Variations in explainers' gesture deixis in explanations related to the monitoring of explainees' understanding

Stefan Lazarov (stefan.lazarov@uni-paderborn.de)
Faculty of Arts and Humanities, Psycholinguistics Group
& TRR 318 "Constructing Explainability", Paderborn University
33098 Paderborn, Germany

Angela Grimminger (angela.grimminger@uni-paderborn.de)
Faculty of Arts and Humanities, Psycholinguistics Group
& TRR 318 "Constructing Explainability", Paderborn University
33098 Paderborn, Germany

Abstract

In this study on the use of gesture deixis during explanations, a sample of 24 videorecorded dyadic interactions of a board game explanation was analyzed. The relation between the use of gesture deixis by different explainers and their interpretation of explainees' understanding was investigated. In addition, we describe explainers' intra-individual variations related to their interactions with three different explainees consecutively. While we did not find a relation between interpretations of explainees' complete understanding and a decrease in explainers' use of gesture deixis, we demonstrated that the overall use of gesture deixis is related to the process of interactional monitoring and the attendance of a different explainee.

Keywords: explanation; gesture deixis; monitoring; understanding

Introduction

Explanations are co-constructive interactions in which an explainer provides a less-knowledgeable person (*explainee*) with information about an entity or a process (*explanandum*) to increase their knowledge and understanding (Rohlfing et al., 2021). To increase explainees' knowledge and understanding, and to resolve understanding-related problems, explainers use verbal and non-verbal modes of communication, such as speech and gestures, simultaneously. Both modalities form an integrated system, which becomes apparent in the tight temporal and semantic coupling (Kendon, 2004; Kita, 2009; McNeill, 2005). Co-speech gestures, which express semantically related content to the spoken parts of utterances, can provide interactional guidance and support understanding via pointing, representing and highlighting certain aspects (de Ruiter, 2000). Although previous empirical research has provided findings about gestures' role in contributing to addressees' understanding (Congdon et al., 2017; Habets et al., 2011; Kelly et al., 2004; Kelly et al., 2010), it is not yet known how explainers use particular gestural functions in relation to their interpretations of explainees' different levels of understanding.

In this paper, we address this open question and focus particularly on explainers' use of gesture deixis in board game explanations in the physical absence of an explanandum, i.e., the board game. Because the game was not present first, the explainers organized the interaction by applying skills of memories and constructive imagination (Bühler, 1982; West, 2014). The goals of our study are to discover 1) how gesture deixis is used by different explainers in the temporal relation to their interpretations of explainees' understanding (assessed retrospectively), and 2) how this relation could be explained by explorations of explainers' intra-individual gestural behavior during interactions with different explainees. For this purpose, we analyzed the behavior of eight explainers, each of them explaining a board game to three different explainees consecutively (in total, 24 explanatory dialogues). We want to clarify that even though the present study focusses on gestures, the analyzed gestural forms co-occurred with speech in a natural dialogue situation.

The different dimensions of gestures

The absence of a physical explanandum may hamper addressees' comprehension of the spatial organization of unknown objects. Speech and gesture deixis play an essential role in solving the problem of spatial orientation (Bühler, 1965). Co-speech gestures may serve different functions such as highlighting or drawing on a surface shared by the interlocutors. For example, drawing invisible objects by performing gestures is essential in the successful establishment of joint imagined spaces (Kinalzik & Heller, 2020). In cases when the explanandum is absent from the shared referential space, an imaginary presentation of the explanandum by explainers' pointing and drawing behavior may be required. Therefore, applying McNeill's (2006) assumption that gestures represent multidimensional functions ("iconicity", "metaphoricity", deixis, "temporal highlighting" (for beats), "social interactivity", p. 301) to our coding seems to be more appropriate than an application of McNeill's (1992) classical formal and functional categorization of the different gesture types. The current study focuses concretely on the dimension of deixis together with other dimensions (e.g., deixis and iconicity, or deixis and highlighting) and includes hybrid gestural forms.

In general, deictic gestures represent behavior (such as pointing using extensible body parts) which establishes an

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

indexical link between a reference and a referent (McNeill, 1992; de Ruiter, 2000). They aim at attracting interlocutors' attention and at contributing to the understanding of spoken references (Clark, 2003; Stojnic et al., 2013). Regarding McNeill's dimensions, gesture iconicity represents certain features of a referent and is semantically related to the co-occurring speech (de Ruiter, 2000; McNeill, 1992; Poggi, 2008). Like gesture deixis, gesture iconicity contributes to the attraction of addressee's attention, and to their memory recall and comprehension (Dargue et al., 2021; Kandana-Arachchige et al., 2021; McKern et al., 2021). In contrast to the dimensions of deixis and iconicity, temporal highlighting, realized by beat gestures, does not convey semantic information, but it emphasizes information by being temporally aligned with a related part of a spoken utterance and with prosodic marking (Beege et al., 2020; Dimitrova et al., 2016). Some research has reported that beat gestures may contribute to understanding, however at a much lower degree than deixis or iconicity do in native speaking contexts (Austin & Sweller, 2014; Dimitrova et al., 2016; Rohrer et al., 2020). Thus, we investigate the dimension of gesture deixis in observable hybrid forms of co-speech gesture categories to account for the multidimensionality of gestures. Together with other dimensions, such as iconicity and highlighting, we relate explainers' use of gesture deixis to their interpretations of explainees' understanding.

