# TERMS-Bench

## Diagnosing LLM Negotiation Agents Beyond Deal Rate

Erica Zhang[♠], Fangzhao Zhang[♠], Aneesh Pappu[♠], Batu El[♠], Jose Blanchet[♠], Susan Athey[♣,♥], Jiashuo Liu[§,†], James Zou[♠,†]

[♠] Stanford School of Engineering [♣] Stanford Department of Economics [♥] Stanford Graduate School of Business
[§] Independent [†] Equal Advising

**ABSTRACT**

Negotiation is a central mechanism of economic exchange, shaping markets, procurement, labor agreements, and resource allocation. It is also a canonical testbed for agentic language models, requiring multi-turn interaction under hidden preferences, strategic communication, and binding constraints. These properties make negotiation hard to evaluate: unlike math or code, it has no intrinsic verifier. Existing LLM negotiation evaluations rely on LLM-vs.-LLM interaction or aggregate outcomes such as deal rate, leaving failures opaque. We introduce TERMS-BENCH (**T**estbed for **E**conomic **R**easoning in **M**ulti-turn **S**trategy), a Bayesian-game framework that makes the environment itself the verifier by specifying the counterpart's latent type, policy, and payoff structure. We instantiate it in bilateral price negotiation, where the counterpart's private state and simulator policy are hidden from the agent but observable to the evaluator. This turns the counterpart from a black-box opponent into a diagnostic instrument, enabling agent-attributable failure analysis and oracle-reference optimality gaps. Evaluating 13 LLM agents spanning frontier systems from major providers, TERMS-BENCH turns negotiation evaluation from aggregate ranking into actionable diagnosis: *where* agents fail, *why* they fail, and *what* to strengthen. Empirically, frontier models saturate deal rate yet diverge in surplus extraction, cue use, belief calibration, and compliance, revealing agent-specific bargaining bottlenecks masked by prior benchmarks.

**Environment-as-Verifier**
Specify the counterpart's latent type, policy, and payoff structure to make the environment verifiable.

**Bayesian Bargaining**
A Bayesian-game framework for bilateral price negotiation under hidden information.

**Attributable Diagnostics**
Pinpoint where agents fail, why they fail, and what to strengthen with oracle benchmarks.

**Leaderboard**
https://terms-bench.github.io

**Code**
github.com/zou-group/terms-bench

**Correspondence**
{yz4232,jamesz}@stanford.edu

## 1 Introduction

Negotiation lies at the intersection of reasoning, communication, and social cognition, and serves as a canonical setting for agentic intelligence. It pervades commercial workflows such as procurement, contracting, pricing, logistics, where decisions are made under incomplete information, asymmetric incentives, and operational constraints [Raiffa, 1982, Bazerman and Neale, 1992]. Evaluation of LLM agents sits along a verifiability spectrum. At one end, math [Hendrycks et al., 2021] and code [Chen et al., 2021, Jimenez et al., 2024] have intrinsic verifiers that yield automatic correctness signals and enable reinforcement learning from verifiable rewards [Lambert et al., 2024, DeepSeek-AI, 2025]. At the other end, open-ended generation has no reference

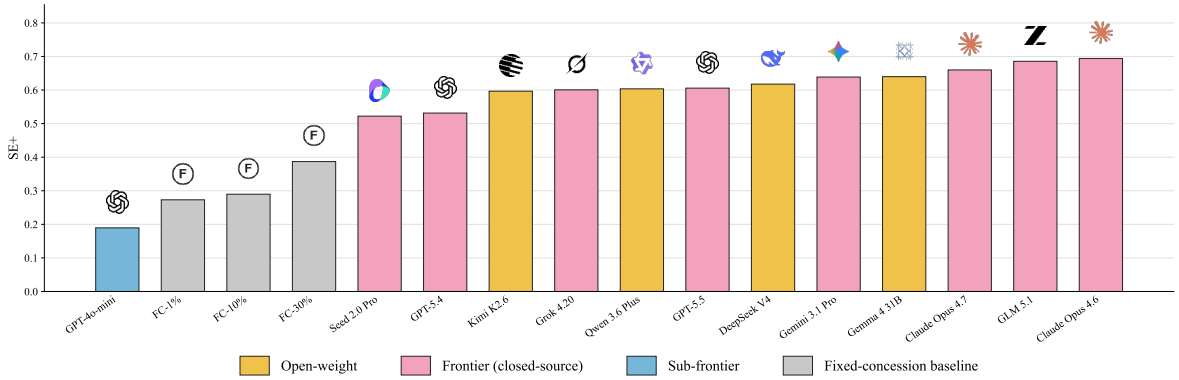



Figure 1: **Teaser terminal performance on the synthetic bilateral negotiation benchmark.** Surplus efficiency on feasible episodes ($SE_\pi^+$, normalized by ZOPA width) for 13 LLM agents and three fixed-concession baselines (FC-1%,10%,30%), each evaluated on the same 1,800 seeded episodes against the TERMS-BENCH simulator. Bars are colored by tier. $SE_\pi^+$ is one of six diagnostic metrics spanning terminal value, agreement calibration, opponent modeling, and protocol compliance reported in §4.2; product-grounded and other instantiations are deferred to §3,§4.

solution and relies on human or LLM-as-a-judge proxies [Zheng et al., 2023, Jia et al., 2025]. Negotiation occupies a *semi-verifiable* middle ground: terminal outcomes such as agreement are objective, but the multi-turn reasoning and language strategy that produce them are not. This makes negotiation both a stress test for agentic systems and a methodological challenge for evaluation: the verifier must be constructed, and how it is constructed determines what can be diagnosed.

