

A Divergence Model of Scaffolding in Dialogue via Negation

Amit Singh^{1*}, Sinem Görmez^{2,3}, Britta Wrede^{2,3}, Katharina J. Rohlfing¹

Abstract—Dialogue is not only a matter of aligning with a partner’s current state. In knowledge-asymmetric settings, such as explanations or tutoring, interlocutors must also build shared understanding over time by contributing complementary information to an emerging semantic space. We propose a divergence model of scaffolding that formalizes this process at the event level using Figure–Ground representations. The model captures how interlocutors update their semantic ground through each other’s contributions and how divergence between their evolving grounds changes over time. We use this model to investigate whether linguistic negation functions as a scaffolding resource when divergence increases. We analyze 67 dyads engaged in a Quarto game explanation task. Figure–Ground representations are derived both from human annotations and from LLM-based extraction, embedded in a distributional semantic space, and used to estimate divergence change across dialogue turns. A Bayesian hierarchical model shows that increases in divergence predict a higher probability of negation. This pattern holds for both LLM-derived and human-annotated representations, although the effect is stronger for the LLM-derived pipeline. These findings support a processual account of dialogue in which negation helps reorient locally diverging interaction toward shared coherence.

Index Terms—Dialogue process, Scaffolding, Co-Construction, Negation, Interaction Modeling

I. INTRODUCTION

Many models of dialogue, including models used in human-AI interaction, treat communication as a problem of alignment. From this perspective, a system uses the interaction history and the current utterance to infer the user’s state and to generate an appropriate next response [1]. This view has been highly influential, and it explains many aspects of successful communication. However, it becomes limited in knowledge-asymmetric settings, such as explanations, tutoring, or instructional dialogue. In such settings, the less knowledgeable partner does not simply need an aligned response. Rather, the partners must gradually build shared understanding over the course of the interaction [2]–[5]. In the literature, these shortcomings have been attributed primarily to two qualitative factors: scaffolding and distributed processes [5]–[10].

Scaffolding is a process through which a more informed partner provides tailored guidance to a less informed partner, often through repeated support, until the task goal is achieved

¹Psycholinguistics, Faculty of Arts and Humanities, Paderborn University, Germany.

²Medical Assistance Systems, Medical School OWL and ³Center for Cognitive Interaction Technology, CITEC from Bielefeld University, Germany. *Corresponding Author: Amit Singh, amit.singh@uni-paderborn.de

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under TRR 318/3 (2026), project number 438445824.

[2]; crucially, this support gradually fades away as the partner becomes competent [6]. The “distributed cognition” perspective further argues that semantics does not reside in individual minds, but is instead grounded in a shared environment to which both partners contribute [7], [9]. Nevertheless, current dialogue systems still lack these capabilities. Taken together, these two aspects form the backbone of dialogue in situations characterized by knowledge asymmetry and learning. In situations with knowledge asymmetry, humans co-construct with and for each other through mutual scaffolding [2], while simultaneously establishing and jointly contributing to an externally available semantic space without an explicit need to represent one another’s internal states [7], [10], [11]. Only when these two characteristics are considered together does *complementarity* emerge as a core organizing principle of dialogue [8], [12]. Hence, scaffolding is a processual aspect of dialogue, constituted through the complementary acts of the partners upon an emerging semantics, rather than through mere alignment with internal representations [12]. Complementary actions thus become indispensable in dialogues characterized by knowledge asymmetry, where divergences from joint goals are both more frequent and more likely to arise than in many other forms of interaction [2], [13]. In such cases, scaffolding serves as a key means through which divergences emerging during dialogue are guided back toward global coherence [14]. Put differently, partners jointly shape the semantic space through divergence by complementing each other. On the one hand, divergences along the complementarity path are inevitable and inherent to many forms of interaction. On the other hand, interaction also affords ecological means through which such divergences can be meaningfully scaffolded. Negation is one such scaffolding mechanism that is pragmatically rich, universal, and frequently used to reorient the emerging dynamics toward global coherence [15], [16]. Consider the following excerpt from a dyadic game explanation, in which two participants jointly work toward understanding *Quarto* game (Table I). From the foregoing dialogue, it becomes apparent that participants not only align their representations across different linguistic levels, but also complement one another by contributing what is not yet available in the shared semantic space. From this distributed perspective, interaction proceeds not merely through alignment, but through completion by the other, that is, through a form of complementarity. For the evolving semantic space, we will consider how partners contribute new information under focus (which we refer to as Figure) against some already established *Ground*. The Figure always acts against the Ground, so they relate to and complement each other. Although the idea of Figure

