Coupling of Task and Partner Model: Investigating the Intra-Individual Variability in Gaze during Human–Robot Explanatory Dialogue

Amit Singh
SFB/Transregio 318 Constructing Explainability
Paderborn University, Germany
amit.singh@uni-paderborn.de

Katharina J. Rohlfing
SFB/Transregio 318 Constructing Explainability
Paderborn University, Germany
katharina.rohlfing@uni-paderborn.de

Abstract

In a successful dialogue in general and a successful explanation in specific, partners need to account for both, the task model (what is relevant for the task) and the partner model (what one can contribute). The phenomenon of coupling between task and the partner model becomes especially interesting in the context of Human–Robot Interaction where humans have to deal with unknown capabilities of the robot, which can momentarily be perceived when the robot is unable to contribute to the task. Following research on the path over manner prominence in an action [31–33], a robot explained actions to a human by emphasizing two aspects – the path ("where" component) and the manner ("how" component). On critical trials, the robot occasionally omitted one of these components where participants sought missing information for the path or the manner. Participants' information-seeking and gaze behaviour were analysed. Analysis confirms the initial predictions for, a) task model (path over manner prominence), i.e., earlier information-seeking for path-missing than manner-missing trials, and b) partner model, i.e., while information-seeking is predominantly tied to the attention on the robot's face, when robot fails to provide resolution, attention shifts more often towards its torso – a behavior likely to indicate an exploration of the robot's capabilities. An individual-level analysis further confirms that the intra-individual variation in the task model is partly influenced by the perceived capability of the robot.

CCS Concepts

• **Human-centered computing → Empirical studies in collaborative and social computing**.

Keywords

Explanation, Scaffolding, Eyetracking, Partner Model, HRI

ACM Reference Format:
Amit Singh and Katharina J. Rohlfing. 2024. Coupling of Task and Partner Model: Investigating the Intra-Individual Variability in Gaze during Human–Robot Explanatory Dialogue. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI Companion '24), November 04–08, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3686215.3689202

ICMI Companion '24, November 04–08, 2024, San Jose, Costa Rica
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0463-5/24/11
https://doi.org/10.1145/3686215.3689202

1 Introduction

Explanation is a process as well as an outcome, which does not conform to a one-size-fits-all approach [30]. In this process, the involved partners dispose of distinct expectations about their partner (explainer) and the task [9, 18, 25]. Whereas the expectation about the partner relates to the questions of who is making what kind of contribution (partner model), the expectation about the task concerns which aspect of the task is relevant to achieve the task goal (task model) [37]. Although the coupling of the two models was postulated for effective tutorial interactions [37], little is known about how it can be achieved in other forms of dialogue, e.g., in Human–Robot Interaction. The coupling of such models becomes particularly significant in a human–robot explanatory dialogue, since in this context, humans as explainees not only question their partner's capabilities more often but also attempt to continuously determine whether and to what extent the robot can contribute to the task. For this purpose, humans usually test the robot with multiple probes on which basis they finally establish a partner model, i.e., what kind of capabilities they can rely on in an interaction[27]. Following this, the task model can be co-constructed during the course of interaction, tailored for both the parties based on their current state on understanding [25]. For example, studies in the past suggested that humans adapt to the robot's cues that are interpretative of what the robot is capable of [8]. In this sense, the dialogue "provides subtle clues to the robot's functionality and thus to adequate partner modeling" [8, p. 35]. Other studies have shown that humans are able to align with the robot's actions depending on their partner model of the robot [36]. However, previous studies did not model the individual differences to study how the coupling of these models can be achieved. For an individual, the task conceptualization could rely on the partner model, such that participants can structure or adapt their linguistic behavior about the task based on the perceived capability of the robot [35]. In this context, the perceived capability at a specific moment in the interaction can influence how the utterances about the source of misunderstanding are structured [25]. Whereas most studies consider one of the aspects, the challenge lies in systematically define and relate the task and the partner models. Hence, here our aim was to first lay down the definitions of these models in scope of our study and subsequently investigate the relationship between them.

1.0.1 Task model. We define a task model requiring information on two dimensions to perform the given action – Path and Manner. Here, the path characterizes the "where" component in an action, whereas the manner specifies "how" the action is to be performed. This task model is grounded on an established finding in psycholinguistics research [29, 31], where each action was performed with attention to the path and manner. The explainer (robot) provided the information on how and where the given object has to be placed through a verbal guidance (section 2.4). In numerous studies, it has been found that in an action event while path remains in the foreground of attention the manner component tends to be attenuated, resulting in a more robust memory for the path compared to the manner [31–33]. This phenomenon is consistently observed across studies in adults as well as developmental studies in infants [14, 22]. Thus, we expected to observe the effect of path vs manner prominence in the information-seeking behaviour of the participants, consequently informing us about the task conceptualization. This approach first allows us to establish our initial predictions for the task model, and then systematically investigate how the partner model influences it during an interaction.