The design of LLM negotiation agents has evolved from end-to-end neural systems [Lewis et al., 2017] to modular architectures that decouple strategic reasoning from language realization [He et al., 2018, Yarats and Lewis, 2018], strengthened through domain adaptation, structured reasoning, game-theoretic objectives, and behavioral conditioning [Chatterjee et al., 2024, Yao et al., 2023, Hwang et al., 2023, Hua et al., 2024, Cohen et al., 2025]. This progress in agent design has been accompanied by parallel advances in evaluation. Building on early dialogue corpora [Lewis et al., 2017, He et al., 2018, Chawla et al., 2021], recent LLM negotiation benchmarks have advanced along several fronts: multi-turn, multi-agent, and multi-product market simulation with private reservation values and welfare-based scoring [Bianchi et al., 2024, Liu et al., 2026]; multi-party stakeholder games with cooperative, greedy, and adversarial incentives that surface manipulation and coalition dynamics [Abdelnabi et al., 2024]; and human-preference-validated utility metrics that move beyond profit-only evaluation [Oh et al., 2026]. In parallel, game-theoretic benchmarks such as `GTBench` [Duan et al., 2024] and `Alympics` [Mao et al., 2024] probe abstract strategic competence in canonical games, largely via LLM-vs-LLM play. Despite their differences, these frameworks share a common evaluation paradigm: the evaluated agent negotiates against another LLM, and competence is summarized by terminal outcomes such as deal rates or surplus shares. This paradigm has two structural limits. First, because the counterpart is itself a black-box policy with unspecified latent preferences, outcomes confound agent competence with counterpart variability and preclude attribution to specific reasoning failures. Second, even multi-component or human-preference-validated metrics ultimately collapse performance into scalar rankings, obscuring where in the negotiation workflow an agent succeeds or fails.

Together, these limits point to a broader gap: existing evaluation measures *what* outcome is reached, but not *how* it arises from the agent's reasoning process. The LLM-vs.-LLM paradigm constructs verifiability by using another language model as a proxy judge [Zheng et al., 2023], preserving non-verifiability inside the counterpart rather than resolving it. For deployable agents operating under uncertainty and operational constraints, this opacity matters: failures that *would* be diagnosable under a transparent evaluator stay hidden. Empirical work shows that failures often arise within the negotiation process itself, through unstable reasoning, prompt sensitivity, and latent behavioral variability [Kwon et al., 2024, Schneider et al., 2024, Huang and Hadfi, 2024]. Meaningful evaluation therefore requires an evaluator-transparent counterpart: one against which outcomes can be attributed to specific agent behaviors and systematic failures surfaced before deployment.

We introduce TERMS-BENCH, a Bayesian-game framework for negotiation evaluation that constructs the environment itself as the verifier. This is a third approach to evaluating semi-verifiable agentic tasks: where LLM-as-judge methods [Zheng et al., 2023, Dubois et al., 2024] construct verifiability through a proxy model and outcome-rule benchmarks [Zhou et al., 2024, Liu et al., 2024] through hand-specified criteria, TERMS-

BENCH constructs it through the environment, so agent behavior can be attributed rather than merely scored. Benchmark instantiations are specified by varying the action space, observation modality, reward structure, and counterpart policy. In this paper, we instantiate TERMS-BENCH in bilateral price negotiation under incomplete information. By fully specifying the negotiation kernel, the benchmark makes the counterpart a diagnostic instrument: the evaluator observes the latent state and policy needed to explain agent success or failure, enabling simulator-defined reference policies and optimality gaps. Our contributions are:

- **A Bayesian-game framework with attributable, reference-based evaluation.** We introduce TERMS-BENCH, a framework that specifies the counterpart as a parameterized kernel with latent type and language-mediated cues, supporting benchmark instantiations along action-space, modality, and reward axes. We instantiate the framework in bilateral price negotiation, where a fixed counterpart policy enables agent-attributable failure analysis as well as simulator-defined oracle reference policies and optimality gaps.
- **Capability-isolating counterpart families.** We introduce six counterpart families that vary economic reactivity, cue reliability, noise, and pressure. Cross-family degradation isolates failures in cue use, latent-type inference, calibration under noisy evidence, and robustness to adversarial pressure.
- **Diagnostics and interventions beyond deal rate.** We evaluate feasible and infeasible episodes separately across four axes: terminal value, agreement calibration, opponent modeling, and protocol compliance. Oracle-posterior and revealed-type interventions further decompose performance gaps into inference, uncertainty, and control failures.

Together, these contributions turn negotiation evaluation into a diagnostic tool: not a leaderboard, but a controlled foundation for failure attribution that shows practitioners *where* agents break down and *what* to strengthen.

# 2 Negotiation as a Diagnostic Game

In this section, we formalize TERMS-BENCH, the Bayesian-game framework that makes the environment itself the verifier, and instantiate it in bilateral price negotiation under incomplete information. The framework and instantiation are grounded in economic bargaining and negotiation-analysis primitives; see Appendix A.