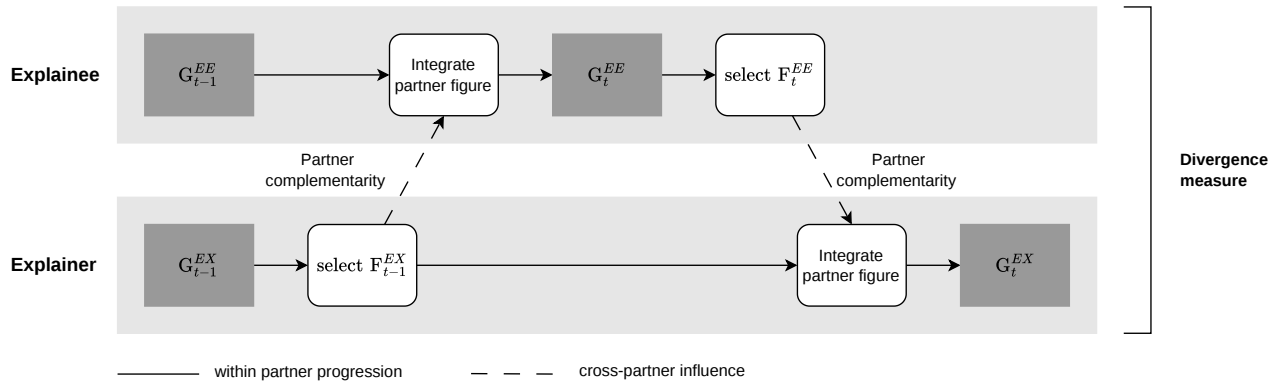


Fig. 1: Complementarity: interlocutors maintain ground semantics (G) performing figural actions (F). Figure (F) at time $t - 1$ serves as complementary signal to the ground semantics at t , thereby changing ground semantics as dialogue unfolds.

TABLE I: Excerpt of transcription from the game explanation task. P1 and P2 designate the interlocutors.

Original German transcription	English translation
P1: "...wie Vier gewinnt?"	P1: "...like Connect Four?"
P2: "das ist <i>nicht</i> wie bei Vier gewinnt... eher wie Tic-Tac-Toe..."	P2: "it is <i>not</i> like Connect Four... it is more like Tic-Tac-Toe..."
P1: "es ist <i>nicht</i> wie Vier gewinnt ... sondern ..."	P1: "so it is <i>not</i> like Connect Four ... but rather ..."
P2: "nein, nein ... die Positionierung ist wie bei Tic-Tac-Toe"	P2: "no, no ... the positioning is like in Tic-Tac-Toe."
P1: "... sonst hätte ich gedacht, dass es aufeinander ist ... aber"	P1: "... otherwise I would have thought that it was stacked upright ... but"

and Ground was originally developed in Gestalt psychology and later adapted in Talmy’s seminal work on Cognitive Semantics [17] to investigate the conceptual structuring of two events in relation to one another, here, we treat Figure and Ground as event schemas that continuously evolve along the dialogue. Talmy describes Figure and Ground as an abstract semantic relation between a focal element and a backgrounded reference entity [17]. According to Talmy, “the Ground is a reference entity, one that has a stationary setting relative to a reference frame, with respect to which the Figure’s site, path, or orientation is characterized” [17, p. 184]. For example, in the utterance by P1 (“it is not like Connect Four”), the Ground can be the ongoing global event [Game Comparison], while the Figure can be the principal constituent under focus [Connect Four].

In relation to scaffolding, we conceptualize negation as becoming salient at moments when interaction dynamics begin to diverge and existing presumptions need to be redirected. With such scaffolding, the explainer aims to regulate elements of knowledge that the explainee needs for successful understanding and task completion [6]. In our suggested terms, negation can be defined as a contrastive guidance mechanism through which current information is contrasted against the already established context, thereby reorienting the dialogue toward coherence [18], [19]. Motivated by these observations in dialogue, our aim is to model a specific

form of dialogue characterized by inherent knowledge asymmetry. More particularly, our aim in this paper is twofold: (a) to derive a model that formalizes complementarity and divergences during dialogue, and (b) to investigate how and when negations are used during dialogue to scaffold the partner. To this end, we use data from a human–human explanation setting in which divergences arise on multiple occasions in the course of the dialogue. The core components of our model are organized around two principles identified above, namely *complementarity* and *scaffolding*. With respect to negation, we hypothesized that participants use negation primarily at those moments in dialogue when divergence in the interaction becomes prominent.

II. MODEL ASSUMPTIONS

The model component and its processes are depicted in Figure 1. Building on the preceding discussion, the model encapsulates two processes: Complementarity and Divergence. The complementarity principle [8], [12], states that shared semantics are distributed in the joint environment and continuously gain from (inter)actions. This complementarity is formalized in the model through cross-partner actions on the established Ground and is depicted by dashed lines. These actions are highlighted as the white boxes and labeled as Figure (F). Thus, we define Figure as the currently relevant event schema, that is, the focal entity under attention and brought up by the partner to complement the emerging semantics or Ground, as denoted by G .

A. Figure and Ground as evolving event schemas

Extending Talmy, we define Ground as a dynamic semantic structure on which the complementarity builds through participants’ actions [17]. Since, the focal entity in Talmy’s framework is denoted by Figure (F), for consistency, we keep this definition as well. Crucially, as compared to Talmy’s definition where the Ground is a static entity, in our model, the Ground is complemented by Figure. Together, Figure and Ground form a dynamic relation: the Figure is relevant to the Ground, and the Ground affords the Figure. The emergence of Ground through the Figure’s complementarity allows us to formalize divergence in latent semantics over time. This

is important for dialogues with knowledge asymmetry. In the literature, the process of scaffolding has been suggested to bridge knowledge divergences [6]. Subsequently, we use semantic divergence to show that scaffolding through negation is more plausible when knowledge divergences become apparent between two partners.