1.0.2 Partner model. We view the partner model as a dynamic process, evolving as the interaction progresses. In our study, the robot predominantly serves as the explainer, given that it possesses more information about the task than the explainee. Consequently, we assume that its capability is continually assessed by the participants (explainee) throughout the interaction, and it may be perceived as less capable when it fails to provide resolution occasionally. Therefore, the explainee develops an expectation or "need" for the robot to effectively scaffold the task. The partner model thus represents the explainee's perception of the explainer's (robot) capability to act as a scaffolding agent. Crucially, this initial perception of the explainer can change throughout the interaction based on the explainer's demonstrated ability to provide –or fail to provide – adequate explanations when requested. Hence, the partner model predominantly draws inspiration from the concept of scaffolding, which is characterized as a supportive and dynamic form of assistance provided by a more knowledgeable partner, tailored to the learner's abilities [37]. In HRI, this emphasizes that the explainer (robot) must be perceived as a partner possessing sufficient characteristics of a scaffolding agent before the explainee can rely on this support. In situations where the conversation breaks down due to a lack of sufficient information or misunderstanding, to keep the loop going, the explainee should be able to offload the task demands to the explainer (robot) by assigning the missing-information in the task. This assignment could manifest in rather various ways, such at looking at the robot's face when in doubt or before raising requests [2, 7, 34], which is considered as a proactive function of gaze during information-seeking in conversation [17, 24].

1.1 Present Study

The setting was created by incorporating the abovementioned two models. In each trial, the robot asked the participant referring to an object to place in a specific manner (vertically or horizontally) on a destined location (path; on letters). To elicit questions from the participant, the robot occasionally omitted either the path or manner information. Our predictions were based on population as well as individual level variations for both the task and partner models, which we specify hereunder.

1.2 Predictions

1.2.1 Task Model: Building upon the findings in psycholinguistics that the path remains salient than manner in an action [21, 29, 31], we predicted that the absence of path information would elicit quicker information-seeking behaviour from participants compared to the instances where the manner information is missing. Crucially, this prediction serves as a validation of our task model. The prediction is also supported by the previously proposed notion that the asymmetrical conceptualization of the path and manner arises from the intrinsic goal-directed nature of cognition, with a primary emphasis on the path [21, 32, 33].

1.2.2 Partner Model: Perceiving a partner as a legitimate scaffolding agent would entail focusing more on the partner's face than other parts of the body especially while information-seeking phase. Crucially, this prediction would be supplemented by comparing the gaze behavior in the condition where the robot successfully provides a resolution (successful scaffolding), compared to a condition where the robot fails to provide a resolution (unsuccessful scaffolding), prompting the explainee to re-ask the question or proceed without support. We further predicted that an unsuccessful scaffolding would lead to an exploration of the partner capabilities where the attention shifts from the robot's face to the other parts of the body. Moreover, evidence from human–human conversation suggests that not looking into the face while information-seeking is often interpreted as a sign of avoidance and lack of interest in conversation [11, 16], such as in the case of unsuccessful scaffolding. Thus, at the population level, we predicted that while the overall looks to face would be higher than other parts of the body during information-seeking phase, this facial engagement would decrease when the robot shows a lack of capability to scaffold the task.

1.2.3 Relationship between the task and partner model. At the individual level, we hypothesized that participants showing more exploratory behaviour while information-seeking — a behavior that is likely to reflect an exploration of partner's capabilities following unsuccessful scaffolding — would exhibit more delay in information-seeking. Here our aim is to find out whether the source of the variability in the task model (delay in information-seeking) is partly determined by the uncertainty about the partner's role and its potential scaffolding capabilities.

2 Method

The methods in this study are approved by the Review Board of the university and the informed consent from all the participants were obtained. The study was preregistered on Open Science Framework (OSF) before data acquisition, and the associated analysis code is available online and the data would be available on request.