## 2.1 TERMS-BENCH: A Bayesian-Game Framework

We model negotiation as an *extensive-form, incomplete-information* game, following classical Bayesian bargaining formulations [Myerson, 1984, Ausubel et al., 2002]. TERMS-BENCH separates the *bargaining environment and protocol* from the *agent policy* that acts within it. This fixed environment enables diagnosis: performance differences can be attributed to agent policies rather than counterpart variability.

**Bargaining environment and protocol.** A bargaining environment is

$$\Gamma = (F, T_A, T_B, u_A, u_B, \mu), \tag{1}$$

where $F$ is the set of feasible outcomes (including disagreement $\perp$), $T_i$ is player $i$'s private type space, $u_i$ is the utility function, and $\mu \in \Delta(T_A \times T_B)$ is the prior over latent types.[1] Each type $t_i = (r_i, e_i, \kappa_i, \eta_i)$ encodes a reservation value $r_i$, an outside-option payoff $e_i$, urgency $\kappa_i$, and a behavioral parameter $\eta_i$ such as strategic stance. A protocol $\mathcal{P}$ is common knowledge and fixes (i) the move ordering, (ii) admissible action space $\mathcal{A} = \mathcal{D} \times \mathcal{L}$, (economic acts paired with natural-language realizations), (iii) information revelation rules, and (iv) termination conditions.

**Agent policy.** Given $(\Gamma, \mathcal{P})$, an agent's behavior is characterized by a policy $\pi : \mathcal{S} \to \mathcal{A}$ mapping information states to actions. At round $k$, the agent's information state $s_k = (h_k, x_A, b_k) \in \mathcal{S}$, contains the public interaction history $h_k$, private side information $x_A$ (e.g., the agent's reservation value and market context), and the agent's belief $b_k \in \Delta(T_B)$ over the counterpart's type.[2] The policy therefore captures both counterpart inference from observed signals and strategic action selection. Together with the fixed counterpart policy $\pi_B$ and prior $\mu$, $\pi$ induces terminal outcomes, allowing evaluation to compare policies under the same environment. We defer the full information-state notation, belief-update mechanics, and policy class to Appendix B.

TERMS-BENCH is intentionally general: instantiations may vary the action space, observation modality, reward structure, and counterpart policy. In this paper, we study a bilateral price-negotiation instantiation in which every component of $(\Gamma, \mathcal{P}, \pi_B)$ is available to the evaluator.

[1] We present the two-player form for clarity; the formulation extends naturally to multi-party and multi-product settings by replacing $(T_A, T_B)$ with $\{T_i\}_{i \in N}$ over a player set $N$, with $\mu \in \Delta(\prod_{i \in N} T_i)$.

[2] Throughout, subscript $A$ denotes the evaluated agent and $B$ denotes the counterpart.

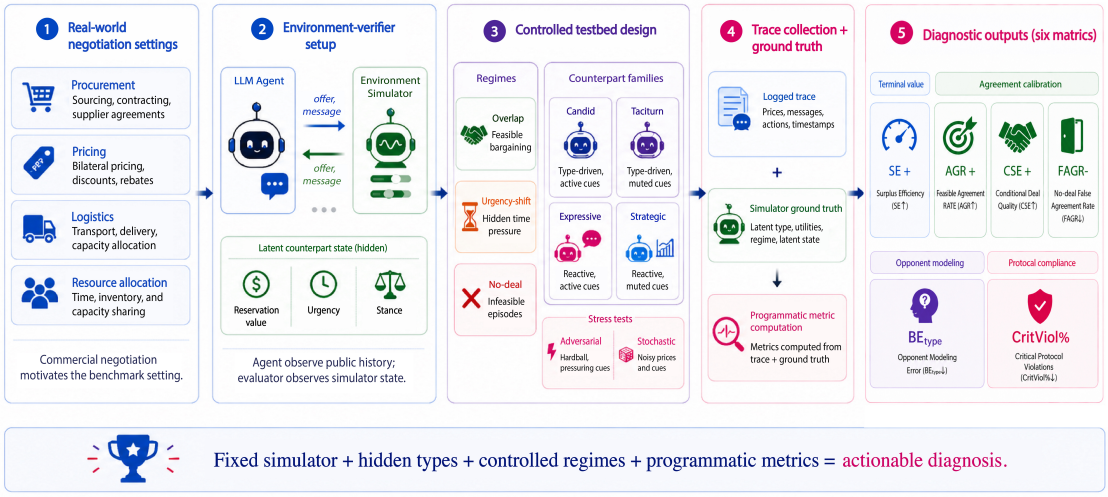



Figure 2: **Overview of TERMS-BENCH.** Commercial negotiation settings motivate an environment-verifier evaluation pipeline: an LLM agent negotiates with a fixed simulator whose latent type, policy, and payoff structure are hidden from the agent but observable to the evaluator. Controlled regimes test feasible bargaining, urgency shifts, and no-deal cases, while simulated counterpart families vary whether behavior is driven mainly by hidden private constraints or recent agent behavior, and whether language cues are informative or muted. Additional adversarial and stochastic families provide stress tests. The evaluator combines logged traces with simulator ground truth to compute six diagnostic metrics for surplus, agreement, no-deal recognition, opponent modeling, and protocol compliance (§2.3, §2.2.2).