III. METHODOLOGY

Figure 3 depicts the concrete pipeline underlying our method. We first describe the operations used to extract Figure and Ground semantics from dialogue, and then formalize the steps through which the measures of Complementarity and Divergence are obtained.

A. Figure–Ground annotation

Following Talmy, we obtained human-annotated Figure–Ground labels from game-explanation dialogues ($N = 67$). In addition, we obtained these representations through automated extraction using an instruction-tuned LLM (Meta-Llama-3.1-8B-Instruct) [20]. We draw on the human annotation scheme introduced by Fisher et al. [21], who manually coded the Quarto explanation dialogues in ELAN using a hierarchical explanation-node scheme. In this scheme, each utterance is linked to content nodes organized into top-level block nodes (e.g., *Players*, *Board*, *Features*) and increasingly specific subordinate nodes, resulting in a structured semantic map of the explanation content (Figure 2). To make these annotations comparable to our theoretical Figure–Ground conception, we collapsed the hierarchy into a binary distinction: all non-terminal nodes in an annotated branch were treated as Ground, whereas only terminal leaf nodes were treated as Figure.

For the automated pipeline, the model was prompted to produce exactly one Figure and one Ground for each speaker turn block. The model was provided with Talmy’s definitions of Figure and Ground and instructed to follow these rules strictly during extraction¹. The outputs were constrained to short English text so that each turn was associated with a single compact event-schema representation. This step yielded text-level Figure and Ground descriptions aligned with each turn unit, making them comparable to the human-annotated Figure–Ground labels.

For further quantitative operations, we mapped these representations into a high-dimensional latent semantic space using the sentence-transformer embedding model (all-mpnet-base-v2) [22], [23], which is specifically trained to produce semantically meaningful sentence embeddings. Concretely, let u_t denote the extracted Figure string at time t , and let v_t denote the extracted Ground string at time t . We compute their embeddings as:

$$\mathbf{F}_t = \text{embed}(u_t), \quad \mathbf{G}_t = \text{embed}(v_t).$$

This procedure yields 768-dimensional vector representations for both Figure and Ground at each time step and for each partner. All Figure and Ground embeddings were

¹The full LLM prompt is provided in the Supplementary Materials.

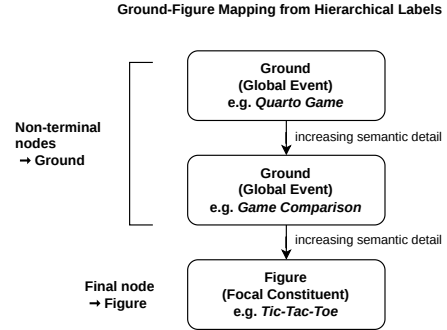


Fig. 2: Representation mapping based on explanation-node scheme [21]. All non-terminal nodes are treated as Ground, whereas terminal leaf nodes are treated as Figure.

L2-normalized to unit norm, so the representation preserves semantic direction while removing magnitude effects.

B. Partner Complementarity

According to our theoretical assumptions of complementarity in dialogue [8], [12], the currently evolving joint semantics is complemented by the actions of the partners involved. In formal terms, this current action is denoted by Figure \mathbf{F} for the respective partners \mathbf{F}^{EE} (Explainee) and \mathbf{F}^{EX} (Explainer) and directly integrates to the Ground \mathbf{G} established so far in the dialogue. This is shown by the dashed line in Figure 1 and labeled as *partner complementarity*. Since we assume that Ground semantics evolve across turns and across partners, we define complementarity at time t as a linear integration of three sources: the partner’s immediately preceding Figure \mathbf{F}_{t-1} , the interlocutor’s own previous Ground \mathbf{G}_{t-1} , and the current observed Ground $\tilde{\mathbf{G}}_t$. In this way, the updated Ground state is not treated as a standalone representation, but as shaped by the partner’s foregrounded action \mathbf{F}_{t-1} , the interlocutor’s own semantic continuity \mathbf{G}_{t-1} , and the currently available evidence $\tilde{\mathbf{G}}_t$. Formally, for the explainer and explainee, respectively, we write:

$$\mathbf{G}_t^{\text{EX}} = \alpha \mathbf{F}_{t-1}^{\text{EE}} + \beta \mathbf{G}_{t-1}^{\text{EX}} + \gamma \tilde{\mathbf{G}}_t^{\text{EX}},$$

$$\mathbf{G}_t^{\text{EE}} = \alpha \mathbf{F}_{t-1}^{\text{EX}} + \beta \mathbf{G}_{t-1}^{\text{EE}} + \gamma \tilde{\mathbf{G}}_t^{\text{EE}}$$