Gestures and comprehension

Previous research relating speakers' co-speech gestures to addressees' understanding has shown that gestures have a general positive effect on understanding (Congdon et al., 2017; Grimminger et al., 2010). As mentioned in the previous section, gesture deixis and iconicity bear semantic information, i.e., they convey meaning (McNeill, 1992; 2006). The semantic congruency between gestures and speech has been also related to a faster reaction time and gesture interpretation, during addressees' observation of speakers' gestures (Habets et al., 2011; Kelly et al., 2010; Ping et al., 2013). Furthermore, observing gestures has been shown to reduce learners' cognitive load and foster social engagement (Li et al., 2021). Although studies have found that co-speech gestures can increase understanding, it is not yet clear how the use of deictic gestures (and hybrid forms) by explainers is related to the dynamics of explainers' interpretations of explainees' understanding, also with respect to interacting with different explainees consecutively.

Monitoring explainees' understanding

Understanding is defined as a cognitive process with gradual qualities (levels) ranging between non-understanding, partial understanding and complete understanding (Bazzanella & Damiano, 1999; Vendler, 1994). In addition to the levels of non-understanding and partial understanding, there is another state, misunderstanding, which refers to an incorrect reception of information. Misunderstandings could be resolved after a detection of the problem and the initiation of a repair by the explainer (Vendler, 1994).

In interaction processes, interlocutors monitor each other's (non-)verbal behavior continuously and elicit information about the achieved level of understanding of an explanandum (Clark & Krych, 2004). Because levels of understanding are gradually changing, monitoring explainees' (non-)verbal signals by explainers could lead to a dynamic variation of explainers' strategies of explaining, including variations in gesturing. Following this assumption, a dynamic variation in gesturing may be observed in relation to explainers' interpretations of explainees' understanding. Monitoring explainees' understanding could be a challenging task for explainers due to the possibility of misinterpretations of explainees' (non-)verbal feedback. Previous research on the interpretations of (non-)verbal feedback signals has shown that (non-)lexical backchannels (Allwood et al., 1992; Arnold, 2012; Bavelas et al., 2000; Ward & Tsukahara, 2000; Yngve, 1970) and head nods (Allwood & Cerrato, 2003; Gander & Gander, 2020) evoke ambiguous interpretations towards either unconditional understanding or solely attention. Furthermore, gaze aversions from an explaining interlocutor can be misinterpreted by explainers as disengagement from a task (Doherty-Sneddon & Phelps, 2007; Jongerius et al., 2022) rather than as a signal of ongoing cognitive processing (Glenberg et al., 1998). Even though explainees' various multimodal signals may lead to misinterpretations because they have been reported to be ambiguous, it is yet interesting how explainers' interpretations of different levels of understanding may be related to characteristics of gesture use on the dimension of deixis.

One way of documenting explainers' interpretations of explainees' understanding in explanatory dialogues moment by moment is the collection of protocolled retrospective accounts from the explainers (Kuusela & Paul, 2000). An applicable related procedure is the conduction of video-recall. Video-recall is a post-test procedure after the main interaction study which aims at stimulating interaction partners' short-term memory of an interaction that has already taken place. Video-recalls can be conducted, for example, by presenting a videorecording of an interaction to the interaction partners and providing the participants with instructions about the demanded focus on specific aspects and events of an explanation (see Methods for a detailed description of the video-recall procedure in this study).

The individuality of gestural behavior

In addition to the relation between explainers' gesture deixis and their interpretations of explainees' understanding, we are also interested in the individual behavior of each explainer towards three different explainees. Previous research on formal gesture features, such as form and path, has shown that gesturing is idiosyncratic, i.e., speaker-individual (Bergmann & Kopp, 2009; Priesters & Mittelberg, 2013). However, Bergmann & Kopp (2009) suggest that the idiosyncratic gesture production by different speakers may also vary in relation to the dialogue situation and the presence of a different addressee. Further, individuals' higher gesture

rates have been reported when there is a greater the degree of expertise between interlocutors (Holler & Stevens, 2007; Jacobs & Garnham, 2007; Kang et al., 2015), or when the explanandum is not present during an explanation (Holler & Stevens, 2007). Based on the previous findings on individual gesturing behavior, we would like to extend the research on this topic by describing the intra-individual variations of different explainers' gesture deixis related to their interpretations of the levels of understanding of three different explainees.