2.1 Participants

Thirty-three university students (mean age = 22.30, SD = 2.63) were recruited via classroom advertisements and flyer distribution. All participants demonstrated native to fluent proficiency in German. Data from 3 participants could not be analysed due to track-loss

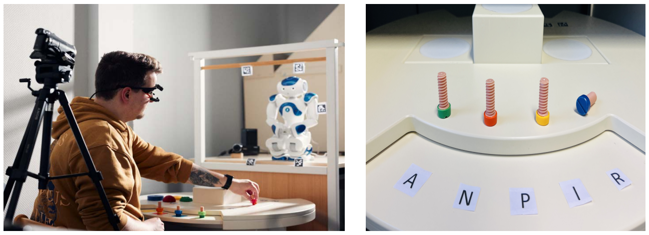

Figure 1: Setup and example stimulus: manners (vertical and horizontal) and paths (A N P I R); Three screws needed to be vertical, while the screw on R needed to be horizontal.

of eye-tracking sample points, hence the final sample size remains thirty participants.

2.2 Stimuli

The stimuli consisted of four objects in the shape of screw of different colours, Green, Red, Yellow, Blue (Fig. 1). Each object can be placed in two possible manners (we will refer to it as a "How"-component) vertically or horizontally (referred to as “hochkant" or “quer" in German). For complexity consideration, we limited the options to these two manner-affordances. Additionally, for each of the four objects, there were five potential positions marked on the table by unique alphabets where the object could be placed. These positions served as the path (we will refer to it as a "Where"-component). The combination of manners (vertical or horizontal) and paths (identified by unique alphabets) was generated randomly for each trial, resulting in a diverse array of configurations, offering distinct patterns concerning both manner and paths and maintaining variability in the task at the same time.

2.3 Procedure

As illustrated in the Figure 1, participants were seated in front of the NAO robot while wearing Pupil Labs glasses for eye-tracking (120 Hz), with a 5-point calibration yielding an accuracy of 0.60° and precision of 0.02°. The room was controlled for illumination and noise isolation. The objects were placed in front of the participant in a transparent glass container on a table. Alphabets stickers were pasted in a row to act as destination path of the object. Two stationary cameras were placed on the sides to capture the experimental setup. To ensure an unbiased interaction between the robot and the participant, an experimenter zone was established behind the participant, concealed by a divider. This setup ensured that the experimenter remained out of sight, facilitating uninterrupted interaction solely between the robot and the participant. The experiment started by calibrating the eyetracking and then introducing the participant to the task.

2.4 Task Instructions

Each participant was instructed to follow the robot for guidance on each placement of the object and were told to ask questions when they feel that the information for the task is insufficient. Since the missing information was only regarding the path or the manner, we expected the participants to ask questions along those dimensions (e.g., where, how, and so on). However, participants were given the liberty to request a repetition whenever they feel it necessary, enhancing the flexibility of the interaction. After introducing the participants to the experiment, the experimenter retreated to the experimenter zone and initiated the experiment script. The robot was programmed with semiautonomous capabilities, where the control can be taken over by the experimenter in case the robot fails or conversation breaks down. On instances when the robot was unable to hear or respond, participants repeated the commands. The robot first introduced itself and then reiterated the task instructions again. After asking for the participant's readiness, the robot commences with the first trial. Each trial began with a specific instruction for placing an object in a designated manner and on a specific path. The instruction template was structured as follows:

> *Place the green Object* ***[Manner]*** *on* ***[Path]***
> *Lege bitte das grüne Objekt* ***[Hochkant or Quer]*** *auf* ***[A]***

Following each instruction, the robot gazed briefly to the task and then to the participant and sought confirmation: “Hast du das erledigt?” (“Have you done it? in English). Upon receiving affirmation, the robot progressed to the next object. The trial concluded when all four objects were placed on the designated path with specific manner. Subsequently, the robot asked participants to put back the objects and then turn back to solve a simple addition or subtraction task. This was done to offload the working memory and to keep participants engaged in the task rather than just passively following it. After solving the task, participants turned back towards the robot, and completed a recall task where they placed back the objects as instructed during the trial. Upon confirmation, the robot proceeded to the next trial. The experiment had four blocks, two for path and two for manner critical trials. Thus, a block included total five trials, and each trial had 4 moves corresponding to 4 objects, resulting in a total of 20 moves in a block generated randomly. In each block, information regarding the path or manner was intentionally omitted (missing information trials) randomly on 4 moves and rest of the moves (16) were of full information (full information trials). This was done to prompt questions from the participants for the missing information occasionally. The entire experimental session lasted approximately 45 minutes, during which eye-gaze and video data were continuously recorded.

3 Data Analysis

Eyetracking data was exported using the Pupil Player software. The videos were initially coded using ELAN [1] and later exported for further analysis. Since we were only interested in the behaviour for the time window from instruction onset till the moment of asking the question, we focus our analysis on this time window. All preprocessing and analysis were conducted using an R script [23], which is openly accessible in an online repository for reproducibility.

