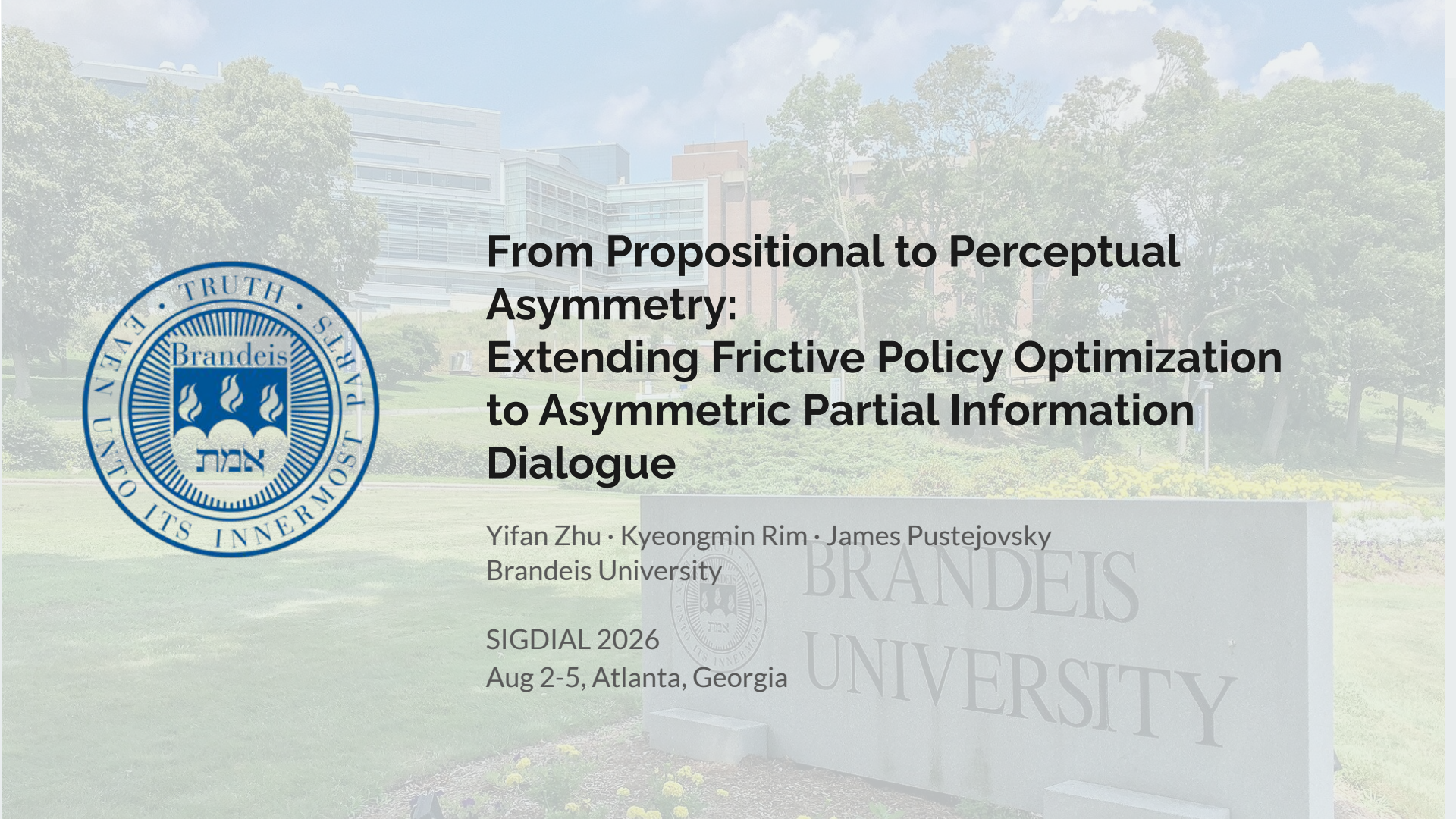


From Propositional to Perceptual Asymmetry: Extending Frictive Policy Optimization to Asymmetric Partial Information Dialogue

Yifan Zhu · Kyeongmin Rim · James Pustejovsky
Brandeis University

SIGDIAL 2026
Aug 2-5, Atlanta, Georgia

Motivation

A Communication Breakdown in interaction when participants having different information.