The parameters α , β , and γ determine the relative contribution of each source in shaping the updated Ground state. Specifically, α weights the contribution of the partner’s immediately preceding Figure, capturing cross-partner influence through the most recent foregrounded action; β weights the contribution of the interlocutor’s own previous Ground state, capturing local semantic continuity; and γ weights the contribution of the observed Ground, that is, the immediate semantic evidence available at time t . In our implementation, we initially set these parameters to $\alpha = \beta = \gamma = 1$, thereby assigning equal weight to each component. Theoretically, however, these weights could also be learned and allowed to change as the dialogue progresses, allowing the model to differentially attune to cross-partner

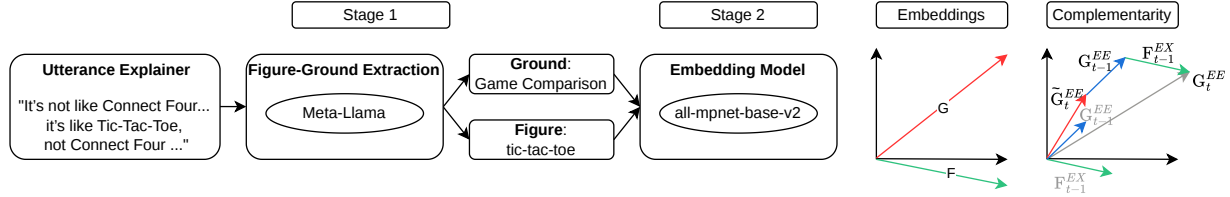


Fig. 3: Pipeline for Figure–Ground extraction from an explanatory utterance and mapping its embeddings, along with a geometric illustration of the complementarity update for the explainee’s ground state.

input, self-memory, and current evidence. We leave this aspect open for future examination. Figure 3 geometrically illustrates the complementarity operation for the evolving Ground representation of one partner.

C. Divergence

To quantify whether emerging semantics are currently operating on a similar representational ground, we measure the distance between explainee’s and explainer’s ground states after each complementarity operation at time t (section III-B). These Grounds states are continuously updated latent semantic representations that emerge through Figure’s complementarity. Hence, we define Ground divergence as cosine distance:

$$D(\mathbf{G}_t^{EE}, \mathbf{G}_t^{EX}) = 1 - \cos(\mathbf{G}_t^{EE}, \mathbf{G}_t^{EX}),$$

where cosine similarity is given by:

$$\cos(\mathbf{G}_t^{EE}, \mathbf{G}_t^{EX}) = \frac{\mathbf{G}_t^{EE} \cdot \mathbf{G}_t^{EX}}{\|\mathbf{G}_t^{EE}\| \|\mathbf{G}_t^{EX}\|}$$

Cosine similarity in a distributional semantic space captures directional properties. Since our grounds are continuously updated through participants’ actions (\mathbf{F}) as dialogue unfolds and may differ depending on how divergent they are at a given point in time, as divergence increases, we expect the similarity value to decrease. For easier interpretation we converted similarity into a distance via $1 - \cos(\cdot)$, so that larger values directly indicate stronger divergence. Intuitively, $D(\mathbf{G}_t^{EE}, \mathbf{G}_t^{EX}) \approx 0$ indicates low divergence where the partners are effectively standing on the same or a nearly similar semantic ground and larger values correspond to an increase in divergence.

Since the divergence model relies on processuality rather than instantaneous semantic similarity, we do not treat raw distance as an indicator of divergence, but rather how it is changing with time. We therefore define divergence change ΔD_t as a processual measure of divergence as the dialogue unfolds:

$$\Delta D_t = D(\mathbf{G}_t^{EE}, \mathbf{G}_t^{EX}) - D(\mathbf{G}_{t-1}^{EE}, \mathbf{G}_{t-1}^{EX}).$$

The ΔD_t was also suitable for measuring the effect of divergence change on scaffolding through negation which needed a processual scale rather than static to capture divergence and its change.

IV. SCAFFOLDING DIVERGENCE THROUGH NEGATION

We theorized that negation, acting as a scaffolding mechanism, reorients divergence toward global coherence [6], [15]. More specifically, we predicted negation to be more frequent in the dialogue when divergence becomes prominent. To investigate this question, we used the measure for change in divergence ΔD_t to predict negations across turns in the dialogue. At the preprocessing stage, we excluded negations that were not purely semantically informative from human annotations, such as negations that were just literal yes–no questions or were fillers such as “I don’t know, maybe...”. The goal was to retain negations that scaffolded previous moves at a meta-level or revised already established assumptions. This distinction was made based on descriptive and metalinguistic negation used for different purposes in dialogue [15], [24].

