Experiment Results:
Same Information, Different Source —
Does Cheap Talk Outperform Verified Disclosure in Bargaining?

February 2026   ·   Status: Post-analysis

1   Research Question

Does the source of partial information about the bargaining zone — game-verified disclosure vs. the opponent's unverified claim — produce opposite effects on bargaining outcomes, even when information content is identical?

We hypothesized a *source-channel reversal*: game-verified disclosure of the seller's approximate cost would activate exploitative anchoring (the buyer pushes toward the seller's limit), reducing deal rates. The same information framed as the opponent's voluntary, unverified claim would activate reciprocal trust, increasing deal rates. The core prediction was a deal-rate ranking: opponent-claimed > no-info (baseline) > game-disclosed.

The hypothesis scored 21/25 on our triviality scorecard (see hypothesis memo), with high marks for prediction surprise (the "less credible information produces better outcomes" reversal) and mechanism specificity (two competing mediators: exploitative anchoring vs. reciprocal trust).

2   Experimental Design

Three-condition between-subjects bargaining experiment with content equivalence across informed conditions.

- **Condition A — No Information (Baseline):** Buyer knows own valuation ($80) but nothing about seller's cost.
- **Condition B — Game-Disclosed (Verified):** Game rules state the seller's cost is approximately $45–$55 (verified information provided by the game).
- **Condition C — Opponent-Claimed (Unverified):** Before Round 1, the AI seller voluntarily states: "My cost is around $45–$55." The game does not confirm this.

All conditions share identical payoff structure (buyer value = $80, seller cost = $50, ZOPA = $30), alternating-offer protocol (max 6 rounds), and AI seller strategy (anchored concession from $70, floor at $55). The AI player prompt is identical across conditions — only the information source differs.

3 Data Overview

Sample

90 simulated sessions: 30 per condition. All sessions used seed 42 for variable draws. Simulated human buyer (gpt-5-nano) played against AI seller (gpt-5-nano), managed by gpt-5.

Data Quality

All 90 transcripts passed the 9-criterion automated checker at 100%. No exclusions.

Summary Statistics

Metric	No Info	Verified	Claimed
N	30	30	30
Deal rate	100%	100%	100%
Mean deal price	$69.47	$69.40	$68.63
(SD)	(1.17)	(1.90)	(1.96)
Mean rounds to deal	2.4	1.5	2.8
(SD)	(1.9)	(1.1)	(1.8)
First-round accept rate	80%	83%	63%
Mean 1st counteroffer[a]	$33.20	$40.48	$44.27
(SD)	(2.94)	(11.49)	(13.67)
Mean buyer earnings	$10.53	$10.60	$11.37
Mean seller earnings	$19.47	$19.40	$18.63
Efficiency	100%	100%	100%

Table 1: Summary statistics by information condition. All 90 sessions reached agreements, yielding full surplus capture (efficiency = 100%) in every case. [a]Among sessions where the buyer counteroffered (n = 6, 5, 11 respectively).

4 Results

Primary Outcome: Deal Rate

The hypothesis predicted differential deal rates across conditions: opponent-claimed > no-info > game-disclosed. **We observed 100% deal rates in all three conditions.** Every one of the 90 sessions reached an agreement. The primary outcome has zero variance, making the central hypothesis untestable with these data.

Diagnosis: The ZOPA ($30 on a $100 price range) was too generous. Combined with the AI seller's concession strategy (willing to accept $\geq$\$55) and the simulated human's acceptance threshold ($\leq$\$75), deals were near-certain regardless of information condition. The design lacked the "zone of possible disagreement" needed to produce variation in deal rates.