Hypotheses

In the present study, we investigate the dynamics in explainers' gesture deixis in relation to the monitoring of explainees' levels of understanding. Because explainers were required to organize their explanations in the physical absence of an explanandum, also drawing on memories and imagination (Bühler, 1982; West, 2014) about the spatial organization of the board game, we expected the occurrence of hybrid gesture forms combining gesture deixis with iconicity (e.g., drawing) or highlighting.

Based on previous studies on the comprehension providing function of gestures (Congdon et al., 2017; Grimminger et al., 2010; Kang et al., 2015), and specifically on deictic (Clark, 2003; Stojnic et al., 2013) and iconic gestures (Dargue et al., 2021; Kandana-Arachchige et al., 2021; McKern et al., 2021), we assume that explainers change the frequency of their gestures based on their interpretations of explainees' understanding. Although previous studies have analyzed gesture use in experimental conditions in which speech is less accessible or less informative, we assume that this also accounts for naturalistic conversation settings, such as those in the present study. We hypothesized that:

(1) Following explainers' interpretation of explainees' complete understanding, explainers' gesture deixis decreases while following explainers' interpretations of explainees' non-, partial or misunderstanding, explainers' gesture deixis increases.

Second, we are interested in intra-individual differences in forms of gesture deixis. Because of the scarcity of empirical work on speakers' gesturing related to interpretations of addressees' (levels of) understanding, this is addressed in an exploratory manner. Based on the previous findings on the individual use of gestures by different speakers (Bergmann & Kopp, 2009; Priesters & Mittelberg, 2013) and variations depending on the addressee (Holler & Stevens, 2007; Jacobs & Garnham, 2007; Kang et al., 2015), we hypothesized that:

(2) The gesture deixis of individual explainers varies depending on the attendance of a different explainee.

We will explore the effect of three different explainees on the gesture deixis of one explainer.

Methods

Data Corpus and Procedure

The sample analyzed in the present study has been randomly selected from the MUNDEX corpus ("Multimodal understanding of explanations") (Türk et al., 2023). MUNDEX is a large video-corpus which contains 87 dyadic, explanatory interactions about the board game *Deep Sea Adventure* in German language. It has been collected to investigate the monitoring of multimodal signals of understanding of explanations.

Dyadic interactions The interactions were videorecorded from six different camera angles (two at each participant's face area, two directed towards each participant's torso, hands and head, one side angle, and one top angle over both interaction partners). The speech of both interlocutors was additionally audio-recorded with individual headsets. In the dyadic interactions, an explainer explained a board game either to three or two explainees consecutively. The interlocutors were unknown to one another. The game was given to the explainers one or two days prior to the study, so that they could learn it on their own. No guided instructions as how to learn the game or additional instructions of the game were provided to them by the experimenters in order to avoid modeling a way of explaining the game during the study. All explainers were thus free to organize the explanations by themselves without any guidance by the experimenters because the study focused on explanatory phenomena natural conversations. The only guidance that the explainers received was to begin the explanations without presenting the board game to the explainees, then to freely choose the moment at which they present the board game to the explainees, and finally to play the game interactively. Thus, each interaction consists of three timely varying phases: game absent, game present and a game play. For the present analysis, we randomly selected eight different explainers, resulting in 24 explanations in total. The mean duration of all 24 explanations overall (incl. all three phases) was 26:49 min (*SD* = 05:30 min). The mean duration of the analyzed phases with the board game absent was 07:04 min (*SD* = 03:44 min).

Video-recall task Following the dyadic interactions, each explainer and each explainee took part in a video-recall task, in which they individually watched the recorded dyadic interaction (side angle camera). Before this task, both interaction partners were instructed to comment on any moment from the interaction for which they recognize explainees' (for the explainers) or their own (for the explainees) different levels of understanding, and to use the key terms *understanding, partial understanding, non-understanding,* and *misunderstanding*. For this analysis, only explainers' comments were used. Each explainer participated three (or two) times in the video-recall task, depending on the number of explainees to whom they explained the game.

Participants

The subsample used for the current analysis consists of eight explainers, who were German native speaking adults (*M* = 23.6, *SD* = 3.38). Among them, two were males, and six were females. Only 18 of the 24 explainees provided socio-

demographic information about age (M = 26.0, SD = 9.75), gender (7 male and 11 female) and native language (also German). All participants signed a consent form. The study had been approved by the Ethics Board of the university.

Data coding

All data analyzed in the present study were annotated using *ELAN* software (Max Plank Institute for Psycholinguistics, The Language Archive). Three coders annotated the data. Coder A annotated explainers' hand gestures in the dyadic interactions and explainers' comments on explainees' understanding from the video-recall task. Coders B and C annotated 10% of the data, respectively, to assess reliability.