The **difference in perception** results in a **difference in understanding.**

AI-Generated image

Motivation

AI-Generated images

Research Question

When two people are working from different information, **can LLMs spot their misunderstandings better by seeing only one person's view, or by seeing both?**

- Finding: Restricting the model to a single participant's view beats, in all three models we tested.
- Contribution: We show how to evaluate friction from inside one participant's view and that the same signal predicts the repairs the dialogue actually makes.

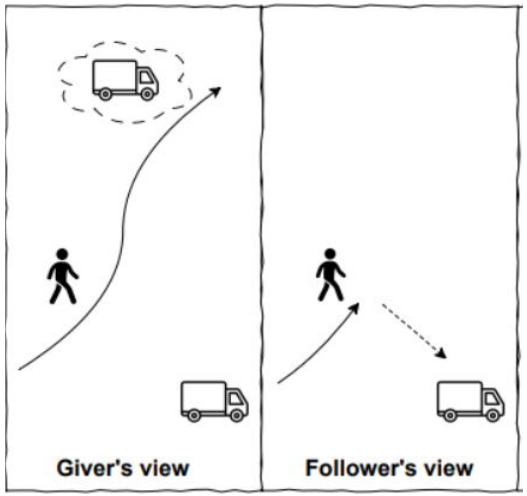
Outline of the talk

- Dataset: where the data comes from, and what counts as a misunderstanding.
- Probing methodology: how we test whether perspective helps.
- Results and discussion: what we found, and what follows from it.

Dataset - Maptask[1]

- 2 participants, 2 maps, different landmarks
  - Giver: has a route marked on the giver's map and must guide the other participant
  - Follower: reproduce this route on the follower's map through verbal communication alone.

- Why this breakdown happened?
  - Because participants have different information.



Giver: Go to **the parked van**

the upper van

Follower: ok, to the **bottom.**

The bottom van

Giver: NO!

The bottom van != The upper van

[1]Anne H. Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson, and Regina Weinert. 1991. The HCRC Map Task corpus. Language and Speech, 34(4):351–366

What's the Problem?

- Each participant perceives different evidence

  → each constructs different propositional content from what they perceive

  → their belief states diverge

  → that divergence surfaces as ambiguity, misreference, and misunderstanding in the dialogue.
- The giver's evidence set contains two vans; the follower's contains one. Same words, different grounding, because different evidence.

What's the Extension?

- Work on common ground and common-ground tracking with LLMs (e.g., DeliData, Karadzhov et al., 2023) , including our own (WTD, Khebour et al., 2024), assumes the evidence is spoken or written text, shared by all participants.
- Two different Asymmetries
  - **Propositional asymmetry:** participants receive the ***same*** **information** and draw different conclusions from it. Addressed in prior work.
  - **Perceptual asymmetry**: participants have private perceptual access, so their ***evidence itself*** **differs.** Not addressed in prior work.
- We evaluate friction from within each participant's own perspective rather than over a shared record.

Ibrahim Khalil Khebour, Mariah Bradford, Kenneth Lai, Christopher Tam, Jingxuan Tu, Benjamin A.Ibarra, Nathaniel Blanchard, Nikhil Krishnaswamy,and James Pustejovsky. 2024. Common ground tracking in multimodal dialogue. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 3587–3602.

Georgi Karadzhov, Tom Stafford, and Andreas Vlachos.2023. DeliData: A dataset for deliberation in multiparty problem solving. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2):1–25.

Corpus Evidence

HCRC MapTask, 13,077 reference expressions across 128 dialogues, annotated per expression for the giver's *intended* and the follower's *interpreted* landmark. Each expression is aligned, pending, or misunderstood. (Li et al 2026)

- Finding one, **multiplicity:** The parked-van example is a *multiplicity* case: one participant's map has two instances of a landmark, the other's has one. Multiplicity is 7% of reference expressions but produces 12% of misunderstandings, six times the corpus average.
- Finding two, **pending:** Misalignment surfaces in two places: outright misunderstanding, and *pending*, where grounding never completes. 38% of pending cases are multiplicity.