3.1 Preprocessing gaze data

Since we were interested in the looks to the face, torso, and the objects, three rectangular area of interests, referred to as surfaces, were defined around the robot's face, torso and the objects respectively. Fixations were computed using a time range from 80 to 350 ms with

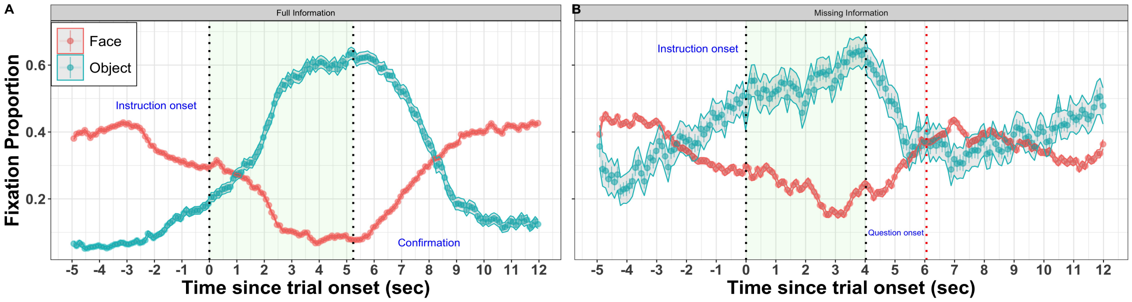

Figure 2: Gaze on robot's face and objects; (A) full information trial and, (B) missing path or manner information trial (right). The green-shaded region shows the instruction duration. Verticle red-line in B depicts the average question onset time

a maximum dispersion set at the default value of 1.5 degrees. To calculate the fixation proportion on a surface within a specified time range, the entire time window was divided at a 50 ms resolution. The number of fixations falling within a surface was then divided by the total number of fixations within that time bin. The delay in the information-seeking was calculated for the time window after the instruction offset by the robot for each trial. Our aim was to model the time-dependent changes in the gaze for the full and missing information trials. The full-information trial spanned from start of the instruction till the confirmation was provided by the participant. And the missing information trial spanned from start of the instruction till the offset of the participants' question. We modeled these behaviours on a continuous time scale using 2nd order Growth Curve Analysis (GCA), a methodology previously applied in analyzing time-dependent changes [19].

3.2 Preprocessing ELAN data

The video data was manual coded on ELAN by three student assistants independently, who were unaware of the hypothesis of the study. The codings included the annotations of robot instructions, participant-posed questions and successful vs unsuccessful scaffolding. A successful scaffolding was considered when the participant sought information and the robot responded successfully (No-error trials), otherwise an unsuccessful scaffolding when the robot did not respond to the request (error trials).

4 Results

4.1 Attention on the objects and the face

For the full-information condition (Fig. 2. A), we observed an increase of attention to the objects as soon as the instruction was presented (vertical line at 0). This continued throughout the instruction period (green shaded region), where a transient decrease in fixation on the face was observed, showing the participants' anticipatory fixation to the objects upon encountering the instruction. Crucially, the attention to objects remained till the end of instruction, after which there was a shift towards the face, as a result of providing confirmation for the completed task (around 10 sec). For the missing information trial (Fig. 2. B), there was an increase in attention towards the objects as soon as the instruction was encountered (vertical line at 0). However, at the end of the instruction, and after realizing that the information was insufficient, participants quickly redirected their gaze back to the robot's face, likely to seek information for the missing aspect in the task. This information-seeking is shown by a transient increase in gaze on robot's face starting just after the offset of instruction which suggests that information-seeking is tightly coupled with looking at the partner, and in this case specifically on the face. This increase in looks to the face was followed by an information-seeking (red vertical line). Importantly, the delay in the onset of this information-seeking was depended upon the type of information missing in the task, i.e., the path or the manner (task-model). According to our first prediction we expected this delay to be higher for the manner than the path, which we analysed subsequently.