A. Analysis

We fitted a Bayesian hierarchical logistic regression model by treating the divergence change score ΔD_t as a predictor of negation [25]. Each turn was coded as a binary outcome $Negation_t \in \{0, 1\}$, with 1 = negation present and 0 = non-negation. Because ΔD_t is a change score, its sign is directly meaningful in our setting: $\Delta D_t > 0$ indicates increasing divergence, whereas $\Delta D_t < 0$ indicates decreasing divergence. Before fitting the models ΔD_t was scaled as $\Delta D_t^{(z)} = \Delta D_t / \text{SD}(\Delta D_t)$, so that a one-unit increase corresponds to an increase of one standard deviation in divergence change. Formally, for observation i nested in dyad $d[i]$, we used a Bernoulli likelihood with a logit link:

$$\begin{aligned} y_i &\sim \text{Bernoulli}(p_i), \\ \text{logit}(p_i) &= \eta_i, \\ \eta_i &= \beta_0 + b_{0,d[i]} + (\beta_1 + b_{1,d[i]}) \Delta D_{z,i}. \end{aligned}$$

Here, β_0 and β_1 are the population-level intercept and effect of divergence-change on negation. The dyad-level terms $b_{0,d}$ and $b_{1,d}$ capture baseline differences in how frequently dyads use negation, and allow for the dyads to vary for the effect of the ΔD on negation. Random effects were modeled with partial pooling, so that dyad-specific estimates are regularized toward the population estimate. Priors for the intercept and slope were set to be weakly regularized and we assessed the robustness of our inference using power-scaling sensitivity analysis, calibration analysis, and posterior

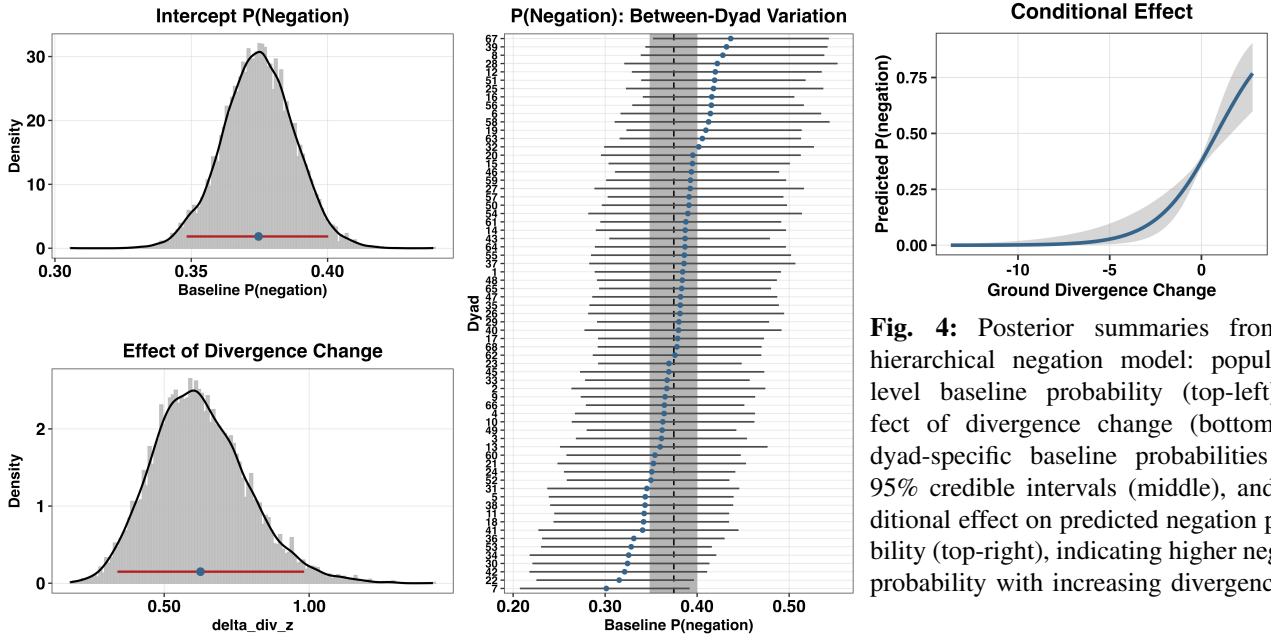


Fig. 4: Posterior summaries from the hierarchical negation model: population-level baseline probability (top-left), effect of divergence change (bottom-left), dyad-specific baseline probabilities with 95% credible intervals (middle), and conditional effect on predicted negation probability (top-right), indicating higher negation probability with increasing divergence.

predictive checks [26]–[28].²

TABLE II: Posteriors for the LLM-derived representations.

Parameter	Mean	Est.Error	95% CrI
baseline $P(\text{negation})$	0.375	0.013	[0.348, 0.400]
$b_{\text{delta_div_z}}$	0.625	0.164	[0.339, 0.983]

V. RESULTS

Model posterior summaries for the baseline negation probability and the effect of divergence change are shown in Figure 4 and Table II. The model-predicted baseline probability of negation use was $P(\text{negation} = 1) \approx 0.375$ (95% CrI [0.348, 0.400]), indicating that, when divergence change is at its mean (i.e., $\Delta D_z = 0$) negation is already a fairly common resource used in the dialogue. This baseline pattern was observed across dyads, with dyad-specific intercepts varying around the grand mean (Figure 4, middle), suggesting stable between-dyad differences in overall negation use.

Crucially, the posterior for the divergence-change effect was positive. The estimate for $b_{\text{delta_div_z}}$ indicates that increases in divergence (positive ΔD_z) are associated with a higher probability of negation ($\beta = 0.625$, 95% CrI [0.339, 0.983]). In other words, as divergence becomes more prominent, negation becomes increasingly plausible as a scaffolding mechanism. This relationship is also visible in the conditional effect plot, as the divergence score increases, the predicted probability of negation increases (Figure 4, right).

