Use of spatial reference frames for motion events in Balinese co-speech gesture

Danielle Naegeli (d.s.u.naegeli@tilburguniversity.edu)
DCC, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands

David Peeters (d.g.t.peeters@tilburguniversity.edu)
DCC, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands

Emiel Krahmer (e.j.krahmer@tilburguniversity.edu)
DCC, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands

Marieke Schouwstra (m.schouwstra@uva.nl)
ILLC, University of Amsterdam, Spuistraat 134, 1012 VB Amsterdam, The Netherlands

Made Sri Satyawati (srisatyawati@unud.ac.id)
Fakultas Ilmu Budaya, Universitas Udayana, Jl. Pulau Nias No.13, Denpasar, Bali 80113, Indonesia

Connie de Vos (c.l.g.devos@tilburguniversity.edu)
DCC, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands

Abstract

Spatial cognition and spatial language are core sites for diversity, both within and across language communities. For instance, when describing motion events, speakers may anchor information either (egocentrically) to their body or (allocentrically) to geographical landmarks in the environment, through both speech and gesture. Here we investigate whether the use of such egocentric versus allocentric frames of reference in co-speech gesture indeed depends on both bodily and environmental axes. In a real-world experiment, members from the traditionally allocentric Balinese community were shown small-scale motion events and asked to retell them. To evaluate the potential influence of both types of axes on gestural frame of reference use, in a 2x2 between-participant design they were assigned to conditions that contrasted the body-anchored axis the motion events unfolded on with the underlying geographical environment-anchored axis. It was observed that the type of body-anchored axis significantly predicted frame of reference representation in participants' gestures, consistent with previous research. The type of environment-anchored axis, however, did not affect characteristics of participants' gestures. These findings advance our understanding of the intricate interplay between language, space, culture, and environment.

Keywords: frame of reference, co-speech gesture, multimodal communication, spatial information, motion events

Introduction

Whenever we describe an event or communicate the location of a referent, expressions like "on the left", "toward", or "in front of" critically rely on adopting a so-called Frame of Reference (henceforth: FoR). In languages such as English or Dutch speakers tend to predominantly adhere to an *egocentric* FoR (Majid et al., 2004), meaning that spatial observations are described relative to the perspective of the speaker. The use of an expression like "on the right" is indeed typically relative to the perspective of the person making the description. Any possible addressee, depending on their location, may have to perform some sort of mental rotation, acknowledging the perspective of the speaker, to understand the communicated information correctly.

Cross-cultural research has made clear that employing an egocentric FoR is not the only possible option. In fact, many communities across the globe use an *allocentric* FoR for spatial description (e.g., Cooperrider, Slotta, & Núñez, 2017; Levinson & Wilkins, 2006; Majid et al., 2004; Wassman & Dasen, 2006), relying on cardinal directions or geographical landmarks that are independent from the speaker's perspective. In these cases, spatial observations could include terminology such as "south of", "downhill", or "seaward". While language communities can vary in which FoR they predominantly use, several FoR may co-exist within a given community. For example, depending on the scale of the spatial array, different FoR might be preferred. As established above, the egocentric FoR is dominant in English and so egocentric descriptions for small-scale descriptions like "it's the book on the left of the vase" are the norm. However, for geographically larger scale descriptions English speakers, like many egocentric-dominant language users, often comfortably transition to using allocentric descriptions like "Germany is to the east of the Netherlands". This shows that even within communities different types of FoR are used, potentially based on size and configuration of a given spatial array.