Secondary Outcome: Negotiation Speed

This was the only statistically significant finding. A Kruskal–Wallis test revealed significant differences in rounds to deal across conditions ($H(2) = 11.24$, $p = .004$, $\epsilon^2 = .126$). Pairwise comparisons
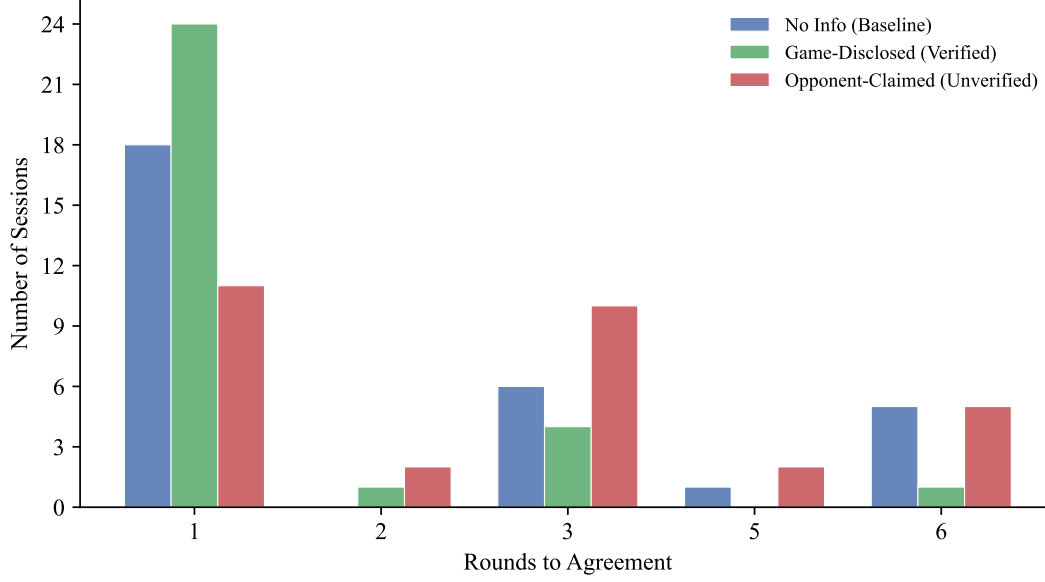
Figure 1: Distribution of rounds to agreement by information condition. Verified information produces rapid convergence (83% settle in Round 1), while unverified opponent claims generate prolonged negotiation (only 63% settle in Round 1, with sessions extending to Round 5–6).

(Bonferroni-corrected) showed:

- **Claimed vs. Verified:** $U = 656.5$, $p = .002$, $r_{rb} = -0.46$ (medium–large effect). Unverified claims produced significantly more rounds than verified disclosure.
- **None vs. Verified:** $U = 555.0$, $p = .170$ (n.s.). Baseline took marginally more rounds than verified.
- **Claimed vs. None:** $U = 511.5$, $p = .593$ (n.s.). Claimed and baseline were not significantly different.

The ordering was: Verified (1.5 rounds) < None (2.4) < Claimed (2.8). Verified information *accelerated* agreement; opponent claims *slowed* it.

Secondary Outcome: Deal Price

A one-way ANOVA found no significant effect of condition on deal price ($F(2, 87) = 2.19$, $p = .118$, $\eta^2 = .048$). The Claimed condition showed a marginally lower mean price ($68.63 vs. $69.47 for baseline), with Cohen's $d = -0.52$ for the pairwise comparison, but this did not survive Bonferroni correction ($p_{adj} = .149$).

Mechanism Check: Counteroffer Anchoring

Among sessions where the buyer counteroffered (n = 22 total), the mean first counteroffer was $33.20 (No Info), $40.48 (Verified), and $44.27 (Claimed). The ANOVA was not significant ($F(2, 19) = 1.85$, $p = .184$), but the effect size was medium–large ($\eta^2 = .163$) and the Claimed–None contrast had a large Cohen's $d = 0.98$. The small sample sizes (n = 6, 5, 11) left this test severely underpowered.

