Experiment Results:
The Information Precision Valley Under Tight Bargaining (v2)

February 12, 2026 · Status: Post-analysis · Iteration of v1

Research Question

Does tightening the zone of possible agreement (ZOPA) reveal an "information valley" in deal rates — where range information about opponent valuations *reduces* deal rates below both exact-information and no-information baselines?

The v1 experiment found 100% deal rates across all conditions under generous ZOPAs ($10–$25), but discovered a significant monotonic price gradient ($F(2, 27) = 7.99$, $p = 0.002$, $d = 1.54$). This v2 iteration shrinks the ZOPA to $2–$5 and reduces rounds from 6 to 3, hypothesizing that tight bargaining conditions would force impasses — particularly in the range-information condition, where the provided range ($35–$50 for seller cost) spans values where no deal is feasible.

The core mechanism is **feasibility uncertainty**: when the buyer is told the seller's cost could be as high as $50 but the buyer only values the item at $42, the range includes values where no deal is profitable, potentially causing pessimistic anchoring and demands outside the true ZOPA. Triviality score: 21/25 (hypothesis memo v2).

Experimental Design

Three between-subjects conditions, each with a within-subject ZOPA manipulation via `buyer_value` $\in \{42, 45\}$ (seller cost fixed at $40, giving ZOPA = $2 or $5). Maximum 3 rounds (AI offers Rounds 1 and 3, human offers Round 2). Both agents are earnings maximizers with hard guardrails only.

Condition	Game folder	Information structure
No Info	`bargain_tight_none`	Neither player knows opponent's valuation
Range Info	`bargain_tight_range`	Buyer told $35–$50; seller told $38–$55
Exact Info	`bargain_tight_exact`	Both know both valuations

All conditions share: seller_cost = $40, 3 rounds, alternating offers (AI first), price range $0–$100. The key design change from v1: ZOPA shrunk from $10–$25 to $2–$5, and rounds reduced from 6 to 3.

Data Overview

Sample: 30 simulations total — 10 per condition (5 at ZOPA = $2, 5 at ZOPA = $5). Factorial design with seed 1022.

Checker results: 27/30 nominal pass (90%), 30/30 substantive pass (100%). All 3 failures are the known checker false positive: deals at exactly `buyer_value` ($42) yield $0.00 human earnings, and the checker hallucinates an arithmetic error ("incorrectly $0.00 instead of $0.00"). No substantive game issues.

Condition	N	Deal Rate	Mean Price	Mean AI Open	Mean Rounds
No Info	10	100%	\$41.85	\$49.50	2.6
Range Info	10	100%	\$42.25	\$45.00	2.5
Exact Info	10	100%	\$41.80	\$41.75	1.2

Table 1: Summary statistics by condition. All 30 games reached deals.

Results

Primary Outcome: Deal Rate

The hypothesis predicted a valley: Range Info deal rate < No Info < Exact Info. **We observed 100% deal rates across all conditions and all ZOPA levels.** The ceiling effect from v1 persists despite shrinking the ZOPA to \$2 and reducing rounds to 3. There is zero variance in the primary outcome — no statistical test is possible.

All four directional predictions are **refuted**: the valley does not exist, not even under the tightest ZOPA (\$2). The feasibility uncertainty mechanism did not produce impasses.

Secondary Outcomes

While the primary hypothesis failed, the secondary outcomes reveal how information precision shapes the *process* of bargaining even when it does not affect the outcome.

AI Opening Offers (Figure 1, Panel A). Information precision dramatically affects anchoring behavior. One-way ANOVA: $F(2, 27) = 8.75$, $p = 0.001$, $\eta^2 = 0.39$. The AI seller opens at the fair price when it knows the buyer's exact value (\$41.75), at a moderate anchor when given a range (\$45.00 — exactly at the midpoint of the range \$38–\$55), and at an aggressive anchor when uninformed (\$49.50). Pairwise: exact vs. none $d = -1.50$ ($p = 0.008$); exact vs. range $d = -5.81$ ($p < 0.001$); range vs. none $d = 0.89$ ($p = 0.03$).