An example of a community that predominantly uses an allocentric FoR, for both large- and small-scale descriptions, can be found on the island of Bali, Indonesia (e.g., Aryawibawa et al., 2018; Wassman & Dasen 2006). Traditionally, in Balinese culture space is conceptualised and divided along two axes anchored to environmental

landmarks, and this is reflected in the direction system of the Balinese language. One axis, *kangin-kauh*, can roughly be translated as "where the sun rises"-"where the sun sets" (see Wassman & Dasen, 2010 for in-depth discussion on the peculiarities of this axis). The other axis, *kelod-kaja*, can be translated as "seaward"-"mountainward", in association with the sea that surrounds Bali and the mountainous terrain around its largest volcano, Mount Agung, at the centre of the island. Therefore, *kelod* and *kaja* are relative to where on Bali the speaker is and are reversed for North and South Bali. The Balinese direction system is deeply intertwined with Balinese culture and religion. Temples, villages, and houses are built in adherence to the axes, and even one's position when sleeping is dictated by them (Aryawibawa et al., 2018).

Nevertheless, despite the cultural importance of the allocentric FoR, most Balinese people are also apt to using an egocentric FoR, especially in the more urban southern part of the island where there are generally many foreigners, Indonesians from different islands, and a mix of people from all over Bali (Dasen & Mishra, 2010). This competence in multiple FoR use is further supported by the linguistic situation. While Balinese is used in daily social life and for cultural practices, the national language of Indonesia, *Bahasa Indonesia*, used in schools and for official purposes, predominantly uses egocentric terms, as one may expect for a language spoken across a wide range of islands, each with their own specific landmarks (Dasen & Mishra, 2010). Therefore, many Balinese people are accustomed with making both allocentric and egocentric spatial descriptions.

Furthermore, while we can describe spatial relations using spatial terms like "*kelod*", "*kaja*", "left", or "right", we often also convey spatial information through bodily gestures accompanying our speech (Kendon, 2004). Previous research suggests that gestures reflect the type of FoR also dominant in the community's spoken language. Hence, a community predominantly using an allocentric FoR for spoken spatial descriptions may also gesture allocentrically (Calderón, De Pascale, & Adamou, 2019; Le Guen, 2011). Our gestures might even supplement spatial information that is absent from our spoken utterances. Kita and Özyürek (2003) found

that for motion events, while manner of the motion is typically conveyed in speech, information on the direction of motion is often omitted in the spoken description but present in a speaker's co-speech gestures.

Building on this, Marghetis et al. (2020) designed a task to investigate FoR use in co-speech gestures independent from spoken spatial language. They conducted their study in Juchitán de Zaragoza, Mexico with speakers of the Isthmus Zapotec, an allocentric FoR-dominant language, who had varying degrees of exposure and proficiency levels of Spanish. Participants watched small-scale motion events, physically unfolding in front of them, and were subsequently asked to recount the events when positioned at a 90° angle to the location from which they had observed the event. This way they anticipated that spatial information would be backgrounded resulting in little being conveyed verbally but substituted in the form of directional gestures. Participants were split into two conditions, depending on the directions the events happened. These unfolded along one of two body-anchored axes, either on the lateral axis (from left to right and right to left), or on the sagittal axis (away from the participant and towards the participant). The results showed that the type of FoR gestures adhered to differed due to two factors: participants' competence in using left-right distinctions and the axis on which the motion events unfolded. Specifically, participants were more likely to use egocentric gestures for descriptions of motion events that unfolded on the sagittal (vs the lateral) axis. This suggests that, similarly to array scale, the axis on which spatial information is located influences the choice of FoR and therefore, causes variation in FoR use within communities. Pitt et al. (2022) also conducted research focusing on the potential effect the different body-anchored axes have on the way people conceptualise space. Similar to Marghetis et al. (2020), they found that members of the Tsimane', another allocentric-dominant community indigenous to the Bolivian Amazon, showed varied FoR use depending on the body-anchored axis the information was presented on. In both a linguistic and non-linguistic task, participants showed a preference for an egocentric FoR for
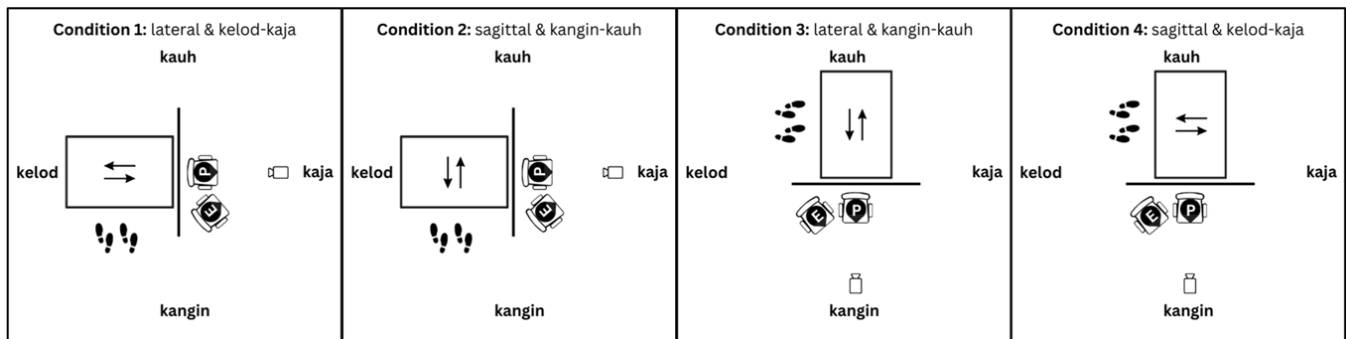