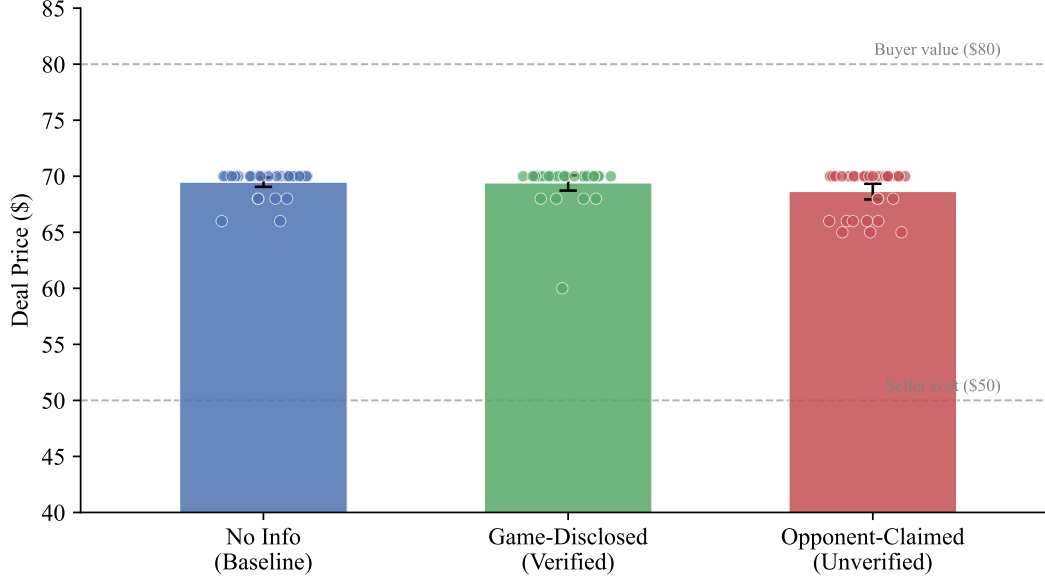The direction is noteworthy: *both* information conditions raised anchoring relative to baseline, and

3

Figure 2: Deal price by information condition, with individual session outcomes overlaid on condition means (±95% CI). Dashed lines mark the seller's cost ($50) and buyer's valuation ($80), defining the zone of possible agreement. Prices cluster near the AI's opening offer of $70, with a slight downward shift in the opponent-claimed condition.
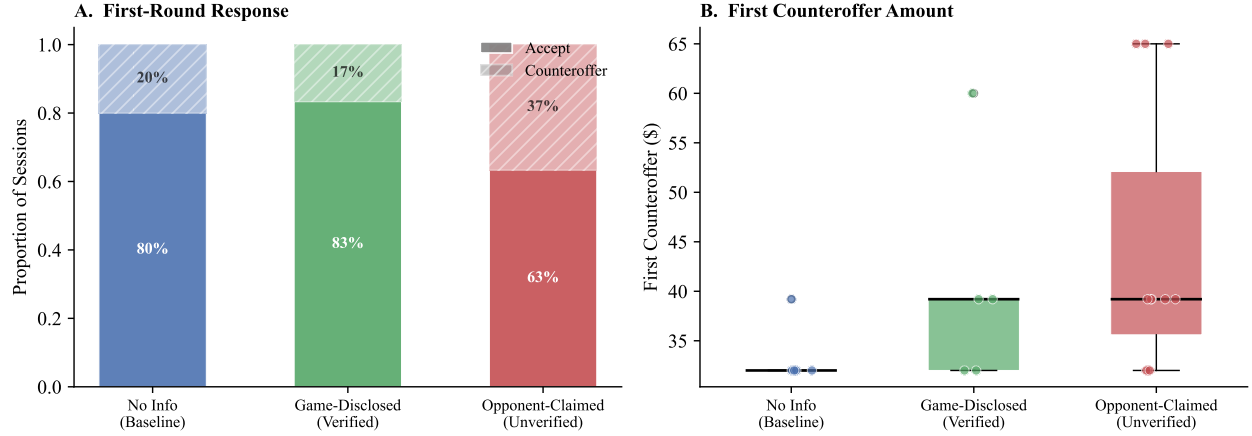

Figure 3: Bargaining dynamics. **Left:** Proportion of first-round accepts vs. counteroffers by condition. The opponent-claimed condition shows the lowest acceptance rate (63%), generating more active negotiation. **Right:** Among sessions with counteroffers, buyers anchored higher when they had cost information — particularly in the opponent-claimed condition ($44.27) vs. baseline ($33.20), consistent with information-driven anchoring. Sample sizes are small (n = 6, 5, 11).

the unverified claim raised it the most. This is consistent with information-driven anchoring but inconsistent with the hypothesis that verified information would produce *more* exploitation than unverified claims.

4

5 Evaluation

Did We Learn Something Interesting?