Rounds to Deal (Figure 1, Panel B). Information precision also affects convergence speed. Kruskal-Wallis $H = 16.1$, $p = 0.0003$, $\epsilon^2 = 0.52$. Exact-information pairs settle in a median of 1 round (the AI opens near fair price, the buyer accepts). No-info and range pairs require 2–3 rounds of negotiation. Exact vs. none and exact vs. range are highly significant ($p < 0.001$), but none vs. range do not differ.

Deal Price (Figure 1). Despite different opening offers, final deal prices do not significantly differ across conditions: $F(2, 27) = 1.04$, $p = 0.37$. All conditions converge near the fair price (\$41–\$42.50 depending on ZOPA). The negotiation process corrects for aggressive anchoring within 3 rounds.

Buyer Surplus Share. Marginally significant: $F(2, 27) = 3.36$, $p = 0.0497$. Buyers capture more surplus under exact info (47%) than range info (30%), with no info intermediate (45%). Range vs. exact: $d = 1.04$, $p = 0.04$. This suggests range information creates an asymmetry that sellers exploit — the buyer, uncertain whether a deal is even feasible, accepts seller-favorable prices.

ZOPA Interaction

ZOPA size does *not* interact with information condition for any outcome. Two-way ANOVA on deal price: condition × ZOPA interaction $p = 0.68$. For AI opening offer: condition is the sole

