norms are perpetuated, then the effects of gestures that encode frame-of-reference information — or at least of habitual exposure to such gestures — must last beyond the interactional encounter. In ongoing work, we are testing the timescale of the gestural priming effects demonstrated here.

Learning to align spatial frames of reference

As speculated by Le Guen (2011), gesture might be one means by which spatial frames of reference can be transmitted to younger generations — a transmission pathway that complements language. The findings presented here suggest that this is possible, at least on relatively short timescales. This suggests that adapting to a community's preferred frame of reference could happen rather quickly. The task used here is, however, simple and constrained; habits of spatial thought in more complex and natural communicative contexts might take longer to establish.

Why might individuals bother to align their spatial frames of reference? One reason to align your spatial frame of reference (either implicitly or explicitly) with the most statistically likely option in your community is to reduce the effort required to communicate. Successful communication depends on common understanding (Clark, 1996), so there is a cost associated with adopting default preferences that differ from community norm. Sharing common ground on preferences for spatial framing should lead to more successful communication of spatial relationships.

A cognitive-ecological perspective

What we are advocating for is a cognitive-ecological perspective (Marghetis, et al., 2020), one in which the canny spatializer is embedded in a rich interwoven ecosystem of communication (Kendon, 2004), one of multiple semiotic resources — language, gesture, cultural artifacts, and visual environment, among others. Through this mutually constraining network of resources default preferences for spatializing in terms of the ego or without arise within a community and become the dominant means of communicating and thinking about spatial relationships.

If we adopt this ecological metaphor, we might wonder whether different "species" in the ecosystem — gesture, language, etc. — play qualitatively different roles. Are there keystone species, for instance? In natural ecosystems keystone species are species that play an outsized role in shaping the dynamics of the ecosystem as a whole. If gesture is a potential pathway of transmission, then what would this system look like without it? Are language and environment sufficient to explain all cross-cultural variation in spatial cognition? One reason to suspect that gesture might be especially influential within this ecosystem is that the gestural modality is intrinsically spatial. When we gesture, we move our hands through space, and when referring to object relationships in space, we are forced into adopting a particular perspective, one that is either egocentric or allocentric. Gesture may thus play an outsized role in shaping spatial cognition — situated at the top of the "semiotic food chain" of spatial reasoning.

Future directions

Our experimental design allowed us to investigate the impact of gesture on spatial framing even after gestural cueing ceased — suggesting that the effects of gesture on spatial reasoning last beyond the immediate referential context. Follow-up work could look at whether the effect of gesture persists not just to later trials of the same task but to entirely different tasks. This would support the conclusion that gesture can shape our internalized spatial models of object relationships. That is, participants are not merely mimicking motor movement of the experimenter on a trial-by-trial basis. Rather, they are adopting a coherent spatial frame of reference to make sense of space — even when no longer cued by an interlocutor's gesture.

The paradigm introduced here could investigate gestural relativity across cultures. The participants sampled here were primarily speakers of English, who tend to prefer egocentric frames of reference. Our procedure could be used to test the gestural relativity of spatial cognition in communities around the world — include those that speak languages that rely exclusively on allocentric frames of reference, that live in rural environments, or that rely on diverse cultural artifacts.

Conclusion

When we think about object relationships in space, we cannot help but adopt a particular spatial frame of reference. It is a necessary communicative window into our spatial world. As we engage in the cultural practices of our communities, we develop certain preferences for which window we choose. We have demonstrated here that co-speech gesture can shape these preferences. The way our hands give shape to patterns in spatial reasoning points toward a theory of *gestural relativity*: Habitually encoding spatial relationships in gesture using a particular frame of reference may transmit that frame of reference to others, shaping their preferences in potentially lasting ways.

References

Brown, P., & Levinson, S. C. (1993). “Uphill” and “Downhill” in Tzeltal. Journal of Linguistic Anthropology, 3(1), 46–74.

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.

Clark, H. H. (1996). Using language. Cambridge university press.

Haun, D. B., 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(46), 17568-17573.

Haviland, J. B. (1993). Anchoring, Iconicity, and Orientation in Guugu Yimithirr Pointing Gestures. Journal of Linguistic Anthropology, 3(1), 3–45.

Kendon, A. (2004). Gesture: Visible action as utterance. Cambridge University Press.

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(1), 16-32.

Le Guen, O. (2011). Speech and Gesture in Spatial Language and Cognition Among the Yucatec Mayas. Cognitive Science, 35(5), 905–938.

Levinson, S. C. (1996). Language and space. Annual review of Anthropology, 25(1), 353-382

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity.* 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(2), 155–188.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning.

Pitt, B., Carstensen, A., Gibson, E., & Piantadosi, S. (2021). Variation in spatial concepts: Different frames of reference on different axes. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 43, No. 43).

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(3), 108–114.

Marghetis, T., McComsey, M., & Cooperrider, K. (2020). Space in Hand and Mind: Gesture and Spatial Frames of Reference in Bilingual Mexico. Cognitive Science, 44(12), e12920.

Marghetis, Ortega, Holmes (2024). Ecological relativity of spatial cognition: Humans think about space egocentrically in urban environments. Proceedings of the 46th Annual Meeting of the Cognitive Science Society.