Partially — we found suggestive evidence that information source affects bargaining *process* **(how long deals take) but not bargaining** *outcomes* **(whether deals happen or at what price), and the process effects run opposite to what the hypothesis predicted.**

1. **The primary hypothesis was untestable.** The 100% deal rate across all conditions means we cannot evaluate the predicted deal-rate ranking. This is a design problem (ZOPA too generous), not a hypothesis problem.
2. **The one significant finding inverts the prediction.** We expected verified information to *slow* negotiation (via exploitative anchoring) and unverified claims to *speed* it (via reciprocal trust). Instead, verified information *accelerated* deals ($p = .002$ vs. claimed) while unverified claims *slowed* them. This suggests:
   - Verified info provides common ground that facilitates rapid convergence (a coordination effect, not an exploitation effect).
   - Unverified claims introduce uncertainty that *extends* negotiation — buyers may want to test the claim's credibility through offers.
3. **Counteroffer anchoring is suggestive but underpowered.** The large effect size ($d = 0.98$) for Claimed vs. None counteroffers hints that cost information (from any source) shifts buyer anchoring upward. But with only 22 counteroffering sessions, we cannot draw firm conclusions.
4. **The "source-channel reversal" was not observed.** Both information conditions shifted behavior in the same direction relative to baseline (higher counteroffers, slight price decrease). They differed in magnitude, not direction. The hypothesized reversal — verified disclosure hurts, unverified claims help — did not materialize.

Interestingness re-assessment: The finding that verified info speeds up deals while unverified claims slow them down is itself somewhat surprising (score: 3/5 on prediction surprise). It suggests a *coordination* vs. *deliberation* mechanism rather than the hypothesized exploitation vs. reciprocity mechanism. However, the evidence is limited by the design's inability to produce deal failures.

Limitations

1. **Ceiling effect on deal rate.** The generous ZOPA ($30) and accommodating acceptance thresholds made deals near-certain, eliminating variance on the primary outcome.
2. **Simulated human, not real.** The gpt-5-nano simulated buyer follows a scripted strategy with limited strategic sophistication. Real humans might exhibit stronger exploitation or trust responses.
3. **AI opponent uniformity.** The AI seller plays identically across conditions. In reality, a seller who voluntarily reveals cost information might also adjust their bargaining strategy.
4. **Small subsamples for mechanism tests.** Only 22/90 sessions produced counteroffers, severely limiting power for anchoring analysis.
5. **Single parameter point.** All sessions used the same buyer value ($80) and seller cost ($50). Effects might differ at other ZOPA sizes.

6 Next Experiment

Archetype: DEEPEN — Same hypothesis, redesigned game to produce variance on the primary outcome and increase power for mechanism tests.

Proposed Changes

1. **Shrink the ZOPA.** Set buyer value = \$65, seller cost = \$50 (ZOPA = \$15 instead of \$30). With a tighter surplus, the AI's opening offer of \$60 leaves the buyer only \$5 of surplus, making rejection and no-deal outcomes plausible.
2. **Raise the AI's opening anchor.** Set opening price at 90% of buyer value (e.g., \$58.50) to create more initial resistance.
3. **Stiffen the simulated human.** Lower the buyer's acceptance threshold to accept only offers ≤ \$57 (vs. current \$75), and raise the initial counteroffer anchor to 60% of buyer value. This creates a real risk of impasse.
4. **Add a "verified-exact" condition.** The current verified condition reveals an approximate range (\$45–\$55). A fourth condition revealing the exact cost (\$50) would test whether precision of verified information intensifies exploitation.
5. **Increase N to 60 per condition.** The design memo estimated 60–80 sessions needed per condition for 80% power to detect 15–20 percentage-point differences in deal rates.

Updated Hypothesis

With a tight ZOPA, information should matter more: knowing the seller's cost allows the buyer to extract nearly all surplus (offer \$51), while ignorance forces higher offers. The source-channel reversal becomes testable: does the buyer exploit verified info more aggressively than claimed info?

Concrete Next Step

Run:

```
/hypothesize Does information source (verified vs. claimed)
  differentially affect deal rates when the ZOPA is narrow
  ($15) and deal failure is plausible?
```

Then `/design-experiment` → `/create-2-player-game` → `/run-experiment` with the tighter parameterization.