Hypothesis Memo: Cheap Talk × Information Asymmetry in Bargaining

February 12, 2026 · Status: Pre-design

Research Question

Does cheap talk increase deal rates under one-sided private information but *decrease* deal rates under two-sided private information, in bargaining games where a mutually beneficial deal exists?

Standard theory predicts that asymmetric information causes bargaining failures even when gains from trade exist (Myerson & Satterthwaite, 1983). A natural prescription is to allow communication: let players talk before making offers. But cheap talk—costless, non-binding messages—creates a dual channel. Players can use it to *signal* their position credibly (facilitating deals) or to *misrepresent* strategically (destroying trust). Which channel dominates likely depends on the information structure. Under one-sided asymmetry, only one player has private information and incentive to bluff; the uninformed party can partially update on signals. Under two-sided asymmetry, *both* players have incentive to misrepresent, creating a fog of strategic communication that may leave both parties worse off than silence.

Interestingness Argument

Triviality Scorecard

No sharpening required. The hypothesis already features a crossover interaction effect, competing mechanisms, and a clean experimental design. The prediction that cheap talk *backfires* under two-sided asymmetry is the key novel claim.

Causal Model

Variable Definitions

Testable Implications

1. **Main interaction:** Deal rate in the (cheap talk + one-sided asymmetry) condition > (no talk + one-sided) condition, *but* deal rate in the (cheap talk + two-sided asymmetry) condition < (no talk + two-sided) condition.
2. **Mechanism test:** In the two-sided/cheap-talk cell, the rate of strategic misrepresentation (M_1) should be significantly higher than in the one-sided/cheap-talk cell.
3. **Mediation:** Controlling for misrepresentation intensity (M_1) should attenuate the negative effect of cheap talk under two-sided asymmetry.
4. **Boundary:** If ZOPA is very large (easy deals), the negative effect of cheap talk under two-sided asymmetry should shrink—misrepresentation matters less when there is ample surplus.

Dimension	Score	Reasoning
Prediction surprise	4/5	The crossover interaction—cheap talk *helps* under one-sided but *hurts* under two-sided asymmetry—is counter-intuitive. Most practitioners assume communication always helps. Experts would split on the direction under two-sided asymmetry.
Literature gap	4/5	Cheap talk in bargaining is studied (Farrell & Gibbons, 1989; Valley et al., 2002), and asymmetric-information bargaining is textbook. But the specific 2×2 interaction—cheap talk $\times$ one-sided vs. two-sided asymmetry—lacks direct experimental evidence, especially with LLM agents.
Mechanism specificity	4/5	Two competing mechanisms identified: *credible signaling* (positive channel) vs. *strategic misrepresentation* (negative channel). The design distinguishes them by varying information structure, which shifts the balance between channels.
Boundary conditions	5/5	This *is* an interaction effect: cheap talk $\times$ information structure. The predicted direction *reverses* across conditions—not a universal main effect.
Testability in games	5/5	Deal/no-deal is a clean binary outcome directly observable in transcripts. The treatment is a simple 2×2 between-subjects factorial.
Total	**22/25**	Score ≥ 18: proceed as-is.

Identification Strategy

- **Randomization:** Participants are randomly assigned to one of four conditions in a 2×2 between-subjects design: {one-sided, two-sided} $\times$ {cheap talk, no talk}.
- **Held constant:** Payoff structure (ZOPA = \$20), number of bargaining rounds, AI opponent strategy (identical across conditions), and game framing.
- **Confounds ruled out:** Random assignment eliminates selection effects. Identical ZOPA rules out difficulty differences. Same AI strategy rules out strategic adaptation confounds.
- **Limitations:** (1) Misrepresentation (M_1) and signaling (M_2) are measured from transcripts via content coding, which introduces measurement noise. (2) The design cannot distinguish whether cheap talk backfires because of *active* misrepresentation or *passive* confusion (players talking past each other). (3) LLM agents may not misrepresent in the same way humans do.

Next Steps

This hypothesis is ready for `/design-experiment` to map the 2×2 factorial to concrete game specifications (config, manager prompt, player prompt). The key design challenge is operationalizing "one-sided" vs. "two-sided" private information in a way that feels natural to players and produces enough variation in deal rates to detect the interaction effect in $\sim$50–100 sessions per cell.