VI. MODEL VALIDATIONS

For internal validation, we used dyad-wise cross-validation (leave-one-dyad-out) via PSIS-LOO and additionally inspected out-of-sample calibration [27], [29], [30]. Overall, PSIS-LOO indicated good out-of-sample reliability for the hierarchical model. Table III compares three models: a

null model with only dyad-level intercepts, a fixed-slope model including divergence change (delta_div_z), and a random-slope model allowing dyads to vary in the strength of this effect. Both models that include delta_div_z outperform the null model in expected out-of-sample predictive accuracy (fixed slope: $\Delta \text{elpd}_{\text{loo}} = 10.27$; random slope: $\Delta \text{elpd}_{\text{loo}} = 13.74$). The random-slope model is nominally best, but its advantage over the fixed-slope model was small relative to the uncertainty ($\Delta \text{elpd} = 3.48$, $\text{SE} = 4.33$). Taken together, divergence change provided clear predictive value for negation, while allowing dyad-specific slope variation added only a limited additional gain.

TABLE III: Leave-one-out model comparison for negation prediction. Higher elpd_{loo} shows better prediction.

Model	elpd_{loo}	$\text{elpd}_{\text{diff}}$	$\text{SE}(\text{diff})$	LOOIC
m_{rs}	-1483.63	0.00	0.00	2967.25
m_{fix}	-1487.10	-3.48	4.33	2974.20
m_{null}	-1497.37	-13.74	6.48	2994.74

$$m_{\text{null}} : 1 + (1 \mid \text{dyad_id}); m_{\text{fix}} : \text{delta_div_z} + (1 \mid \text{dyad_id}); m_{\text{rs}} : \text{delta_div_z} + (1 + \text{delta_div_z} \mid \text{dyad_id}).$$

A. Validation on Human Annotations

For external validation, we applied our best-fitting model m_{rs} to the human-annotated Figure–Ground representations. Using the full model with random effects, the model showed above-chance discrimination and good calibration (Table IVa). In addition, the mean predicted negation probability closely matched the observed event rate, indicating reliable average calibration (Figure 5A).

Beyond validating the original model on human-annotated data, we next asked whether the same theoretical relationship between divergence and negation would also hold when Figure–Ground representations are manually derived from human coding. To test this, we refitted the existing random-slope model to the human-annotated Figure–Ground labels

²Model diagnostics results are provided in Supplementary Materials

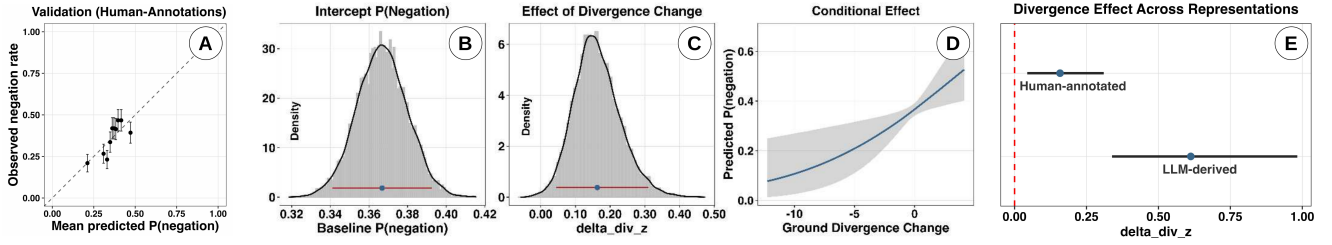


Fig. 5: Validation and posterior results for the human-annotated Figure–Ground representation. (A) Calibration plot of LLM-model m_{TS} on human-annotations. (B) Baseline negation probability; (C) Effect of divergence change; (D) Conditional effect on predicted negation probability; (E) Posterior effects comparisons across Human-annotated and LLM derived models.

TABLE IV: Validation and posterior summary for the human-annotated Figure–Ground representation. (A) Validation performance of model m_{TS} . (B) Posterior summary for the human-annotated Figure–Ground model.

(a) Validation performance

n	Event rate	Pred. $P(\text{neg})$	AUC	Brier
2290	0.362	0.360	0.596	0.228

(b) Posterior summary

Parameter	Mean	Est.Error	95% CrI
baseline $P(\text{neg})$	0.367	0.013	[0.341, 0.393]
$b_{\text{delta_div_z}}$	0.163	0.067	[0.044, 0.310]

(m_{TS}^{nodes}), using the identical model structure as in the original analysis, but with divergence-change values computed from the human-derived Figure–Ground representations.

As shown in Table IVb, the estimated effect of divergence change remained positive ($\beta = 0.16$, 95% CrI [0.067, 0.310]). This indicates that a positive increase in ground divergence was again associated with a higher probability of observing negation (Figure 5C). The same pattern is visible in the conditional effects plot, where positive divergence-change values predict higher negation probabilities (Figure 5D). Although the human-annotation-derived effect is smaller than the LLM-derived effect, crucially, the direction of the relationship is preserved across annotation sources (Figure 5E), validating our theoretical position that participants are more likely to use negation as a scaffolding mechanism when divergence during dialogue is higher.