Hand gestures For annotating explainers' hand gestures, coder A segmented and annotated gesture phrases (McNeill, 1992), that is, gestural movements constituted of gesture strokes and optional preparation and retraction phases of the arm and hand. The recordings from the camera perspective directed towards the torso, hand and head of the explainers were used (with the audio turned on) because it allowed observing explainers' hand shapes and movements over the shared referential space. To ensure reliability, 10% of the data were annotated by coder B (κ = 0.94). Coders identified first explainers' pointing behavior based on explainers' hand / finger shape, and then they annotated the relevant gesture functions according to the feature definitions provided by McNeill (1992, 2006). We observed the dimension of gesture deixis not only in the one-dimensional form of deictic gestures, but also in hybrid forms including iconicity or beats, i.e., deictic-iconic or deictic-beat gestures. *Deictic gestures* were coded based on a single pointing towards a direction or a location where an invisible object would be placed, and co-occurring with the related spoken reference. *Deictic-iconic gestures* were coded based on the criteria for categorical deictic gestures complemented by hand or finger shapes or movements depicting an object, features of an object, or a path. The explainers from our study were observed to point at locations while depicting objects by either positioning the index finger and the thumb in an object related form or drawing objects on the shared referential space by the index finger (Streeck, 2008). *Deictic-beat gestures* were coded based on the criteria for categorical deictic gestures, complemented by (repetitive) biphasic rhythmic hand / finger movements in the presence of prosodic highlighting.

Levels of understanding Coder A annotated the explainers' comments during the video-recall task into the four levels of understanding (Vendler, 1994) that the participants were given as key terms: *understanding, partial understanding, non-understanding*, and *misunderstanding*. Many of the comments could be directly coded based on the presence of these key terms. However, there were other types of comments which did not contain the provided key terms for understanding from the instructions, but rather synonymous or colloquial expressions, for example "to make click" (coll. German for understanding) or "to be unable to visualize" (for non-understanding). Those expressions were coded as one of the levels of understanding. Also, there were comments which were not directly related to explainees' understanding, but rather to the quality of explanation, and such unrelated comments were not considered in the analysis. Coders were trained to sort and decode the relevant information related to explainees' level of understanding. Coder C annotated 10% of the data for a reliability check (κ = 0.85).

Data analysis

For the analysis, all forms of deictic gestures (deictic, deictic-iconic, and deictic-beat) were collapsed into a single variable (gesture deixis). The number of all forms of deictic gestures produced in the gaps between the annotated levels of understanding were counted. The gaps represented the time between two documented levels of explainees' understanding by the explainers. The number of explainers' reports on explainees' levels of understanding varied between the individual dyadic interactions (Table 1).

Table 1. Number of reported levels of understanding across the analyzed subsample of 24 dyadic interactions.

reported levels of:	sum	range	*M*	*SD*
understanding	89	1-22	4.94	5.30
partial understanding	58	1-9	3.41	2.53
non-understanding	61	1-10	2.54	2.10
misunderstanding	18	1-8	2.00	2.34

The data frame was structured according to the nested design of data collection, i.e., the random effect was structured hierarchically in two columns (explainer and explainee). Before choosing the appropriate statistical model, we ran Shapiro-Wilk normality test, which indicated a non-normal distribution of explainers' gestures across the 24 interactions (W = 0.90, p < 0.05). Because of non-normal distribution, we ran a Generalized Linear Mixed Effects Model (GLMM) in *Rstudio* (Rstudio Team, 2020), using the *lme4* package (Bates et al., 2015) with the function:

glmer <- GEST_FREQ ~ UNDERSTAND + (1 | EX/EE)

The frequencies of explainers' different forms of deictic gestures were used as the response variable. The monitored levels of understanding (four-level) were the fixed effect applying a simple contrast, comparing the levels of partial understanding, non-understanding and misunderstanding to the reference level of understanding. The random effect was defined by the nested study design representing each explainer interacting with a different explainee.

Results

Our statistical model indicated a balanced good fit (AIC = 1271.2; BIC = 1283.9) compared to a null model without the fixed effect (AIC = 1384.8; BIC = 1391.2), a low proportional variance based on the fixed effect (marginal R^2 = 0.165), but

a higher proportional variance in combination with the random effect (conditional R^2 = 0.943). The nested random effect indicated a greater variance of individual explainers' gesture deixis across interacting with different explainees (σ^2 = 0.21, *SD* = 0.45) compared to the variance across the eight different explainers regardless the attendance of three different explainees (σ^2 = 0.08, *SD* = 0.29). The fixed effects summary (Table 2 and Figure 1) suggests that the levels *understanding*, *partial understanding* and *misunderstanding* have a significant effect on the variations of the frequencies of gesture deixis across the explainers.