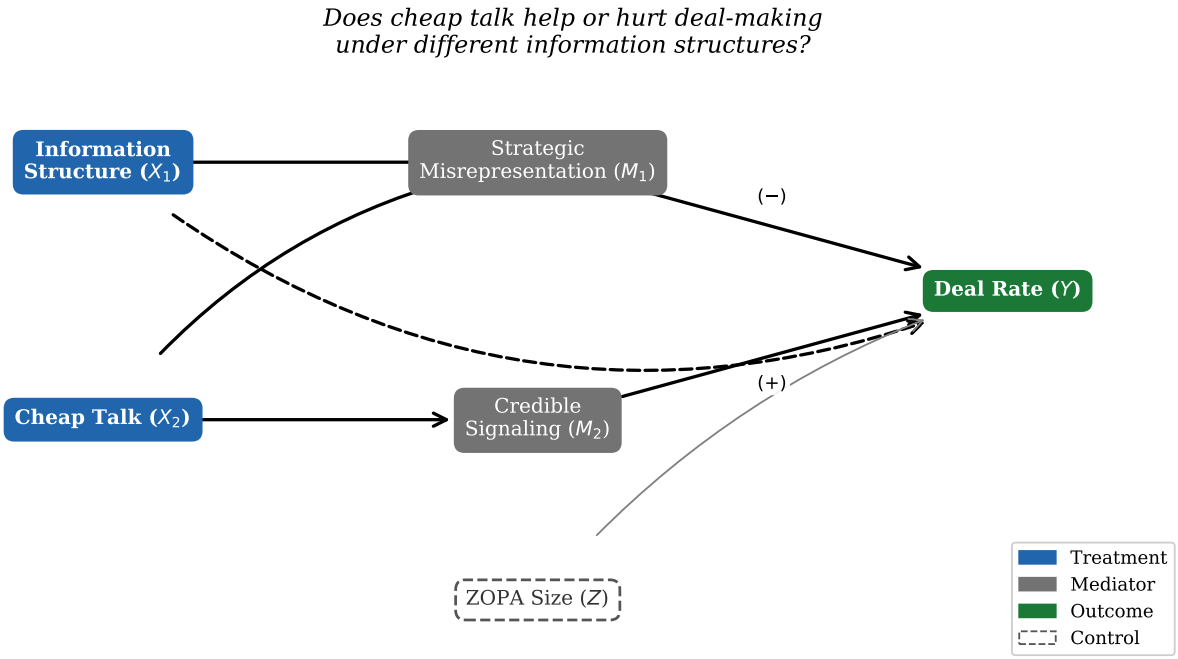

Figure 1: Pearlean causal DAG for the cheap talk × information asymmetry hypothesis. Two treatment variables (blue)—information structure (X_1: one-sided vs. two-sided private information) and cheap talk (X_2: allowed vs. not)—affect deal rate (Y, green) through two competing mediators (gray): strategic misrepresentation (M_1, negative pathway) and credible signaling (M_2, positive pathway). Under two-sided asymmetry, both players misrepresent, making M_1 dominant; under one-sided asymmetry, signaling (M_2) dominates. ZOPA size (Z, dashed border) is held constant across conditions.

Variable	Type	Operationalization	Game Implementation
X_1	Treatment	Information structure: one-sided (only seller knows own cost) vs. two-sided (both have private valuations)	Game variant: one version reveals buyer's valuation to both; the other keeps both valuations private
X_2	Treatment	Communication: cheap talk phase allowed vs. no communication	Game variant: one version includes a free-text chat phase before offers; the other skips to offers
M_1	Mediator	Strategic misrepresentation: degree to which players make false claims about valuations during chat	Measured from chat transcripts: count of claims inconsistent with true valuations
M_2	Mediator	Credible signaling: degree to which players truthfully reveal preference information	Measured from chat transcripts: count of truthful valuation-relevant statements
Y	Outcome	Deal rate: fraction of games where players reach a mutually agreed price within the ZOPA	Binary per game: did both players accept a price? Aggregate: proportion across sessions
Z	Control	Zone of possible agreement (ZOPA) size	Held constant: buyer valuation − seller cost = \$20 in all conditions