## 2.2 Bilateral Price-Negotiation Instantiation

In this instantiation, an evaluated buyer or seller negotiates a single scalar price with an environment-simulated counterpart over up to $K$ alternating-offer rounds. The counterpart has a fully specified latent type that parameterizes its behavior and that the agent must infer from observed prices and messages. The agent, by contrast, is assigned only a reservation value $r_A$ as a hard individual-rationality constraint; its broader negotiation behavior is left to emerge through interaction.

### 2.2.1 Environment Specification

**Counterpart type and regimes.** For each episode, the counterpart draws a private type $t_B = (r_B, \kappa_B, \eta_B) \sim \mu$, where $r_B \in \mathbb{R}_+$ is its reservation value (a buyer counterpart will not pay above $r_B$; a seller counterpart will not accept below $r_B$), $\kappa_B \in [0, 1]$ is its urgency, and $\eta_B \in \{\texttt{conciliatory}, \texttt{neutral}, \texttt{aggressive}\}$ is its strategic stance.[3] We fix the outside option $e = 0$ for both parties. The type $t_B$ fully parameterizes the counterpart kernel $\pi_B$ (§2.2.2) and must be inferred by the agent from observed prices and messages.

To cover diverse bargaining scenarios, we define the environment prior as $\mu = \sum_{m=1}^{M} w_m \, \mu_m$, with $w_m \geq 0$ and $\sum_m w_m = 1$, where each component $\mu_m$ represents a structurally distinct regime. We use three canonical regimes targeting different competencies: (i) *Overlap*, feasible bargaining and surplus extraction; (ii) *Urgency-shift*, time-pressure adaptation; and (iii) *No-deal*, exit discipline under infeasibility. The regime draw specifies the joint reservation geometry and hence both $r_B$ and the agent's reservation value $r_A$. Full sampling distributions are given in §3.1.

**Protocol.** At round $k$, the agent selects $a_k = (d_k, p_k, l_k)$, where $d_k \in \{\texttt{Offer}, \texttt{Accept}, \texttt{Reject}\}$ is the economic decision, $p_k$ is the proposed price (when $d_k = \texttt{Offer}$), and $l_k$ is a natural-language message. An opener variable $\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}$ determines the first move. The agent observes the counterpart's price $p_k^B$ and message $l_k^B$, but not its latent type, nor the structured cues that parameterize $l_k^B$ (see §2.2.2). Offers must respect price bounds and monotonic concession. `Accept` is unavailable until a counterpart offer has been observed. Full constraints, termination cases, and observation-space structure are given in Appendix B.

[3] We use stance rather than personality as the primary behavioral driver; see Appendix B.4 for literature grounding [Ma, 2005, Huang and Hadfi, 2024, Cohen et al., 2025].

**Outcomes and utility.** The outcome space is $F = [p_{\min}, p_{\max}] \cup \{\perp\}$, with public price bounds and disagreement $\perp$. Buyer and seller utilities are $u_{\text{buyer}}(p) = r_{\text{buyer}} - p$ and $u_{\text{seller}}(p) = p - r_{\text{seller}}$, both zero on $\perp$. Agreement is individually rational iff $p \in [r_{\text{seller}}, r_{\text{buyer}}]$, yielding a non-empty zone of possible agreement (ZOPA) when $r_{\text{buyer}} \geq r_{\text{seller}}$.

### 2.2.2 The Counterpart as a Diagnostic Instrument

Given a counterpart family $\mathcal{F}$ and sampled private type $t_B = (r_B, \kappa_B, \eta_B)$, the counterpart is governed by a fixed stochastic kernel

$$(d_k^B, p_k^B, \tilde{s}_k, \tilde{c}_k) \sim \pi_B^{\mathcal{F}}(\cdot \mid t_B, h_{k-1}), \tag{2}$$

where $(\tilde{s}_k, \tilde{c}_k)$ are latent sentiment and strategic-posture cues that parameterize the counterpart's natural-language message but never alter the committed economic action. Fixing $\pi_B$ and $\mu$ across evaluated agents attributes performance differences to $\pi$. Full specification admits a model-based optimal reference policy, yielding a simulator-defined optimality gap (Appendix D).

**Capability-isolating counterpart families.** We instantiate the counterpart kernel $\pi_B$ from §2.2.2 as six parameter presets spanning two diagnostic axes (Figure 2, right): *economic reactivity*, how much economic responses depend on recent agent behavior rather than latent type, and *cue reliability*, whether language cues reflect latent stance. The $2 \times 2$ diagnostic core consists of CANDID, TACITURN, EXPRESSIVE, and STRATEGIC; two stress conditions add STOCHASTIC, with noisy prices and weak cues, and ADVERSARIAL, with hardball behavior, pressuring cues, and an aggressive-skewed stance prior. Family identity is hidden from the agent, so cross-family degradation isolates failures in cue use, latent-type inference, calibration under noise, and robustness to adversarial pressure. Full presets and kernel mechanics are in Appendix C (Table 3).

## 2.3 Objective, Diagnostics, and Attribution

The agent's normative objective is to maximize expected utility against $(\mu, \pi_B)$. Since disagreement has utility zero,

$$\mathbb{E}_{\mu,\pi,\pi_B}[u_A(f)] = \mathbb{P}(f \neq \perp) \cdot \mathbb{E}[u_A(f) \mid f \neq \perp], \tag{3}$$

decomposes performance into *deal attainment* and *value extraction conditional on agreement*. In feasible regimes both are desirable; in no-deal regimes agreement signals failed exit discipline. We therefore report agreement calibration separately for feasible and infeasible episodes.