Figure 1: Experimental layout per condition with Balinese direction labels. "P" and "E" indicate the position of participant and experimenter during the video recording part of the trial. The footsteps indicate the event observation location. In conditions 1 and 3 events unfolded on the lateral axis (from left to right or from right to left); in conditions 2 and 4 on the sagittal axis (away from or towards the body).
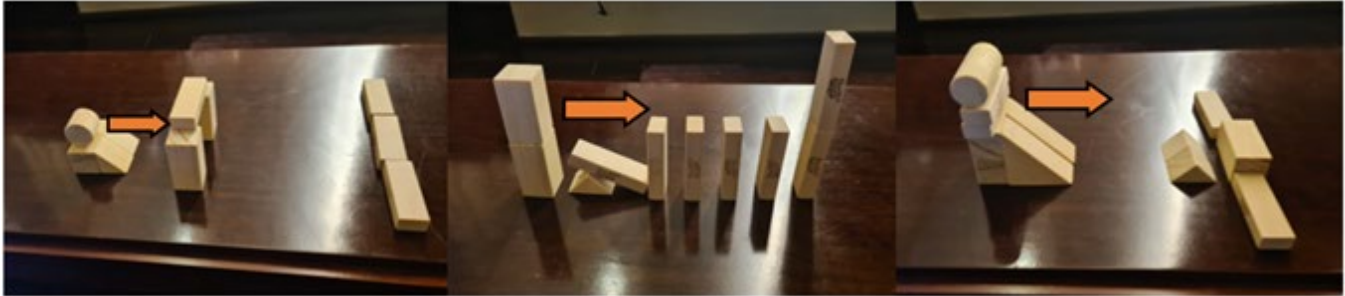
Figure 2: Examples of arrays used in the experiment. The arrows indicate the primary direction of the motion event.

sagittally presented information and an allocentric FoR for laterally presented information.