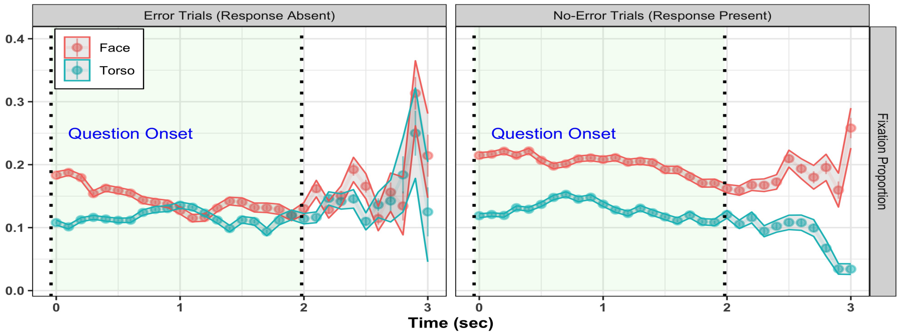

Figure 3: Fixation on face and torso for critical trials where robot did or did not respond to the question. Left plot; unsuccessful scaffolding, Right plot; successful scaffolding. The shaded region shows the mean duration of question onset and offset.

4.2 Information-seeking delay for path and manner

Following our prediction about the task model, we compared the time of question onset for the path and manner critical trials. Here, a delayed onset was observed for manner-related questions compared to path-related questions (Fig. 4), supporting our initial prediction regarding path prominence in event conceptualization. This prediction was confirmed by fitting a hierarchical linear mixed effects model using lme4 package in R [5]. We incorporated a full model with varying intercepts and slopes for each participant to account for main and random effect structure. The predictors were sum contrast coded, with manner coded as +0.5 and path as -0.5, such that the intercept represented the overall mean delay, and the slopes indicated deviations around this mean for each condition. Model predictions are depicted in Fig. 4. Confirming our predictions for the task model, we found a significant delay for the manner than the path related questions (β = 0.44, S.E. = 0.2, p=0.03).

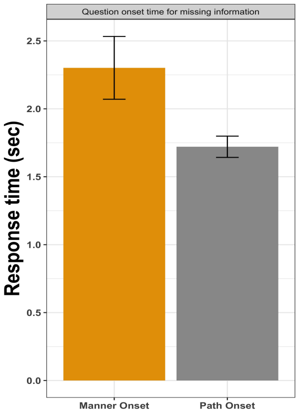

Figure 4: Difference in path vs manner question time. Conditions are on x-axis and response time on y-axis

4.3 Task and partner model

Our population-level partner model suggested that participants redirect their attention towards the partner's face from the object when they intend to ask question.(Fig. 2). Following our partner model predictions, whether the participants' perception about the robot's scaffolding capability (successful vs unsuccessful scaffolding) guided the gaze behaviour towards the robot, we analyzed the gaze in two conditions – where robot successfully provided the resolution (Response present) or did not provide the resolution (Response absent) (Fig. 3). Crucially, we looked into the fixation on the robot's face and the torso during the information-seeking phase. For the unsuccessful scaffolding, we observed an exploratory gaze behaviour where a higher gaze switches from robot's face to the torso and vice-versa was observed (Fig. 3, Response absent or error trial). Importantly, when robot successfully scaffolded the task (Response present), the attention was predominantly was on the face throughout the entire information-seeking phase (Fig. 3, Response present or No-error trial). Further to investigate whether the scaffolding capability contributes to the task conceptualization and thus influence the task model (information-seeking), we modeled the attention allocation on face vs torso as a function of question onset delay. The aim was to look whether the response delay predicted the individual gaze preferences on the face over the torso. For this, we focused our analysis in the information-seeking time window, since this specific window allowed us to examine the proactive role of gaze, when one seeks information particularly about the task. A linear mixed effect model was fitted for logit transformed proportion of looks at the face and torso following [3]. The model included sum-contrast coded Area of Interest (AOI: Face vs Torso) and the question onset delay as a continuous predictor by mean scaling. Our goal was to fit a maximal model by accounting for random effects at the subject level variation with slope correlations[4][1]. Hence, the intercept represented the grand mean of gaze towards the face and torso for mean question onset delay, and deviation around this mean as the effect size of looks toward the face as compared to the torso for the population and the individual subjects. Importantly, we were interested in the random effects (correlation parameter) between the gaze and delay effect sizes for each subject. A full and reduced model comparison was done using log-likelihood test by removing only the correlation parameter from the model [4]. The model predictions for the subjects are depicted in Figure 5. Adding