We report four diagnostic axes (Table 1): (i) *terminal value*, surplus extracted; (ii) *agreement calibration*, agreeing when feasible and exiting when not; (iii) *opponent modeling*, latent-type inference; and (iv) *protocol compliance*, constraint adherence.

| Axis | Metric | Equation | What it captures | Dir. |
|---|---|---|---|---|
| Terminal value | $SE_\pi^+$ | $\frac{1}{\lvert\mathcal{I}^+\rvert}\sum_{i\in\mathcal{I}^+}\frac{u_A(f_i)}{\Delta_i}$ | Surplus extracted, normalized by ZOPA width | ↑ |
| Agreement calibration | $\text{AGR}_\pi^+$ | $\frac{1}{\lvert\mathcal{I}^+\rvert}\sum_{i\in\mathcal{I}^+}\mathbf{1}[f_i \neq \perp]$ | Closes deals when feasible | ↑ |
| | $CSE_\pi^+$ | $\frac{1}{\lvert\mathcal{I}^+_{\text{agr}}\rvert}\sum_{i\in\mathcal{I}^+_{\text{agr}}}\frac{u_A(f_i)}{\Delta_i}$ | Surplus given agreement | ↑ |
| | $\text{FAGR}_\pi^-$ | $\frac{1}{\lvert\mathcal{I}^-\rvert}\sum_{i\in\mathcal{I}^-}\mathbf{1}[f_i \neq \perp]$ | Agrees when infeasible | ↓ |
| Opponent modeling | $BE_{\text{type}}$ | $\frac{1}{3}(BE_r + BE_\kappa + \text{Brier}_\eta)$ | Reservation, urgency, stance inference error | ↓ |
| Protocol compliance | CritViol% | $\frac{1}{N}\sum_i \mathbf{1}[V_i^{\text{crit}} > 0]$ | Price-bound, IR, and invalid-action violations | ↓ |

Table 1: Headline diagnostic metrics, grouped by axis. $f_i$ is episode $i$'s terminal outcome; $\Delta_i := r_{\text{buyer}}^{(i)} - r_{\text{seller}}^{(i)}$; $\mathcal{I}^+ := \{i : \Delta_i > 0\}$, $\mathcal{I}^- := \{i : \Delta_i < 0\}$, and $\mathcal{I}^+_{\text{agr}} := \{i \in \mathcal{I}^+ : f_i \neq \perp\}$. Surplus efficiency decomposes as $SE_\pi^+ = \text{AGR}_\pi^+ \cdot CSE_\pi^+$, separating agreement rate from conditional value extraction. Empirically, we compute belief error from the agent's reported belief. Secondary diagnostics are deferred to Appendix F.

**Attribution via oracle interventions.** Outcome gaps do not reveal whether underperformance comes from poor latent-type inference, residual uncertainty, or failure to act on correct information. Since $(\Gamma, \mathcal{P}, \mu, \pi_B)$ is fully specified, we define a dynamic-programming reference policy $\pi^\star$ (Appendix D). This oracle is simulator-relative:

it is Bayes-optimal for the specified prior, counterpart kernel, protocol, horizon, and utility function, not a claim of human-optimal negotiation. We use it only as a benchmark-internal reference point.

We isolate bottlenecks by varying the agent's information while holding the episode and counterpart fixed. We compare *base*, which infers $t_B$ from prices and messages; *oracle-posterior*, which receives the exact Bayesian posterior $b_k$; and *revealed-type*, which receives the realized $t_B$. Letting $\bar{U}(\cdot)$ denote mean utility,

$$\bar{U}(\pi^\star) - \bar{U}(\pi^{\text{base}}) = \underbrace{\bar{U}(\pi^{\text{post}}) - \bar{U}(\pi^{\text{base}})}_{\Delta_{\text{inf}}\text{: inference}} + \underbrace{\bar{U}(\pi^{\text{reveal}}) - \bar{U}(\pi^{\text{post}})}_{\Delta_{\text{unc}}\text{: uncertainty}} + \underbrace{\bar{U}(\pi^\star) - \bar{U}(\pi^{\text{reveal}})}_{\Delta_{\text{ctrl}}\text{: control}}. \tag{4}$$

Thus $\Delta_{\text{inf}}$ measures the value of replacing the agent's own inference with the simulator posterior, $\Delta_{\text{unc}}$ the value of resolving remaining type uncertainty, and $\Delta_{\text{ctrl}}$ the residual gap to the simulator oracle after type information is no longer hidden. A negative $\Delta_{\text{inf}}$ means that the posterior intervention reduced realized utility relative to the base prompt, while a negative $\Delta_{\text{ctrl}}$ means that the full-reveal LLM condition exceeded the discretized oracle reference on those episodes. Full intervention details are in Appendix E.

# 3 Simulator Design

We factor the simulator into two fixed layers. First, the *economic regime generator* samples episode geometry and hidden state, including reservation values, feasibility, and urgency. Second, the *environment-simulated counterpart policy* governs behavior within the sampled scenario: openings, acceptance, walk-away, counter-offers, and language-facing cues. Because both layers are fixed across evaluated agents, performance differences can be cleanly attributed to the agent policy.