Both these studies hence focused on the body-anchored axes to explain variation in FoR use, however, there might be an additional component that could have influenced participants' behaviour. In both experiments, the body-anchored axes under investigation overlapped with the environment-anchored axes used in the local allocentric FoR. As the experiments were conducted in spaces familiar to the participants there must inevitably have been some overlap between the experimental setup and the allocentric FoR canonically used in the community. Such entrenchment could have also affected participants' behaviour and might be especially interesting when a difference in saliency between the two environment-anchored axes can be presumed. For a community like the one in Bali, one can imagine that the *kangin-kauh* axis, referencing the trajectory of the sun and therefore, being consistent in any location on the island, may be more salient than the *kelod-kaja* axis that shifts as a function of location. Thus, depending on the experimental setup and the overlap of body-anchored (lateral/sagittal) and environment-anchored (e.g., *kangin-kauh*/*kelod-kaja*) axes, differences in participants' behaviour might be mitigated or intensified. To investigate the influence of the two types of axes (body-anchored vs environment-anchored) independently from each other, we ran an experiment, similar to the one in Marghetis et al. (2020), in Bali. In addition to the original two conditions (differing on the body-anchored axis the events unfolded on) we added two conditions in which the entire experimental setup was rotated by 90°. With this setup rotation and the placement of the experimental setup in general we made sure that each body-anchored axis overlapped once with each environment-anchored axis. In line with the findings of both Marghetis et al. (2020) and Pitt et al. (2022), we expected participants to use gestures adhering to different FoR depending on the body-anchored axis the motion event unfolded on. We predicted to find more egocentric gestures on the sagittal axis and more allocentric gestures on the lateral axis. Regarding the environment-anchored axes, we hypothesised that in conditions where the motion event would unfold on the *kangin-kauh* axis we would see more allocentric gestures as it is the more constant and therefore more salient environment-anchored axis.

Methods

Participants
Twenty-four participants (14 female, 10 male) took part in the experiment. They were all students at Universitas Udayana in Bali, Indonesia and their average age was 20 years old (age range 18-23). They were recruited by the local research assistants and university staff members. To participate, students had to be native speakers of Balinese. As compensation for their time, participants received 100'000 Indonesian rupiah.

Materials and Setup
In the experiment, participants saw short, small-scale motion events. The motion events consisted of arrays of wooden blocks in different shapes and sizes which were manipulated to roll or fall over. The block array for each event was built by the experimenter on a rectangular table out of view of the participant. The array was different for each trial, totaling in 12 different motion events that were copied as closely as the material allowed from Marghetis et al. (2020). Each motion event was designed to have one primary direction of motion.

Procedure
The experiment was conducted indoors, in a classroom at Universitas Udayana in Denpasar, Bali. The door was closed and some of the curtains had to be drawn to allow for decent lighting in the video recording. The stimulus events and procedure of this experiment were inspired by Marghetis et al. (2020). The experimental setup comprised of a table and two chairs on one side of the table with a fabric sheet separating the table and the seating area (see Figures 1 and 3). The block arrays for the motion events were built on the table between trials by an experimenter. When seated, participants had their backs to the table while facing a camera with the experimenter sitting at slight angle. The experimenter and the participant observed each motion event from the side of the table. The "event observation location" and the seats were at a 90° angle to each other. Participants observed a total of 12 motion events, in one of two orders. For each trial, the motion event started with a block or cylinder that was pushed or knocked over. Following Marghetis et al. (2020), the first six trials were initiated by the experimenter and in the following six trials the

Figure 3: Stills of participants gesturing with experimenter on the side. The participant in the first two pictures from the left is gesturing motion on the sagittal axis, first away from the body (first picture) and then towards the body (second picture). The participant in the third and fourth picture from the left is gesturing motion starting from his right (third picture) and starting from his left (fourth picture).

participants were instructed how to start the motion event and did so themselves. Once the event had unfolded, the experimenter covered the blocks with a piece of fabric and moved back to the seating area with the participant where the participant took a seat. The experimenter then started the recording on a laptop connected to the camera, took a seat on the chair next to the participant, and prompted the participant to start the retelling. The experimenter asked in Balinese what the participant had seen, without mentioning gesture. Participants retold the events in Balinese and the experimenter reminded them in case they switched into Bahasa Indonesia. Participants were all native speakers of Balinese; however, many young Balinese people do not speak Balinese consistently throughout their everyday lives but mix it with Bahasa Indonesia.

As discussed in the Introduction, in the original experiment in Marghetis et al. (2020) participants were assigned to one of two conditions depending on whether they saw the motion events unfold on the lateral or the sagittal axis. We added two conditions in which participants saw events unfold on one of the two body-anchored axes, but the experimental setup was rotated by 90° to control for the motion event axes overlapping with the Balinese environment-anchored axes (*kangin-kauh/kelod-kaja*). Both experimental setups were built to be parallel with the environment-anchored axes (see Figure 1). In all four conditions, directions alternated from trial to trial. As such, participants in the lateral axis conditions saw trials going from left to right, then right to left, and so forth, while participants in the sagittal axis conditions saw them going away from them and towards them. Each of the four conditions comprised of six participants in a between-participant design. This resulted in a total of 72 trials per condition.