Other discrepancy types exist in the corpus and produce other friction, but multiplicity is the dominant source; the rest of the talk focuses on it.

Nan Li, Albert Gatt, and Massimo Poesio. 2025. Grounded misunderstandings in asymmetric dialogue: A perspectivist annotation scheme for MapTask. LREC 2026

The Perspective-Based friction detection

4 failure modes, corresponding to perspective-based diagnostic (Pustejovsky & Krishnaswamy 2026)

Value conflict - Is the expression I am producing (or accepting) uniquely identifying for me, and if not, have I flagged it?
Hazard - Am I (or my partner) acting confidently on a grounding that I have reason to doubt?
Contradiction - Given the spatial claims I attribute to my partner, is there a candidate on my map that satisfies both?
Uncertainty - Can I commit to a referent, or does my evidence admit multiple candidates?

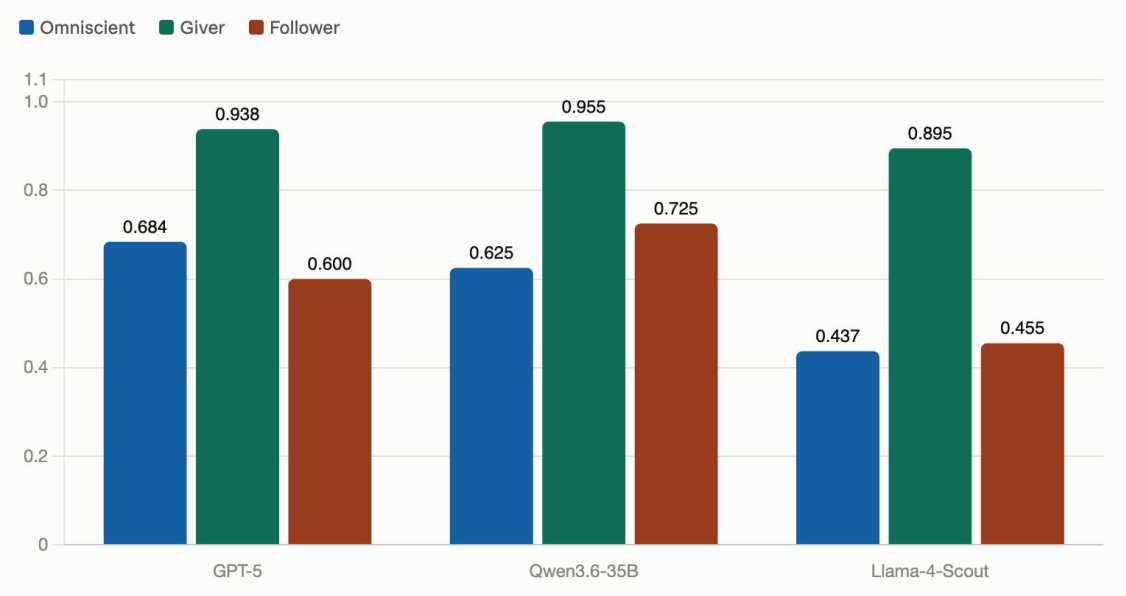
James Pustejovsky and Nikhil Krishnaswamy. 2026.Frictive policy optimization for LLMs: Epistemic intervention, risk-sensitive control, and reflective alignment. arXiv preprint arXiv:2604.25136.

Probe1: Can the LLM understand the forced ambiguity?

- Omniscient: Both maps + dialogue history
  - Question: Do giver and follower refer to the same physical landmark?
- Giver-only: Giver's map only + dialogue history
  - Question: *is the target RE underspecified on the giver's own map?*
- Follower-only: Follower's map only + dialogue history
  - Question: does the giver's spatial description conflict with the follower's map?