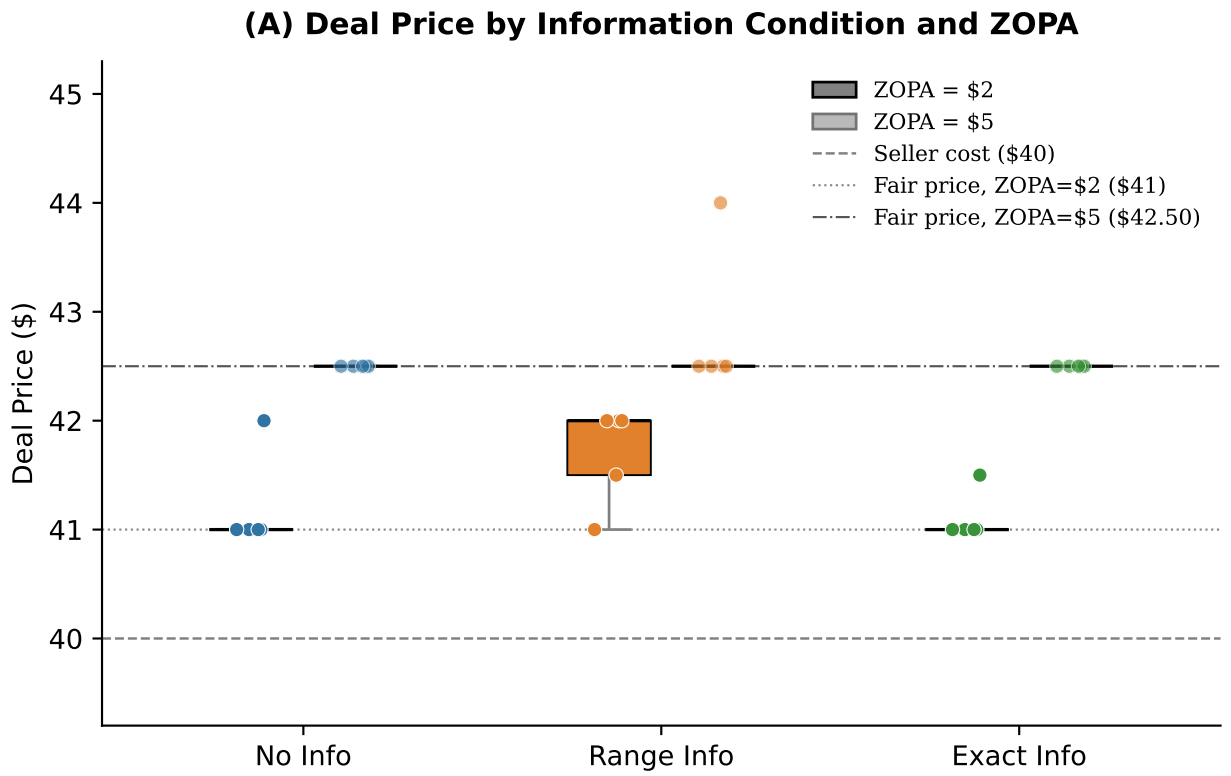

Figure 1: Deal price by information condition and ZOPA. Despite very different opening offers (AI opens at \$49.50 under no-info vs. \$41.75 under exact-info), final prices converge near the fair price in all conditions. Dashed lines show seller cost (\$40) and fair prices (\$41 and \$42.50).

significant predictor ($p = 0.001$); ZOPA has no effect ($p = 0.58$), nor does the interaction ($p = 0.95$). The AI anchors identically regardless of whether the ZOPA is \$2 or \$5.

This directly refutes the moderation prediction: the hypothesized "deeper valley under tight ZOPA" does not exist because there is no valley at all.

Evaluation

Did We Learn Something Interesting?

Partially — we learned about negotiation process but not about deal breakdown.

The primary hypothesis (information valley in deal rates) was comprehensively refuted: 100% deal rates across all conditions, including under the tightest ZOPA (\$2) with only 3 rounds. The LLM agents are too rational and too cooperative to reach impasse. The "feasibility uncertainty" mechanism did not produce the predicted behavior — even when the buyer was told the seller's cost could be as high as \$50 (above the buyer's own \$42 valuation), the simulated human still negotiated successfully.

However, the secondary findings are genuinely informative:

1. **Information anchoring is real and large** ($\eta^2 = 0.39$). The AI seller's opening offer is

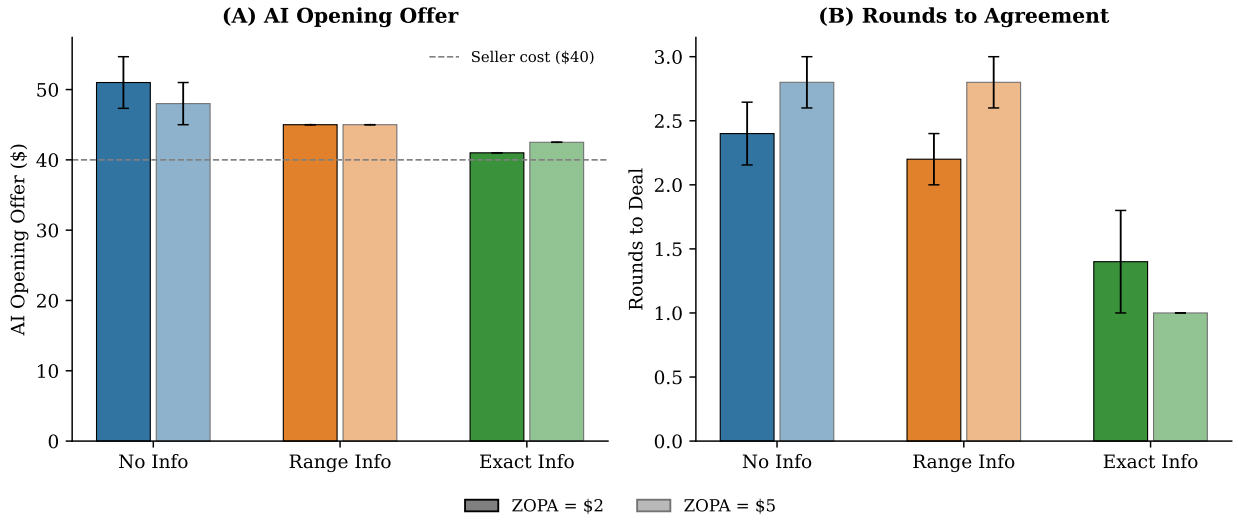

Figure 2: (A) Mean AI opening offer by condition. Information precision anchors the AI: exact info → fair price, range info → range midpoint, no info → aggressive anchor. (B) Mean rounds to deal. Exact information produces near-instant agreement; no-info and range-info pairs negotiate for 2–3 rounds.