VII. DISCUSSION

For asymmetric interactions, we conceptualized dialogue as processual and dynamic, where common knowledge is built and continuously emerges through partners’ complementarity [8], [10], [12]. A shift toward this view, from speaker–listener alignment models, has changed our way of treating interaction no longer primarily as an intent-inference task (e.g., [1], [31]), where the semantics are disjunct and located *within* each participant, but rather as a process of scaffolding common knowledge in a joint semantic space [2]. We also highlighted that complementarity in asymmetrical interactions as such would show a tendency to elicit momentarily processual convergence/divergence, where divergences are scaffolded through linguistic or nonlinguistic

mediums to attain global coherence (see also, [12]) and task completion. In this sense, divergences are not side loops showing conversational breakdowns or repairs [32], [33], but are inherent and graded characteristics of the processuality, particularity in asymmetric interaction.

Given this theoretical background, we formalized complementarity across and within the participants involved in a co-constructive game explanation task. We showed that the complementarity of the participants’ actions towards the common knowledge can be shown as an integration of the partner’s action with the already developing Grounds in a joint semantics space [34]. In formal terms, the additivity of event schemas in high-dimensional space provided an abstract representation of processuality due to associative principles. This is not new, as prior work has shown that the linear additivity of representations in high-dimensional vector spaces can give rise to joint semantics, which we used here to formalize complementarity [34]. A similar principle of distributional semantics also enabled us to derive semantic distance via cosine similarity, which was subsequently used to formalize divergence. Building only on linear operations, these methods remain interpretable and, more importantly, scalable, allowing the model to be extended to multi-party dialogue in which complementarity and divergence continuously emerge. Although we began by modeling complementarity through associative principles and divergence through cosine similarity, the framework leaves open the possibility of estimating both through more sophisticated approaches. Even so, the theoretical framing grounded in Talmy’s notion of Figure and Ground [17], applied to both LLM-derived and human-annotated event schemas, remained consistent across methods: in both cases, positive change in divergence reliably predicted negation. This suggests that the theoretical effect was directionally robust and aligned across the two operationalizations, albeit weaker for the human annotation than for the LLM-derived model. We believe that this difference may be explained by variability between human coders and LLM-generated annotations, although a direct comparison between the two remains an open question for future investigation.

One of our central goals was to investigate how partners are scaffolded through divergences using linguistic means. Following this, we found that *negation* is one of the key scaffolding means emerging in the processual timeline to restore coherence when divergence becomes inevitable. Di-

alogical systems built on principles of complementarity can harness such linguistic devices online to navigate divergence, thereby supporting learning and other forms of dialogue characterized by knowledge asymmetry. Another crucial point will be to extend our account of scaffolding as support that is not only provided but also gradually faded to foster long-term learning [6]. A next step in this direction is to operationalize the present framework as an online scaffolding mechanism. Since the model estimates temporal changes in semantic divergence, it could provide a basis for real-time detection of emerging misunderstandings in interactive systems. Such estimates could in turn support adaptive scaffolding strategies, such as clarification, rephrasing, or contrastive explanation [18]. An important open question is whether these computations remain sufficiently robust under online conditions, where utterances are partial, noisy, and time-critical. Addressing this question would help move the present framework from retrospective analysis toward real-time scaffolding.