Perspectival diagnostics are more reliable than global judgments

- Giver condition's higher accuracy reflects the simplicity of within-map multiplicity detection
- The omniscient view degraded
- Gold: annotation from Li et al. (2026)



More probing experiments to complete the logic

- Omniscient: Both maps + dialogue history
  - Is the target RE underspecified with both map?
- Giver-only: Giver's map only + dialogue history
  - Do giver and follower refer to the same physical landmark?
- Follower-only: Follower's map only + dialogue history
  - Do giver and follower refer to the same physical landmark?

Perspectival strategies help in misalignment and underspecified detection

Qwen 3.6

	Giver only (Acc / F1)	Follower only (Acc / F1)	Both maps (Acc / F1)
Q: is the RE underspecified?	**.955/.950**	.725/.167	.725/.626
Q: is the RE misaligned?	**.645/.413**	**.682/.113**	.625/.370

Probe3: Can the LLM act upon detected friction?

- Omniscient: Both maps + dialogue history
  - Target: contradiction detected over a global perspective
  - $\mathrm{Contr}(h, y)$
- Perspective: Giver's map only + dialogue history vs. follower's map only + dialogue history
  - Target: contradiction detected between two perspectives
  - $\mathrm{Contr}(h^{(i)}, y) = \mathbf{1}\left[\mathrm{ref}(y \mid h^{(g)}) \neq \mathrm{ref}(y \mid h^{(f)})\right]$

Help in Friction detection and Predicting Repairs

	Gemma-4-31B		GPT-5	
	Omni.	Persp.	Omni.	Persp.
Repairs (of 13)	5	**13**	12	**13**

- Evaluated against human ”repair” utterance appeared within next 5 mentions of the target RE
- Gold: annotation from Li et al. (2026)

Summary & Future Work

- Collaborative dialogue is not always a **shared-context** phenomenon: when participants hold private perceptual state, friction lives in the gap between their information horizons.
- **Perspectival re-annotation:** Re-annotate OneCommon by applying the grounding cascade separately to each participant's visual perspective, testing whether friction diagnostics generalize to continuously varying asymmetry.
- **Cross-task transfer:** Evaluate the perspective-based patterns on other structurally asymmetric dialogue tasks, including DPIP (Zhu et al., 2026).
- **Epistemic alignment:** Develop perspectival annotations that distinguish surface coordination from genuine belief alignment, revealing silent accommodation and persistent interpretive divergence.

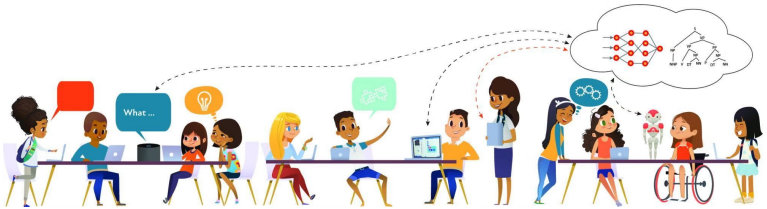
Yifan Zhu, Mariah Bradford, Kenneth Lai, Timothy Obiso, Videep Venkatesha, James Pustejovsky, and Nikhil Krishnaswamy. 2026. Distributed partial information puzzles: Examining common ground construction under epistemic asymmetry. LREC 2026

Acknowledgments

This research was supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grants DRL 2019805 and DRL 2454151.

The opinions expressed are those of the authors and do not represent views of the NSF.

iSAT
NSF AI Institute for Student-AI Teaming



Beyond MapTask

OneCommon (Udagawa & Aizawa, 2019): continuous, not discrete, perceptual overlap — a much harder setting.

- 56% selection-failure rate (vs. 1.8% in MapTask) — but the same principle holds.
- Quantity of friction doesn't predict success: failed dialogues hedge more

Takuma Udagawa and Akiko Aizawa. 2019. A natural language corpus of common grounding under continuous and partially-observable context. In Proceedings of the AAAI Conference on Artificial Intelligence,volume 33, pages 7120–7127.