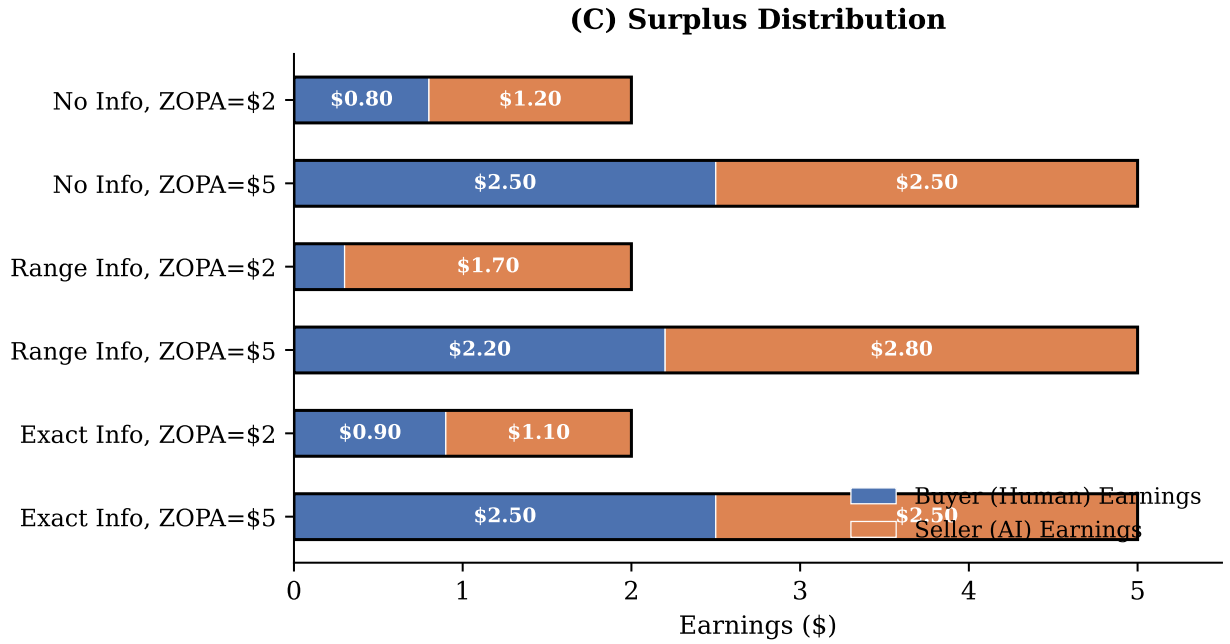

Figure 3: Surplus distribution by condition and ZOPA. In the range-info condition, sellers capture a larger share of surplus ($1.73/$2 under tight ZOPA) than in the exact-info condition ($1.10/$2). Efficiency is 100% in all conditions.

mechanically determined by its information: exact → fair price, range → range midpoint, none → aggressive anchor. This is the v2 analog of the v1 price gradient.

2. **Negotiation corrects anchoring within 3 rounds.** Despite opening offers that differ by

$8 across conditions, final prices converge. The bargaining process is remarkably efficient.

3. **Range info disadvantages buyers** (marginally significant). The buyer surplus share drops from 47% to 30% under range info, suggesting that feasibility uncertainty makes buyers more willing to accept seller-favorable terms even when deals are feasible.

Diagnosis: IMPLEMENTATION. The ceiling effect is not a design problem (the ZOPA *is* tight) but an implementation problem: gpt-5-nano agents reason about payoffs too well. They recognize that any positive earnings beat $0 and accept deals even when the surplus is minimal. The feasibility uncertainty mechanism depends on *behavioral* heuristics (anchoring, loss aversion, pessimism) that LLM agents do not exhibit. This is a fundamental limitation of LLM-as-subject experiments for testing behavioral hypotheses.