REFERENCES

- [1] M. J. Pickering and S. Garrod, *Understanding Dialogue: Language Use and Social Interaction*, 1st ed. Cambridge University Press, 2021.
- [2] K. J. Rohlfing, P. Cimiano, I. Scharlau, T. Matzner, H. M. Buhl, H. Buschmeier, E. Esposito, A. Grimminger, B. Hammer, R. Hab-Umbach, I. Horwath, E. Hullermeier, F. Kern, S. Kopp, K. Thommes, A.-C. Ngonga Ngomo, C. Schulte, H. Wachsmuth, P. Wagner, and B. Wrede, "Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 3, pp. 717–728, 2021.
- [3] M. Chromik and A. Butz, "Human-xai interaction: A review and design principles for explanation user interfaces," in *Human-Computer Interaction – INTERACT 2021*, ser. Lecture Notes in Computer Science, C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, Eds. Cham: Springer, 2021, vol. 12933, pp. 619–640.
- [4] M. Ridley, "Human-centered explainable artificial intelligence," *Journal of the Association for Information Science and Technology*, vol. 76, no. 1, pp. 98–120, 2025.
- [5] U. Schmid and B. Wrede, "What is missing in xai so far?" *KI - Künstliche Intelligenz*, vol. 36, no. 3–4, pp. 303–315, 2022.
- [6] D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving," *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89–100, 1976.
- [7] E. Hutchins, "The distributed cognition perspective on human interaction," in *Roots of Human Sociality: Culture, Cognition and Interaction*, N. J. Enfield and S. C. Levinson, Eds. Oxford: Berg, 2006, pp. 375–398.
- [8] R. Fusaroli, J. Rączaszek-Leonardi, and K. Tylén, "Dialog as interpersonal synergy," *New Ideas in Psychology*, vol. 32, pp. 147–157, 2014.
- [9] H. De Jaegher, E. Di Paolo, and S. Gallagher, "Can social interaction constitute social cognition?" *Trends in Cognitive Sciences*, vol. 14, no. 10, pp. 441–447, 2010.
- [10] H. De Jaegher and E. A. Di Paolo, "Participatory sense-making: An enactive approach to social cognition," *Phenomenology and the Cognitive Sciences*, vol. 6, no. 4, pp. 485–507, 2007.
- [11] H. H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.
- [12] E. Gregoromichelaki, A. Eshghi, C. Howes, G. J. Mills, R. Kempson, J. Hough, P. G. T. Healey, and M. Purver, "Language and Cognition as Distributed Process Interactions," *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue*, 2022.
- [13] J. Rączaszek-Leonardi, A. Dębska, and A. Sochanowicz, "Pooling the ground: Understanding and coordination in collective sense making," *Frontiers in Psychology*, vol. 5, p. 1233, 2014.
- [14] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn Taking for Conversation*," in *Studies in the Organization of Conversational Interaction*, J. Schenkein, Ed. Academic Press, 1978, pp. 7–55.
- [15] L. R. Horn, *A Natural History of Negation*. Chicago, IL, USA: University of Chicago Press, 1989.
- [16] B. Kaup and C. Dudschig, "Understanding Negation: Issues in the processing of negation," in *The Oxford Handbook of Negation*, V. Déprez and M. T. Espinal, Eds. Oxford University Press, 2020.
- [17] L. Talmy, *Toward a Cognitive Semantics, Volume 1: Concept Structuring Systems*. Cambridge, MA, USA: MIT Press, 2000.
- [18] A. Groß, A. Singh, N. C. Banh, B. Richter, I. Scharlau, K. J. Rohlfing, and B. Wrede, "Scaffolding the human partner by contrastive guidance in an explanatory human-robot dialogue," *Frontiers in Robotics and AI*, vol. 10, 2023.
- [19] A. Singh and K. J. Rohlfing, "Contrastive verbal guidance: A beneficial context for attention to events and their memory?" *Cognitive Science*, vol. 49, no. 8, p. e70096, 2025.
- [20] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten *et al.*, "The llama 3 herd of models," *arXiv preprint arXiv:2407.21783*, 2024.
- [21] J. Fisher, A. Robrecht, S. Kopp, and K. Rohlfing, "Exploring the semantic dialogue patterns of explanations - a case study of game explanations," in *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Maribor, Slovenia, 2023.
- [22] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, 2019, pp. 3982–3992.
- [23] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, "Mpnet: Masked and permuted pre-training for language understanding," in *Advances in Neural Information Processing Systems 34 (NeurIPS 2020)*, 2020, pp. 16 857–16 867.
- [24] W. Köller, *Formen und Funktionen der Negation: Untersuchungen zu den Erscheinungsweisen einer Sprachuniversalie*. De Gruyter, 2016.
- [25] P.-C. Bürkner, "brms: An r package for bayesian multilevel models using stan," *Journal of Statistical Software*, vol. 80, pp. 1–28, 2017.
- [26] N. Kallioinen, T. Paananen, P.-C. Bürkner, and A. Vehtari, "Detecting and diagnosing prior and likelihood sensitivity with power-scaling," *Statistics and Computing*, vol. 34, p. 57, 2024.
- [27] B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, and E. W. Steyerberg, "Calibration: The achilles heel of predictive analytics," *BMC Medicine*, vol. 17, p. 230, 2019.
- [28] A. Gelman, X.-L. Meng, and H. Stern, "Posterior predictive assessment of model fitness via realized discrepancies," *Statistica Sinica*, vol. 6, no. 4, pp. 733–807, 1996.
- [29] A. Vehtari, A. Gelman, and J. Gabry, "Practical bayesian model evaluation using leave-one-out cross-validation and waic," *Statistics and Computing*, vol. 27, no. 5, pp. 1413–1432, 2017.
- [30] A. Vehtari, D. Simpson, A. Gelman, Y. Yao, and J. Gabry, "Pareto smoothed importance sampling," *Journal of Machine Learning Research*, vol. 25, no. 72, pp. 1–58, 2024.
- [31] N. D. Goodman and M. C. Frank, "Pragmatic Language Interpretation as Probabilistic Inference," *Trends in Cognitive Sciences*, vol. 20, no. 11, pp. 818–829, 2016.
- [32] M. Dingemanse, S. Roberts, J. Baranova, J. Blythe, P. Drew, S. Floyd, R. Gisladottir, K. Kendrick, S. Levinson, E. Manrique, G. Rossi, and N. Enfield, "Universal principles in the repair of communication problems," *PLOS ONE*, vol. 10, p. e0136100, 2015.
- [33] S. Albert and J. P. de Ruiter, "Repair: The interface between interaction and cognition," *Topics in Cognitive Science*, vol. 10, no. 2, pp. 279–313, 2018.
- [34] D. Kartsaklis, "Compositional operators in distributional semantics," *Springer Science Reviews*, vol. 2, no. 1–2, pp. 161–177, 2014.