**Notation.** Let $R := p_{\max} - p_{\min}$ be the price range and define $\Delta := r_{\text{buyer}} - r_{\text{seller}}$. Then $\Delta > 0$ denotes a feasible episode with ZOPA width $\Delta$, while $\Delta < 0$ denotes a no-deal episode with infeasibility gap $-\Delta$. We use $z > 0$ for sampled feasible widths and $q > 0$ for sampled no-deal gaps.

## 3.1 Economic Regime Generator

The regime generator instantiates three canonical regimes by specifying reservation geometry and urgency. Counterpart urgency is drawn from a baseline law $D_\kappa = \text{Beta}(\alpha_\kappa, \beta_\kappa)$ rescaled to $[0, 1]$; counterpart stance $\eta_B$ is drawn from a family-specific prior (§3.2, Appendix C).

- **Overlap** *(surplus extraction; feasible-agreement calibration).* Sample $z \sim \mathcal{U}[\Delta_{\min}, \Delta_{\max}]$ and midpoint $m$ within price bounds; set $r_{\text{buyer}} = m + z/2$ and $r_{\text{seller}} = m - z/2$. Difficulty varies with $z$.
- **Urgency-shift** *(time-pressure adaptation).* Use Overlap reservations, but draw $\kappa_B \sim D_\kappa^{(s)}$ with mean shift $s := \mathbb{E}_{D_\kappa^{(s)}}[\kappa] - \mathbb{E}_{D_\kappa}[\kappa]$. Since urgency is hidden and payoff-irrelevant, this isolates adaptation to pressure.
- **No-deal** *(exit discipline under infeasibility).* Sample $q \sim \mathcal{U}[q_{\min}, q_{\max}]$ and midpoint $m$; set $r_{\text{buyer}} = m - q/2$ and $r_{\text{seller}} = m + q/2$. Smaller $q$ makes infeasibility harder to distinguish from a narrow ZOPA.

The generator also supports a *data-grounded* variant that swaps synthetic price geometry for empirical statistics from a real catalog; we describe it in §3.3 after the counterpart policy.

## 3.2 Environment-Simulated Counterpart Policy

Conditional on the sampled scenario and counterpart behavior family $\mathcal{F}$ (§2.2.2), the counterpart kernel $\pi_B$ comprises four components: (i) an opener-role protocol, (ii) an acceptance and walk-away response model, (iii) a counter-offer rule, and (iv) a cue-generation interface. Family-specific presets, history-feature definitions, and default hyperparameters are deferred to Appendix C.

**Opening role.** The opener $\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}$ is an episode-level protocol attribute balanced across regime–family cells (§4.1). If the counterpart opens, its first price is sampled from the randomized opening-offer model in Appendix C.4. If the agent opens, the counterpart first applies the response model below. Absent acceptance or walk-away, it samples its first price from the same opening model, since no prior counterpart offer exists.

**Acceptance and walk-away.** Given an agent offer $p_k^A$, define the role-normalized favorability

$$\bar{\Delta}_k := \begin{cases} (p_k^A - r_B)/R, & \text{counterpart is seller,} \\ (r_B - p_k^A)/R, & \text{counterpart is buyer,} \end{cases}$$

so that $\bar{\Delta}_k \geq 0$ iff the offer is individually rational for the counterpart. Using the concave deadline clock $\widetilde{D}_k := \sqrt{k/K}$ and $\widetilde{\bar{D}}_k := 1 - \widetilde{D}_k$, the acceptance probability is

$$a_k = \mathbf{1}\{\bar{\Delta}_k \geq 0\}\, \sigma\big(g_\theta(p_k^A, t_B, k, h_{k-1})\big)\,, \tag{5}$$

$$g_\theta = \alpha\bar{\Delta}_k + \beta\kappa_B - \gamma\widetilde{\bar{D}}_k + \rho_{\mathcal{F}}(\eta_B)\,\mathrm{ConcedeSpeed}_k + \xi_{\mathcal{F}}(\eta_B)\,\mathrm{Rigidity}_k. \tag{6}$$

Conditional on non-acceptance, the counterpart may walk away (recorded $d_k^B = \texttt{Reject}$) with hazard

$$\omega_k = \mathbf{1}\{k \geq k_{\mathrm{walk}}\}\,\mathbf{1}\{\bar{\Delta}_k < 0\}\,\sigma\big(\phi_0 + \phi_\Delta[-\bar{\Delta}_k]_+ + \phi_T\tau_k^W\big)\,, \tag{7}$$

where $\tau_k^W \in [0,1]$ is the walk-away clock active after a grace period. Walk-away is therefore enabled only after $k_{\mathrm{walk}}$ rounds and only when the current offer violates the counterpart's reservation constraint. The resulting response distribution is $(a_k,\ (1-a_k)\omega_k,\ \mathbf{1}\{k<K\}(1-a_k)(1-\omega_k))$ over $(\texttt{Accept}, \texttt{Reject}, \texttt{Offer})$, with remaining mass at $k = K$ resulting in round-limit disagreement.