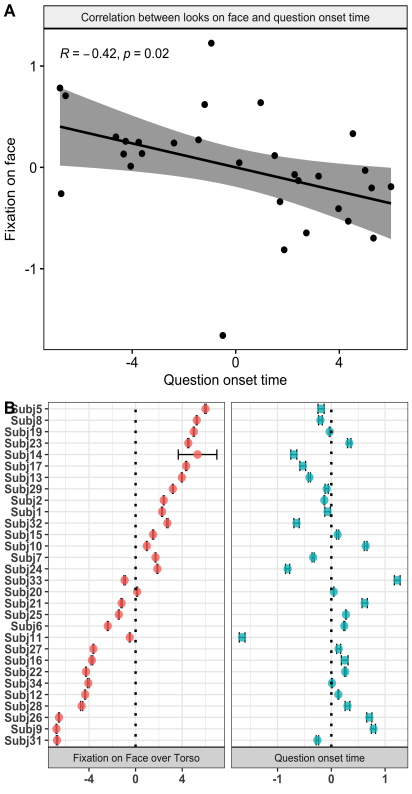

Figure 5: (A) Correlation, looks on face over torso and question time, (B) model prediction for participants

[1] model: lmer(meanLook <- 1+AOI*CentMeanDelay + (1+AOI+CentMeanDelay|Subj))

the correlation random structure between gaze and delay slope parameters significantly improved the model fit (χ^2 = 997.86, Df = 1, p<0.001), suggesting that the correlational parameter effectively captured the variation between onset delay and the gaze on face over torso. Further, there was a significant interaction between AOI and question onset time (β = -0.14, SE = 0.0, p<0.001), such that the delay also negatively predicted the gaze on the face over the torso. Meaning, subjects showing more preference for the robot's face than the torso were quicker in asking questions (R=-0.42, p=0.02), as depicted in Figure 5. In comparison to our population-level partner model, where the majority of participants attended to the face while asking question, on the contrary individuals who exhibited a higher tendency to look at the torso over the face specifically during information-seeking phase, experienced more delay in information-seeking for critical trials. To highlight again, more preference to the torso was observed for the condition where robot was perceived less capable of scaffolding the task (Fig. 3)

5 Discussion

The study builds upon a task model, that requires a differentiation between path and manner of an action. Further, the partner model consists of the expectation that the partner can scaffold the participant to perform those actions. Apart from establishing the validity of these models at the group level, our focus was predominantly to account for the intra-individual variability of the task and the partner model in HRI. Based on our initial assumptions about the task model, we observed that participants were generally quicker in information-seeking for the path compared to the manner. This aligns with a robust finding in psycholinguistics [6, 20–22], which we tested in the context of HRI.

Our investigation into the partner model, with a focus on attentional dynamics, confirmed our initial assumption. At the population level, participants tended to look more at the robot's face during two distinct time segments: a) while receiving the instructions and b) while information-seeking (asking questions). Crucially, while receiving the instructions, we also observed a transient decline in looks at the face which potentially shows the anticipatory attention to the objects as the task instruction is unveiled. Conversely, when participants asked questions (for missing information), their gaze was predominantly directed towards the robot's face–a behavior indicative of perceiving the robot as a scaffolding partner. It is worth to mention that despite the fact that area of interest for the body covered the most part of the visual field of view, the participants' gaze on the face was predominantly higher which might reflect that the robot was ascribed as a scaffolding partner over and above just a point of verbal reference. In this case, one could have sought information by keeping attention on the objects, which might be conceived as a question to oneself than to the partner, given the proactive role of gaze in day-to-day social encounters [11, 15].

Moreover, we observed a delay in manner than path related questions. This underlines our inherent task model, supporting the idea that manner carries more vulnerability necessitating conscious cognitive opearations, as opposed to the relatively automatic processing of path relying on less cognitive resources [21]. As a definition, a conscious cognitive operation on a representation would require more scaffolding than a representation which is already in attention (e.g., path). One of the methods suggested in previous studies to mitigate this effect is contrastive scaffolding of the path and manner [13, 28], which has been shown to improve attention to both components. Provided that in our experiment the path and manner instructions were presented randomly, and the delay was observed for linguistic encoding, than performing an action per se, the likely explanation of the manner related cost could be due to linguistic processes governing the retrieval of these two aspects.

Nevertheless, scaffolding the manner might need more intervention of the scaffolding partner than the path. In this regard, the intra-individual findings from the study highlight the aforementioned aspects, quicker information-seeking when the preference to the robot's face is higher than the torso. This might suggest that the delay in linguistic processing was partly influenced by whether the robot was conceptualized as a partner possessing sufficient scaffolding capability or not. Especially, when the robot did not successfully scaffold the task, the gaze was more exploratory where the attention shifted more often to the torso and other body parts, indicating an exploration.