Gesture coding and data preprocessing

A native Balinese research assistant transcribed the audio of each trial video, translated word-for-word and literally into English, and annotated occurrence and direction of motion gestures. These gestures were then checked by a second researcher (the first author), while a third researcher (an

Indonesian national experienced in data collection on Bali but unfamiliar with the study at hand) was consulted for difficult cases to ensure only motion gestures describing the primary motion event were included in the analysis. Gestures that occurred when participants described e.g., how the experimenter covered the array after the event were excluded from the analysis. Gesture directions were assigned one of five possible variables: ALLO (allocentric gesture), EGO (egocentric gesture), MULTI (multiple directions describing the primary event present in one trial), OTHER (direction of gesture adhered neither to ego- nor allocentric FoR), and NA (for trials in which there was no directional gesture).
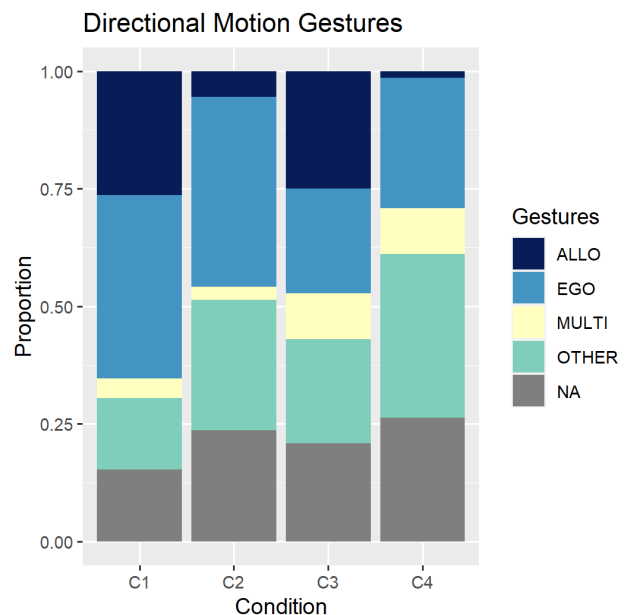

Figure 4: Stacked barplot of proportional distribution of types of motion gestures by condition. Participants in conditions 1 and 3 saw motion events on the lateral axis and participants in conditions 2 and 4 saw them on the sagittal axis. The setup was rotated by 90° for conditions 3 and 4.

Results

We collected data in a total of 288 (4 x 72) trials. Only in 4.9% of the trials ($n = 14$), participants used verbal language such as "to the left" or "from the block closest to the North" to provide information on the direction of the motion event. This is consistent with the previously stated observation that the directions of motion events are often omitted in spoken descriptions. Clear directional gestures were present in 78.5% of trials (226 of 288), meaning that in 21.5% of trials no co-speech gestures at all or no identifiable directional gestures were produced. Of the trials with gestures indicating direction of motion, 19 (8.4%) included more than one direction and 72 (31.9%) contained gestures that were consistent neither with an egocentric nor an allocentric FoR.

Allocentric gestures were used in a total of 42 trials and egocentric gestures in 93 trials (respectively 18.6% and 41.2% of the trials containing directional motion gestures).