**Counter-offer generation.** When $d_k^B = \texttt{Offer}$ and the counterpart has made a prior offer, it forms a type-dependent concession score

$$\begin{aligned}\tilde{\lambda}_B(h_{k-1}) = \lambda_0 + \lambda_1\kappa_B - \lambda_{2,\mathcal{F}}(\eta_B)\,\mathrm{ConcedeMagnitude}_k \\ - \lambda_3\mathbf{1}\{\eta_B = \texttt{aggressive}\} + \lambda_4\mathbf{1}\{\eta_B = \texttt{conciliatory}\},\end{aligned} \tag{8}$$

clipped to $\lambda_B = \min\{1, \max\{0, \tilde{\lambda}_B\}\}$. The noisy concession candidate $\tilde{p}_k^B = p_{k-1}^B - \lambda_B(p_{k-1}^B - r_B) + \varepsilon_k$, $\varepsilon_k \sim \mathcal{N}(0, \sigma_p^2)$, is projected onto the role-dependent monotone feasible interval $\mathcal{M}_B(k)$ to enforce individual rationality and monotonic concession:

$$p_k^B = \Pi_{\mathcal{M}_B(k)}(\tilde{p}_k^B), \quad \mathcal{M}_B(k) = \begin{cases}[r_B, p_{k-1}^B], & \text{seller},\\ [p_{k-1}^B, r_B], & \text{buyer}.\end{cases} \tag{9}$$

If no prior counterpart offer exists, $p_k^B$ is drawn from the opening-offer model.

**Cue generation and language realization.** After committing the economic action, the simulator samples latent cues $(\tilde{s}_k, \tilde{c}_k)$ from a family-specific cue model (Appendix C.5), with $\tilde{s}_k \in \{\texttt{positive}, \texttt{neutral}, \texttt{negative}\}$ and $\tilde{c}_k \in \{\texttt{Concede}, \texttt{Hold}, \texttt{Pressure}\}$. These cues condition the counterpart's message but never alter the committed action or price.

## 3.3 Data-Grounded Extension

TERMS-BENCH admits a *data-grounded* extension that lets practitioners evaluate agents on their own market data while preserving the diagnostic guarantees of the synthetic suite. Only the regime generator and the observable product context are replaced; the counterpart kernel, oracle policy, information-intervention decomposition, and metric definitions are unchanged, so any data-grounded instantiation reuses the same diagnostic axes as the synthetic main experiment and supports direct per-model comparison across instantiations.

**Interface.** A data-grounded instantiation is specified by three practitioner-supplied inputs. (i) A *product catalog* of items $j \in \mathcal{J}$, each with summary price statistics $(\hat{p}_{\mathrm{ref}}^{(j)}, \hat{p}_{\mathrm{lo}}^{(j)}, \hat{p}_{\mathrm{hi}}^{(j)})$ (reference price plus historical low and high), used to calibrate the public reference price and the latent reservation wedges around it. (ii) *Category-level price bounds* $[p_{\min}^{(c)}, p_{\max}^{(c)}]$ that set the public action range. (iii) *Observable product context* (item name, category, salient attributes, observable market range), which is appended to the agent's prompt; private reservations, counterpart urgency, and stance remain hidden. Given these inputs, the regime generator samples a category and product, applies the same overlap, urgency-shift, and no-deal geometries to the latent wedges, and passes episodes to the unchanged counterpart and evaluation pipeline. Thus, practitioners can plug in any catalog by supplying these statistics, without modifying the benchmark machinery.

**Instantiation in this paper.** For our experiments we instantiate the extension on *AmazonHistoryPrice* [Xia et al., 2024], a CamelCamelCamel-derived catalog of 831 products across 14 categories with per-product historical statistics. We use it as both an external-validity check on the synthetic findings and a difficulty stress test, since real catalogs typically present a much wider, long-tailed public action range that makes the product reference price a more decisive anchor than under the synthetic geometry. Per-model results, the geometric comparison to the synthetic suite, and the full construction are deferred to §4.3 and Appendix H.2.