Here, we focused on a specific aspect of the partner model – perceived scaffolding capability – and found that participants' delayed response is partly influenced by their perception of the partner's scaffolding capability (as assessed by the gaze). We should critically note that in future, our assessment needs to be supported by results from either an online measure assessing the dynamical change in perception of partner's role or an offline survey. Our findings bear significance not only in the realm of human–robot interaction but also extend to human–human interactional contexts, i.e., perceiving a partner capable of scaffolding might allow the explainee to flexibly ascribe more responsibility for the task to their partner. This is grounded on the understanding that joint actions involve close collaboration and a shared distribution of the task load among participants [10]. For example, the amount of information an explainee shares about the source of a misunderstanding can offer insights into her commitment to a joint goal. Furthermore, when an explainee use varied constructions in their verbal behaviour, it can also establish the partner model as more capable to scaffold the task as opposed to when they use repetitive requests, since day-to-day language use is hardly deterministic and relies on variability [12]. In future, we aim to capture more qualitative aspects of this perceptual dimension and investigate its meaning for the facial engagement. Other functional aspects of these measures should also not be neglected, which might entirely allude to a different cognitive function in an explanation, such as, monitoring, which has been proposed to play a different role in explanation processes [26]. Including a more holistic scale in such cases additionally requires considering the meta aspects from multimodal interaction (e.g., backchanneling and signs of ignoring the partner) and might offer us a deeper insight about explanation processes in HRI.

6 Conclusion

This study suggests a potential pathway to systematize task and partner models from an interdisciplinary perspective by investigating human Gaze behaviors in HRI. Future research should delve into the proactive role of the partner model in shaping task model, particularly when they are embedded in an interactional context.

Acknowledgments

The work is supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824.

References

[1] 2023. *ELAN (Version 6.7).* https://archive.mpi.nl/tla/elan [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive..

[2] Michael Argyle and Roger Ingham. 1972. Gaze, Mutual Gaze, and Proximity. *Semiotica* 6, 1 (1972), 32–49. https://doi.org/10.1515/semi.1972.6.1.32

[3] Dale Barr. 2008. Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language* 59, 4 (2008), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

[4] Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 3 (2013), 28–38. https://doi.org/10.1016/j.jml.2012.11.001

[5] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[6] Philippe Bourdin. 1997. On Goal-Bias Across Languages: Modal, Configurational and Orientational Parameters. In *Typology: Prototypes, Item Orderings and Universals (Proceedings of LP '96)*, Bohumil Palek (Ed.). Charles University Press, Praha, 185–218.

[7] Michael Cook. 1977. Gaze and mutual gaze in social encounters: How long—and when—we look others "in the eye" is one of the main signals in nonverbal communication. *American Scientist* 65, 3 (1977), 328–333.

[8] Kerstin Fischer. 2011. How people talk with robots: Designing dialog to reduce user uncertainty. *AI Magazine* 32, 4 (2011), 31–38.

[9] Alan Garfinkel. 1982. Forms of Explanation: Rethinking the Questions in Social Theory. *British Journal for the Philosophy of Science* 33, 4 (1982), 438–441.

[10] Margaret Gilbert. 2014. *Joint Commitment: How We Make the Social World.* Oxford University Press.

[11] Erving Goffman. 1964. The Neglected Situation. *American Anthropologist* 66, 6 (December 1964), 133–136. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00030

[12] H. P. Grice. 1971. Intention and Uncertainty. *Proceedings of the British Academy* 57 (1971), 263–279.

[13] André Groß, Amit Singh, Ngoc Chi Banh, Birte Richter, Ingrid Scharlau, Katharina J. Rohlfing, and Britta Wrede. 2023. Scaffolding the Human Partner by Contrastive Guidance in an Explanatory Human-Robot Dialogue. *Frontiers in Robotics and AI* 10 (2023), 1236184. https://doi.org/10.3389/frobt.2023.1236184

[14] Susan J. Hespos, Molly M. Saylor, and Sharon R. Grossman. 2009. Infants' Ability to Parse Continuous Actions. *Developmental Psychology* 45, 2 (2009).

[15] Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26 (1967), 22–63. https://doi.org/10.1016/0001-6918(67)90005-4

[16] Kobin H. Kendrick and Judith Holler. 2017. Gaze Direction Signals Response Preference in Conversation. *Research on Language and Social Interaction* 50, 1 (2017), 12–32. https://doi.org/10.1080/08351813.2017.1262120

[17] Christopher L. Kleinke. 1986. Gaze and Eye Contact: A Research Review. *Psychological Bulletin* 100, 1 (1986), 78–100. https://doi.org/10.1037/0033-2909.100.1.78

[18] Tim Miller. 2019. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence* 267 (2019), 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[19] D. Mirman, J.A. Dixon, and J.S. Magnuson. 2008. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language* 59, 4 (2008), 475–494.