Table 2. Explainers' frequency of gesture deixis related to interpretations of explainees' understanding.

effect	*M*	*SD*	*β*	*SE*	*z*	*p*
U (int.)	46.19	32.59	3.95	0.14	27.59	***
PU	38.0	23.69	-0.33	0.06	-5.83	***
NU	61.36	44.94	0.05	0.05	0.97	ns
MU	27.28	26.48	-0.67	0.09	-7.63	***

*** ($p < 0.001$), ns ($p > 0.05$)
U = understanding (intercept), PU = partial understanding, NU = non-understanding, MU = misunderstanding

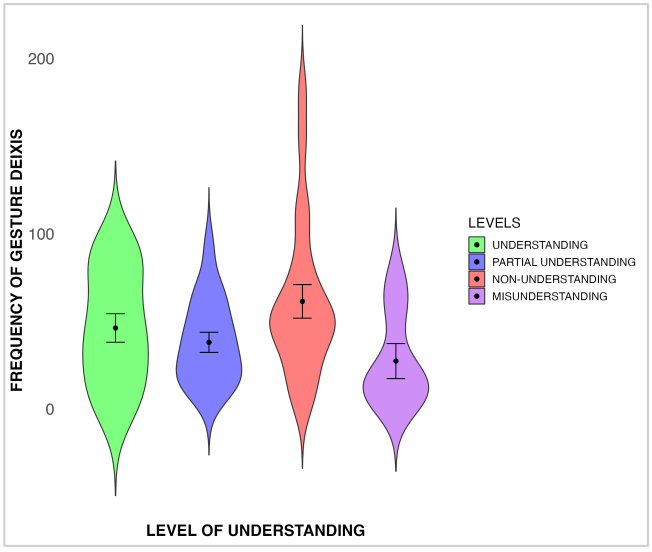

Figure 1: Explainers' gesture deixis related to interpretations of explainees' understanding.

For testing our first hypothesis whether explainers' gesture deixis decreases after monitoring complete understanding or increases after monitored partial, non- and misunderstanding, we looked at the estimated means and conducted post-hoc pairwise comparisons for significant differences. The results are summarized in Table 3. Overall, the results do not suggest that explainers' gesture deixis decreases following the interpretation of complete understanding in explainees' behavior. Gesture deixis after monitoring non-understanding increases slightly compared to gesture deixis after monitoring complete understanding. However, the difference between both extremes is not significant (β = -0.05, *SE* = 0.05, *z* = 0.97, $p > 0.05$). We observed that gesture deixis decreases in relation to explainers' reports of explainees' partial understanding and misunderstanding. The statistical model indicated significant differences for the comparison between understanding and partial understanding (β = 0.33, *SE* = 0.06, *z* = 5.83, $p < 0.001$), as well as for the comparison between understanding and misunderstanding (β = 0.67, *SE* = 0.09, *z* = 7.63, $p < 0.001$).

Table 3. Explainers' frequency of gesture deixis related to interpretations of explainees' understanding: Estimated means and *SE*.

Understanding	*EM*	*SE*	*LCL*	*UCL*
U	3.95	0.14	3.67	4.23
PU	3.62	0.14	3.33	3.90
NU	4.00	0.14	3.73	4.28
MU	3.28	0.16	2.97	3.59

Although the results indicated that monitoring explainees' understanding, partial understanding and misunderstanding is related to variations of explainers' gesture deixis, hypothesis 1 could not be verified. The frequency of explainers' gesture deixis following interpretations of explainees' complete understanding is not significantly different than the frequency of gesture deixis following interpretations of explainees' non-understanding, and it decreases significantly following interpretations of explainees' partial and misunderstanding.

For hypothesis 2, we explored intra-individual differences in explainers' gesture deixis to reveal the random effect variations from our statistical model in a descriptive manner. The first part of our analysis indicated higher intra-individual variations of explainers' gesture deixis regarding the three different explainees compared to inter-individual variations between the eight explainers. The individual charts in Figure 2 illustrate normalized proportions derived from the absolute frequencies of each explainer's gesture deixis related to the reported levels of understanding of each explainee. The variance of monitored levels of understanding for each of the interactions between an explainer (EX) and an explainee (EE) is immediately visible: Explainers have not reported on monitoring all four levels of understanding in each interaction with a different explainee. Thus, we can compare the use of gesture deixis only for non-understanding, partial understanding and understanding. Regarding the level of non-understanding, we observed intra-individual differences in the proportions of gesture deixis for EX12, EX13 and EX16. All explainers who monitored explainees' partial understanding used gesture deixis differently when interacting with a different explainee. Comparable differences between the proportions of explainers' gesture deixis related to monitoring explainees' understanding were observed in EX7, EX9, EX11, EX13 and EX19. Our results indicate that the use of gesture deixis is related not only to the

attendance of a different explainee but also to the monitored level of explainees' understanding by the explainers. The results on the variance at the level of different explainees and the influence of the monitored levels of explainees' understanding on the frequencies of gesture deixis, support hypothesis 2 that explainers exhibit intra-individual variations in gesture deixis regarding the monitored levels of understanding.