Limitations

- **LLM rationality ceiling:** The agents compute that any deal > $0 beats impasse, eliminating the behavioral channel the hypothesis depends on. Human subjects, who exhibit loss aversion and anchoring, might produce the predicted valley.
- **Small N:** With only 10 per condition (5 per cell), we have limited power for secondary analyses. The buyer surplus share finding ($p = 0.0497$) is borderline.
- **Fixed ranges:** The ranges ($35–$50 and $38–$55) were chosen to span infeasible values, but the AI may not process "infeasibility" the way a human would.
- **Two iterations, same ceiling:** This is the second experiment (after v1) to find 100% deal rates. The pattern is robust across ZOPAs from $2 to $25 and rounds from 3 to 6.

Next Experiment

Two iterations have established that LLM agents do not produce impasse in bilateral bargaining, regardless of ZOPA size, round limits, or information structure. The ceiling effect is a property of the agents, not the game design. Three paths forward:

Option A: PIVOT to a different outcome variable. Instead of testing whether information affects *deal rates* (which are always 100%), test whether information affects the *source of value* — e.g., shift from bargaining over price to a game where information determines whether players can coordinate at all (e.g., a coordination game, common-pool resource, or signaling game where miscoordination is the failure mode rather than impasse).

Option B: HUMAN-TEST the current design. The tight-ZOPA bargaining game is well-designed for human subjects. The $2 ZOPA with range info spanning infeasible values should produce genuine feasibility uncertainty in human participants. Run the v2 games with real humans via `interview chat`.

Option C: PIVOT to information *source* effects. The v1 results showed a price gradient driven by anchoring. Rather than manipulating precision (none/range/exact), manipulate the *source* of information: does the same numerical information carry different weight depending on whether it comes from a credible vs. non-credible source? This shifts the mechanism from behavioral heuristics (which LLMs lack) to information processing (which LLMs can model).

Recommended: Option C (PIVOT). The anchoring result (opening offers scale with information) is the most robust finding across both v1 and v2. A source-credibility manipulation directly

tests whether LLM agents *discount* information based on its provenance — a cognitive mechanism that LLMs may actually exhibit, unlike loss aversion or feasibility pessimism.

Next step: Run `/hypothesize` with a new direction: "Does information source credibility moderate anchoring effects in bargaining? When told opponent valuations come from an unreliable source, do agents anchor less aggressively?"

Prior Experiment Context (Machine-Readable)

```
prior_hypothesis: Under tight bargaining (ZOPA <=$5, <=3 rounds), range info
  reduces deal rates below exact and no-info baselines via feasibility uncertainty
verdict: NO_NULL
diagnosis: IMPLEMENTATION
key_finding: 100% deal rates across all conditions even with $2 ZOPA and 3 rounds;
  LLM agents too rational for impasse. Secondary: significant anchoring effect on
  AI opening offers (eta-sq=0.39) and marginal buyer surplus disadvantage under
  range info.
key_statistic: deal_rate = 100% (30/30); AI opening offer F(2,27)=8.75, p=.001
dag_variables: X=info_precision, M=feasibility_uncertainty, Y=deal_rate, Z=zopa_size
testable_implications_results: valley_in_deal_rates=REFUTED,
  no_info_resilience=CONFIRMED (100% deals), exact_info_efficiency=CONFIRMED
  (fastest deals), zopa_moderation=REFUTED (no interaction)
next_archetype: PIVOT
proposed_changes: Abandon deal-rate hypothesis for LLM agents. Pivot to
  information source credibility as a manipulation that LLMs can process.
  Alternatively, test the current design with human subjects.
next_hypothesis_sketch: Information source credibility moderates anchoring in
  bargaining. Valuations from unreliable sources produce weaker anchoring and
  more aggressive counteroffers than identical valuations from reliable sources.
```