Figure 4 shows the proportional use of directional gesture by condition. In conditions 1 and 3 participants saw events on the lateral axis and in conditions 2 and 4 events happened on the sagittal axis. Regarding the environment-anchored axes, in conditions 1 and 4 events unfolded along the *kelod-kaja* axis and in conditions 2 and 3 along the *kangin-kauh* axis. From this graph we can infer that participants in conditions 1 and 3 behaved relatively similarly, as well as participants in conditions 2 and 4. These conditions share the same body-anchored event axes. Participants in the lateral axis conditions used numerically more allocentric gestures than participants in sagittal axis conditions. In condition 1 and condition 3 allocentric gestures were used in 19 and 18 trials respectively. Allocentric gestures were used in 4 trials in condition 2 and 1 trial in condition 4. In contrast, there is no visible clustering in the conditions sharing the same environmental axis (conditions 1 and 4, and conditions 2 and 3, respectively). This suggests that the body-anchored axes could have had an influence on which FoR participants adopted for their gestures while the environment-anchored axes did not.

Analysis

We ran a logistic generalized mixed-effects regression model using the lme4 package in R (Bates et al., 2015; R Core Team, 2022, v4.2.2). Only trials with egocentric and allocentric gestures were included (a total of 135 trials), coded binomially, and are henceforth referred to as the dependent variable GestureType. We included body-anchored axes (lateral/sagittal) and environment-anchored axes (*kangin-kauh*/*kelod-kaja*) as fixed effects, both sum-coded. We also included random by-participant and by-item intercepts (see Table 1).

In line with Figure 4, the model confirmed that the body-anchored axes were a predictor of FoR use as we found a significant main effect of body-anchored axes ($p = .005$). This indicated that there was a significantly lower proportion of egocentric gesture use in lateral-axis conditions ($M = 0.543$, $sd = 0.501$) compared to sagittal-axis conditions ($M = 0.907$, $sd = 0.292$). In turn, in light of the binomial nature of the data, this meant that there was a relative preference for allocentric gestures in lateral conditions. The environment-anchored axes, however, did not predict the use of FoR. The distribution of egocentric/allocentric gestures did not differ significantly ($p = 0.937$) between *kangin-kauh* conditions ($M = 0.655$, $sd = 0.48$) and *kelod-kaja* conditions ($M = 0.713$, $sd = 0.455$). There was also no significant interaction effect between the two types of axes ($p = 0.489$).

Discussion

With this study we have investigated the possible influence of both body-anchored and environment-anchored axes on the use of FoR in co-speech gesture of Balinese participants. Previous research has suggested that the type of body-anchored axis on which spatial information is situated, or events happen on, influences which FoR people adopt when describing them (Marghetis et al., 2020; Pitt et al., 2022). While these previous studies were conducted in communities that exhibit an overall preference for an allocentric FoR, neither of them addressed whether reference frame use was influenced not just by the speakers' body-anchored axes but also the underlying environment-anchored axes of the allocentric spatial system at hand. To control for this and systematically investigate the influence of both types of axes we conducted an expanded version of the small motion events task in Marghetis et al. (2020). Our version took the form of a 2x2 design in which the axes the motion events happened on were parallel to and overlapped with one of the environment-anchored axes of the Balinese spatial FoR or the other. We did this by showing participants events on the lateral or the sagittal axis (as in Marghetis et al., 2020) and by rotation of the entire experimental setup by 90° for two of the four conditions.

In a nutshell, our findings are in line with results from previous research. Participants exhibited variation in FoR use for the directional motion gestures depending on which body-anchored axis the events unfolded on. In the lateral conditions, participants were relatively more likely to use

Table 1: Outcome of logistic generalised mixed-effects regression model.
Model: GestureType ~ BodyAnchoredAxes*EnvironmentAnchoredAxes + (1|Participant) + (1| Item)

	Estimate Std	Error	z value	Pr (>|z|)
(Intercept)	2.141	0.775	2.764	0.005**
BodyAnchoredAxes	-3.829	1.373	-2.788	0.005**
EnvironmentAnchoredAxes	0.094	1.201	0.79	0.937
BodyAnchoredAxes x EnvironmentAnchoredAxes	1.643	2.374	0.692	0.489

allocentric (vs egocentric) gestures (consistent with Pitt et al., 2022), while in the sagittal conditions, participants were more likely to use egocentric (vs allocentric) gestures (consistent with both Marghetis et al., 2020 and Pitt et al., 2022).