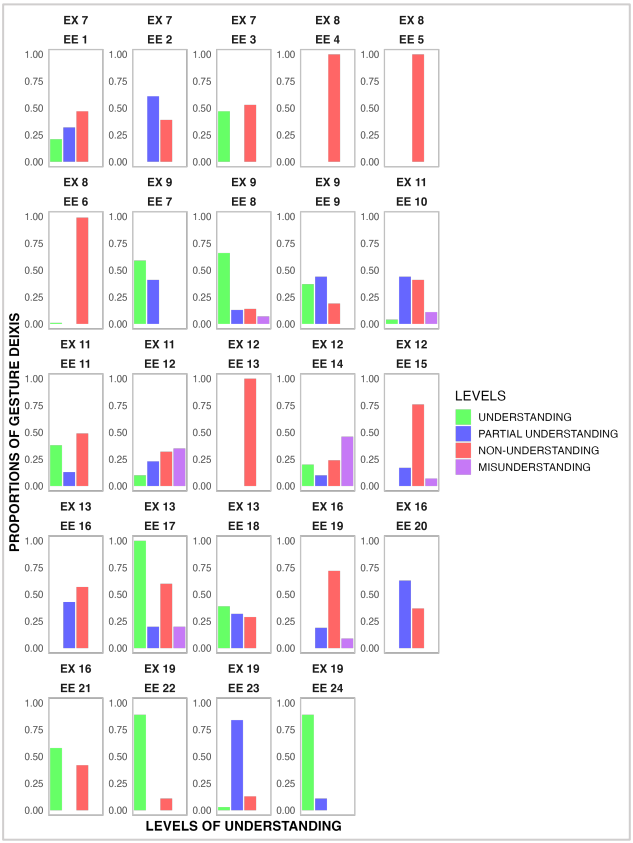

Figure 2: Individual proportional variations of explainers' gesture deixis related to interpretations of explainees' understanding.

Discussion

In this study, explainers' gesture deixis in relation to their interpretations of explainees' levels of understanding when explaining a board game was analyzed. Further, explainers' intra-individual variations of gesture deixis when interacting with different explainees were addressed. Other than hypothesized, the results indicated that monitoring explainees' complete understanding is not followed by a decrease in explainers' gesture deixis. Also, the exploration of individual explainer's gestures revealed that explainers adapted their deictic gestures within each interaction with a different explainee and their interpretation of the level of understanding.

Based on previous research on addressees' increasing comprehension when observing co-speech gestures (Clark, 2003; Congdon et al., 2017; Dargue et al., 2021; Kandana-Arachchige et al., 2021; McKern et al., 2021; Stojnic et al., 2013), we assumed that explainers' interpretations of explainees' complete understanding would be associated with a decrease in their pointing behavior in the interaction. In our analysis, we did not find support for this assumption. Explainers' use of gesture deixis during the explanations in the absence of the board game remained stable, even when interpreting explainees' complete understanding. One possible reason for explainers' continuous use of gesture deixis could be that the absence of the board game required the establishment of joint imagined spaces (Kinalzik & Heller, 2020), also by pointing to invisible locations and referents. This might have been especially pronounced because the explainers familiarized themselves with the game instructions before the study, and thus they had become experts of the board game, in comparison to the explainees who were novices. Because of this knowledge gap during the interaction and the physical explanandum being absent, explainers may have expected a continuous high demand for a visual presentation of the board game components and their spatial organization on the imagined space by the less knowledgeable explainees (Kang et al., 2015).

Our results on explainers' individual use of gesture deixis when interacting with different explainees could be related to previous findings on speakers' individual behavior (Priesters & Mittelberg, 2013) and possible variations depending on the attendance of different addressees (Bergmann & Kopp, 2009; Jacobs & Garnham, 2007). Regarding the findings from the current study, we conclude that gesture deixis is related not only to different explainees, but also to explainers' monitoring of explainees' understanding.

In this paper, we focused only on explainers' retrospective reports on explainees' understanding without considering other forms of dynamics, such as explainees' verbal and nonverbal behavior in the interactions or the topical organization of the explanations. This is a limitation of the study. Therefore, in future analyses we aim to expand our research to explore explainers' gesture deixis within certain topics from the explanations, such as specific game rules. Thus, the consideration of the topical organization (i.e., openings and closures of topics, as well as elaborations of topics) would also allow the analysis of explainers' gesture deixis and their relation to dynamics of explainees' understanding within specific explanation episodes.

Further research will consider more fine-grained statistical analyses including hybrid gestural forms (i.e., deictic-iconic, deictic-beat) and different forms of explainees verbal and nonverbal forms of feedback behavior (e.g., gaze behavior, head gestures, and linguistic backchannels).