[20] Anna Papafragou. 2010. Source-Goal Asymmetries in Motion Representation: Implications for Language Production and Comprehension. *Cognitive Science* 34, 6 (August 2010), 1064–1092. https://doi.org/10.1111/j.1551-6709.2010.01107.x

[21] Sarah Pourcel. 2004. What makes path of motion salient?. In *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society.* Sheridan Books, Ann Arbor, MI, 505–516.

[22] Shannon M Pruden, Tilbe Göksun, Sarah Roseberry, Kathy Hirsh-Pasek, and Roberta M Golinkoff. 2012. Find your manners: How do infants detect the invariant manner of motion in dynamic events? *Child Development* 83, 3 (2012), 977–991. https://doi.org/10.1111/j.1467-8624.2012.01741.x

[23] R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[24] F. Roassano. 2010. Questioning and responding in Italian. *Journal of Pragmatics* 42, 10 (2010), 2756–2771.

[25] Katharina J. Rohlfing, Philipp Cimiano, Ingrid Scharlau, Tobias Matzner, Heike M. Buhl, Hendrik Buschmeier, Elena Esposito, Angela Grimminger, Barbara Hammer, Reinhold Hab-Umbach, Ilona Horwath, Eyke Hullermeier, Friederike Kern, Stefan Kopp, Kirsten Thommes, Axel-Cyrille Ngonga Ngomo, Carsten Schulte, Henning Wachsmuth, Petra Wagner, and Britta Wrede. 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 (Sept. 2021), 717–728. https://doi.org/10.1109/TCDS.2020.3044366

[26] Katharina J. Rohlfing, Philipp Cimiano, Ingrid Scharlau, Tobias Matzner, Heike M. Buhl, Hendrik Buschmeier, Elena Esposito, Angela Grimminger, Barbara Hammer, Reinhold Häb-Umbach, Ilona Horwath, Eyke Hüllermeier, Friederike Kern, Stefan Kopp, Kirsten Thommes, Axel-Cyrille Ngonga Ngomo, Carsten Schulte, Henning Wachsmuth, Petra Wagner, and Britta Wrede. 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 (2021), 717–728. https://doi.org/10.1109/TCDS.2020.3044366

[27] J. Scholtz. 2003. Theory and evaluation of human robot interactions. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the.* 10 pp.–. https://doi.org/10.1109/HICSS.2003.1174284

[28] Amit Singh and Katharina Rohlfing. 2023. Contrastiveness in the Context of Action Demonstration: An Eye-Tracking Study on Its Effects on Action Perception and Action Recall. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 45. https://escholarship.org/uc/item/2w94t4cv

[29] Dan I. Slobin. 1987. Thinking for Speaking. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, Vol. 13. 435–445. https://doi.org/10.3765/bls.v13i0.1826

[30] Karol Sokol and Peter Flach. 2020. One Explanation Does Not Fit All. *Künstliche Intelligenz* 34 (2020), 235–250. https://doi.org/10.1007/s13218-020-00637-y

[31] L. Talmy. 1975. Semantics and syntax of motion. In *Syntax and Semantics*, J. Kimball (Ed.). Vol. 4. New York Academic Press, New York.

[32] Leonard Talmy. 1985. *Lexicalization patterns: Semantic structure in lexical forms.* https://www.degruyter.com/database/COGBIB/entry/cogbib.11708/html?lang=en

[33] Leonard Talmy. 2000. *Toward a Cognitive Semantics.* 44, Vol. II. https://mitpress.mit.edu/9780262700962/toward-a-cognitive-semantics/

[34] Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. 2001. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 301–308. https://doi.org/10.1145/365024.365119

[35] A.-L. Vollmer, K.S. Lohan, K. Fischer, Y. Nagai, K. Pitsch, J. Fritsch, K. Rohlfing, and B. Wrede. 2009. Early Information-Seeking in Human-Robot Interaction: Face vs. Torso Attention Shifts. In *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on Development and Learning.* IEEE, 1–6. https://doi.org/10.1109/DEVLRN.2009.5175525

[36] A. L. Vollmer, K. J. Rohlfing, B. Wrede, and A. Cangelosi. 2015. Alignment to the Actions of a Robot. *International Journal of Social Robotics* 7, 2 (2015), 241–252.

[37] D. Wood, J. S. Bruner, and G. Ross. 1976. The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry* 17, 2 (1976), 89–100.