We note that using allocentric FoR descriptions for lateral spatial information and egocentric FoR descriptions for sagittal spatial information both results in participants showing the direction of motion events on the axis away from and towards their body, therefore the sagittal axis. Both Marghetis et al. (2020) and Pitt et al. (2022) suggest that this is due to the relative difficulty of lateral distinctions. While the lateral axis is symmetric, the sagittal axis is asymmetric and can therefore be considered more salient (Clark, 1973; Tversky, 2011). Egocentric terms like "left" and "right" are only acquired by members of communities in which they are used and even then, children acquire them quite late (Shusterman & Li, 2016). This is also supported in Marghetis et al. (2020) by participant's familiarity with using "left" and "right" being a predictor of their FoR use. Our participants, however, generally used a lot of egocentric gestures even in the lateral conditions (44 compared to 49 in the sagittal conditions) where left-right distinctions are required. In contrast to Marghetis et al. (2020) all of our participants were proficient speakers of a language that commonly uses egocentric terms (Bahasa Indonesia), and therefore an assessment of their comprehension of egocentric terms and a subsequent analysis on the basis of those scores was not needed.

Contrary to our hypothesis on the influence of the environment-anchored axes, we did not find environment-anchored axes to be a significant predictor of FoR use. In essence, this confirms that the environment-anchored axes may not have been a confounding factor in previous research. We had hypothesised that in conditions where the primary event axis overlaps with the *kangin-kauh* axis we would see an increase in allocentric FoR use. Our reasoning behind this was that the *kangin-kauh* axis is based on the locations of the sun rising and setting, and therefore more consistent and more salient than the *kelod-kaja* axis that depends on one's location relative to topographic landmarks. If true, and combined with our findings on the body-anchored axes, this should be especially strong for when the events are shown on the lateral axis, and it is parallel to *kangin-kauh*. As already mentioned, we did not find a statistical influence of environment-anchored axes on FoR use, and neither was there a numerical increase in allocentric gestures in condition 3 compared to the other conditions. We did however find that participants in condition 3 showed the lowest use of egocentric gestures across all conditions. The indoors setting we used had little view of the outside and no culturally significant objects, which might make it difficult for Balinese participants to orient themselves along their allocentric FoR. Future research will investigate whether the current results will replicate in an outdoor setting where the environment-anchored axes are naturally more prominent. As such, the intricate interplay between language, space, culture, and environment can be further elucidated.

References

Aryawibawa, I. N., Putra Yadnya, I. B., Ngurah Parthama, I. G., & Pye, C. L. (2018). Balinese spatial reference frames: Linguistic and non-linguistic evidence from the north of Bali. *Lingua, 215,* 40-52.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software 67*(1), 1-48.

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. (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., Slotta J., & Núñez R. (2017). Uphill and downhill in a flat world: The conceptual topography of the yupno house. *Cognitive Science, 41,* 768–799.

Dasen, P. R., & Mishra, R. C. (2010). *Development of geocentric spatial language and cognition: An eco-cultural perspective.* Cambridge University Press.

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,* 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., & Wilkins, D. (2006). *Grammars of space,* Cambridge University Press.

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., McComsey, M., & Cooperrider, K. (2020). Space in hand and mind: Gesture and spatial frames of reference in bilingual Mexico. *Cognitive Science, 44*, e12920.

Pitt, B., Carstensen, A., Boni, I., Piantadosi, S. T., & Gibson, E. (2022). Different reference frames on different axes:

Space and language in indigenous Amazonians. *Science Advances, 8*(47), eabp9814.

R Core Team (2022). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

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

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.

Wassmann, J., & Dasen, P. R. (2006). How to Orient Yourself in Balinese Space: Combining Ethnographic and Psychological Methods for the Study of Cognitive Processes. In J. Straub, D. Weidemann, C. Köbl, and B. Zielke (Eds.), *Pursuit of Meaning. Advances in Cultural and Cross-Cultural Psychology* (pp. 351-376). Bielefeld: Transcript Publishers.