Acknowledgments

This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) TRR 318/1 2021 - 438445824. We thank all participants for supporting this research and our research assistants for their help in transcribing and annotating the video data.

References

Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics, 9*(1), 1–26. https://doi.org/10.1093/jos/9.1.1

Allwood, J. & Cerrato, L. (2003). A Study of Gestural Feedback Expressions. In P. Paggio, K. Jokinen & A. Jönsson. (Eds.), *First Nordic Symposium on Multimodal Communication* (pp. 7-22).

Arnold, K. (2012). Humming along. *Contemporary Psychoanalysis, 48*(1), 100–117. https://doi.org/10.1080/00107530.2012.10746491

Austin, E. E., & Sweller, N. (2014). Presentation and production: the role of gesture in spatial communication. *Journal of Experimental Child Psychology, 122*, 92-103. https://doi.org/10.1016/j.jecp.2013.12.008

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co narrators. *Journal of Personality and Social Psychology, 79*(6), 941–952. https://doi.org/10.1037/0022-3514.79.6.941

Bazzanella, C., & Damiano, R. (1999). The interactional handling of misunderstanding in everyday conversations. *Journal of Pragmatics, 31*(6), 817-836. https://doi.org/10.1016/S0378-2166(98)00058-7

Beege, M., Ninaus, M., Schneider, S., Nebel S., Schlemmel, J., Weidenmüller, J., Moeller, K. & Rey G. D. (2020). Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Computers & Education, 156*, Article 103955. https://doi.org/10.1016/j.compedu.2020.103955

Bergmann, K., & Kopp, S. (2009). Systematicity and idiosyncrasy in iconic gesture use: empirical analysis and computational modeling. In S. Kopp & I. Wachsmuth (Eds.), *Lecture Notes in Computer Science* (pp. 182-194). https://doi.org/10.1007/978-3-642-12553-9_16

Bühler, K. (1965). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Gustav Fischer Verlag.

Bühler, K. (1982). The deictic field of language and deictic words. In R. J. Jarvella & W. Klein (Eds.), *Speech, place and action: Studies in Deixis and related topic* (pp. 9-30). John Wiley & Sons, Ltd.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243–268). Lawrence Erlbaum. https://doi.org/10.4324/9781410607744

Clark, H. H. & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62-81. https://doi.org/10.1016/j.jml.2003.08.004

Congdon, E., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction, 50*, 65–74. https://doi.org/10.1016/j.learninstruc.2017.03.005

Dargue, N., Phillips, M. & Sweller, N. (2021). Filling the gaps: observing gestures conveying additional information can compensate for missing verbal content. *Instructional Science, 49*, 637-659. https://doi.org/10.1007/s11251-021-09549-2

de Ruiter, J. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture (Language Culture and Cognition)* (pp. 284-311). Cambridge University Press. https://doi.org/10.1017/CBO9780511620850.018

Dimitrova, D., Chu, M., Wang, L., Özyürek, A., Hagoort, P. (2016). Beat That Word: How Listeners Integrate Beat Gesture and Focus in Multimodal Speech Discourse. *Journal of Cognitive Neuroscience, 28*(9), 1255-1269. https://doi.org/10.1162/jocn_a_00963

Doherty-Sneddon, G., & Phelps, F. G. (2007). Teacher's responses to children's eye gaze. *Educational Psychology, 27*(1), 93-109. https://doi.org/10.1080/01443410601061488

ELAN (Version 6.2.) [Computer Software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. https://www.archive.mpi.nl/tla/elan

Gander, A. G., & Gander, P. (2020). Micro-feedback as cues to understanding in communication. Dialogue and Perception – Extended Papers from DaP2018. In C. Howes, S. Dobnik, & E. Breitholtz (Eds.), *CLASP Papers in Computational Linguistics* (pp. 1-11). Gothenburg University.

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & cognition, 26*(4), 651–658. https://doi.org/10.3758/bf03211385

Grimminger, A., Rohlfing, K. J., & Stenneken, P. (2010). Children's lexical skills and task demands affect gestural behavior in mothers of late-talking children and children with typical language development. *Gesture*, *10*(2–3), 251–278. https://doi.org/10.1075/gest.10.2-3.07gri

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The Role of Synchrony and Ambiguity in Speech–Gesture Integration during Comprehension. *Journal of Cognitive Neuroscience, 23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Holler, J. & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology, 26,* 4–27. https://doi.org/10.1177/0261927X06296428

Jacobs, N. & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56,* 291–303. https://doi.org/10.1016/j.jml.2006.07.011

Jongerius, C., Hillen, M. A., Romijn. J., A., Smets, E. M. A., & Koole, T. (2022). Physician gaze shifts in patient physician interactions: functions, accounts and responses. *Patient Education and Counceling, 105*(7), 1-14. https://doi.org/10.1016/j.pec.2022.02.018

Kandana-Arachchige, K. G., Blekic, W., Simoes Loureiro, I., & Lefebvre, L. (2021). Covert attention to gestures is sufficient for information uptake. *Frontiers in Psychology, 12*, 776867. https://doi.org/10.3389/fpsyg.2021.776867

Kang, S., Tversky, B., & Black, J. B. (2015). Coordinating gesture, word, and diagram: Explanations for experts and novices. *Spatial Cognition & Computation*, *15*(1), 1–26. https://doi.org/10.1080/13875868.2014.958837

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253–260. https://doi.org/10.1016/S0093-934X(03)00335-3

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychological Science, 21*(2), 260-267. https://doi.org/10.1177/0956797609357327

Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge University Press. https://doi.org/10.1017/cbo9780511807572

Kinalzik, N., & Heller, V. (2020). Establishing joint imagined spaces in game explanations: Differences in the use of embodied resources among primary school children. *Research on Children and Social Interaction, 4*(1), 28-50. https://doi.org/10.1558/rcsi.12417

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes, 24*(2), 145–167. https://doi.org/10.1080/01690960802586188

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology, 113*(3), 387-404. https://doi.org/10.2307/1423365

Li, W., Wang, F., Mayer, R. E., & Liu, T. (2021). Animated pedagogical agents enhance learning outcomes and brain activity during learning. *Journal of Computer Assisted Learning, 38*(3), 621–637. https://doi.org/10.1111/jcal.12634

McKern, N., Dargue, N., Sweller, N., Sekine, K. & Austin, E. (2021). Lending a hand to storytelling: Gesture's effects on narrative comprehension moderated by task difficulty and cognitive ability. *Quarterly Journal of Experimental Psychology, 74*(10), 1781-1895. https://doi.org/10.1177/17470218211024913

McNeill, D. (1992). *Hand in Mind: What Gestures Reveal about Thought.* The University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought.* University of Chicago Press. https://10.7208/chicago/9780226514642.001.0001

McNeill, D. (2006). Gesture and Communication. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics (Second Edition)* (pp. 60-66). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00798-7

Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2013). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General, 143*(1), 195–204. https://doi.org/10.1037/a0032246

Poggi, I. (2008). Iconicity in different types of gestures. *Gesture, 8*(1), 45–61. https://doi.org/10.1075/gest.8.1.05pog

Priesters, M. A., & Mittelberg, I. (2013). Individual differences in speakers' gesture spaces: Multi angle views from a motion capture study. *TiGeR Workshop*, Tilburg, NL. https://tiger.uvt.nl/pdf/papers/priesters.pdf

Rohlfing, K. J., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H. M., Buschmeier, H., Esposito, E., Grimminger, A., Hammer, B., Häb-Umbach, R., Horwath, I., Hüllermeier, E., Kern, F., Kopp, S., Thommes, K., Ngonga Ngomo, A.-C., Schulte, C., Wachsmuth, H., Wagner, P., & Wrede, B. (2021). Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems, 13*(3), 717-728. https://doi.org/10.1109/TCDS.2020.3044366

Rohrer, P. L., Delais-Roussarie, E. & Prieto, P. (2020). Beat Gestures for Comprehension and Recall: Differential Effects of Language Learners and Native Listeners. *Frontiers in Psychology, 11*, Article: 575929. https://doi.org/10.3389/fpsyg.2020.575929

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Stojnic, U., Stone, M., & Lepore, E. (2013). Deixis (even without pointing). *Philosophical Perspectives, 27*(1), 502–525. https://doi.org/10.1111/phpe.12033

Streeck, J. (2008). Depicting by gesture. *Gesture, 8*(3), 285-301. https://doi.org/10.1075/gest.8.3.02str

Türk, O., Wagner, P., Buschmeier, H., Grimminger, A., Wang, Y., & Lazarov, S. (2023). MUNDEX: A multimodal corpus for the study of the understanding of explanations. In P. Paggio & P. Prieto (Eds.), *Book of Abstracts of the 1st International Multimodal Communication Symposium* (pp. 63–64).

Vendler, Z. (1994). Understanding Misunderstanding. In D. Jamieson (Ed.), *Language, Mind, and Art* (pp. 9-22). Springer. https://doi.org/10.1007/978-94-015-8313-8_2

Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics, 32*(8), 1177–1207. https://doi.org/10.1016/s0378-2166(99)00109-5

West, D. E. (2014). *Deictic Imaginings: Semiosis at Work and at Play*. Springer. https://doi.org/10.1007/978-3-642-39443-0

Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting Chicago Linguistic Society*, April 16-18, 1970, Chicago Linguistic Society, Chicago (pp. 567-578).