| Agent | $SE_\pi^+ \uparrow$ | $\text{AGR}_\pi^+ \uparrow$ | $CSE_\pi^+ \uparrow$ | $\text{FAGR}_\pi^- \downarrow$ | $BE_{\text{type}} \downarrow$ | CritViol $\downarrow$ | %Oracle $\uparrow$ |
|---|---|---|---|---|---|---|---|
| Claude Opus 4.6 | **0.694**$_{\pm 0.014}$ | 99.3$_{\pm 0.5}$ | 0.699$_{\pm 0.013}$ | 0.00$_{\pm 0.00}$ | 0.222$_{\pm 0.013}$ | 0.00$_{\pm 0.00}$ | 76.3$_{\pm 1.5}$ |
| GLM-5.1 | 0.686$_{\pm 0.016}$ | 95.1$_{\pm 1.2}$ | **0.721**$_{\pm 0.014}$ | 0.00$_{\pm 0.00}$ | 0.218$_{\pm 0.008}$ | 1.33$_{\pm 0.53}$ | **76.8**$_{\pm 1.8}$ |
| Claude Opus 4.7 | 0.660$_{\pm 0.014}$ | 98.2$_{\pm 0.8}$ | 0.672$_{\pm 0.014}$ | 0.00$_{\pm 0.00}$ | 0.229$_{\pm 0.006}$ | 0.00$_{\pm 0.00}$ | 72.9$_{\pm 1.6}$ |
| Gemma-4-31B-IT | 0.640$_{\pm 0.014}$ | 99.8$_{\pm 0.2}$ | 0.641$_{\pm 0.014}$ | 0.00$_{\pm 0.00}$ | 0.260$_{\pm 0.009}$ | 0.06$_{\pm 0.11}$ | 69.9$_{\pm 1.5}$ |
| Gemini-3.1-Pro | 0.639$_{\pm 0.013}$ | 99.7$_{\pm 0.3}$ | 0.641$_{\pm 0.013}$ | 0.00$_{\pm 0.00}$ | 0.271$_{\pm 0.012}$ | 0.00$_{\pm 0.00}$ | 69.5$_{\pm 1.4}$ |
| DeepSeek-V4-Pro | 0.618$_{\pm 0.016}$ | 97.5$_{\pm 0.9}$ | 0.633$_{\pm 0.016}$ | 0.00$_{\pm 0.00}$ | 0.228$_{\pm 0.005}$ | 0.61$_{\pm 0.36}$ | 69.1$_{\pm 1.8}$ |
| GPT-5.5 | 0.606$_{\pm 0.014}$ | 99.2$_{\pm 0.5}$ | 0.611$_{\pm 0.014}$ | 0.00$_{\pm 0.00}$ | 0.228$_{\pm 0.010}$ | 0.00$_{\pm 0.00}$ | 66.6$_{\pm 1.6}$ |
| Qwen3.6-Plus | 0.604$_{\pm 0.016}$ | 98.2$_{\pm 0.7}$ | 0.614$_{\pm 0.015}$ | 0.17$_{\pm 0.33}$ | 0.237$_{\pm 0.006}$ | 2.06$_{\pm 0.66}$ | 67.7$_{\pm 1.8}$ |
| Grok 4.20 | 0.601$_{\pm 0.015}$ | 99.1$_{\pm 0.5}$ | 0.606$_{\pm 0.014}$ | 0.00$_{\pm 0.00}$ | **0.212**$_{\pm 0.006}$ | 1.50$_{\pm 0.56}$ | 65.9$_{\pm 1.6}$ |
| Kimi-K2.6 | 0.597$_{\pm 0.016}$ | 97.1$_{\pm 1.0}$ | 0.614$_{\pm 0.015}$ | 0.00$_{\pm 0.00}$ | 0.236$_{\pm 0.009}$ | 0.00$_{\pm 0.00}$ | 66.7$_{\pm 1.8}$ |
| GPT-5.4 | 0.531$_{\pm 0.016}$ | 99.4$_{\pm 0.4}$ | 0.535$_{\pm 0.016}$ | 0.00$_{\pm 0.00}$ | 0.242$_{\pm 0.009}$ | 0.00$_{\pm 0.00}$ | 59.4$_{\pm 1.8}$ |
| Doubao-Seed-2.0-Pro | 0.522$_{\pm 0.014}$ | **99.9**$_{\pm 0.2}$ | 0.523$_{\pm 0.014}$ | 0.00$_{\pm 0.00}$ | 0.247$_{\pm 0.009}$ | 0.06$_{\pm 0.11}$ | 56.6$_{\pm 1.5}$ |
| GPT-4o-mini | 0.189$_{\pm 0.013}$ | 52.2$_{\pm 2.8}$ | 0.363$_{\pm 0.016}$ | 0.00$_{\pm 0.00}$ | 0.251$_{\pm 0.005}$ | 0.00$_{\pm 0.00}$ | 22.2$_{\pm 1.6}$ |
| Fixed 30% | 0.387$_{\pm 0.015}$ | **99.9**$_{\pm 0.2}$ | 0.387$_{\pm 0.015}$ | 0.00$_{\pm 0.00}$ | – | 0.00$_{\pm 0.00}$ | 42.7$_{\pm 1.6}$ |
| Fixed 10% | 0.290$_{\pm 0.013}$ | 94.5$_{\pm 1.3}$ | 0.307$_{\pm 0.013}$ | 0.00$_{\pm 0.00}$ | – | 0.00$_{\pm 0.00}$ | 33.3$_{\pm 1.5}$ |
| Fixed 1% | 0.273$_{\pm 0.012}$ | 92.2$_{\pm 1.5}$ | 0.296$_{\pm 0.013}$ | 0.00$_{\pm 0.00}$ | – | 0.00$_{\pm 0.00}$ | 31.3$_{\pm 1.4}$ |

Table 2: Aggregate agent performance in the bilateral instantiation of TERMS-BENCH with 95% CIs. Green: top-two per metric; red: violations/false agreements (darker = worse). Arrows indicate preferred direction. The roster spans frontier systems from major providers at evaluation time, plus `GPT-4o-mini` as a sub-frontier reference.

# 4 Experiments

## 4.1 Experimental Setup

**Benchmark suite.** We evaluate agents across three regimes (§3.1) and six counterpart behavior families (§2.2.2). For each regime–family pair, we block over agent role {`Buyer`, `Seller`} and opener assignment {`AgentOpens`, `CounterpartOpens`}, with 25 episodes per role–opener cell, yielding 100 episodes per regime–family[4] pair and 1,800 episodes per agent. Across agents, we reuse the same seeded specifications, fixing reservation geometry, latent counterpart type, urgency draws, opening harshness, opener assignment, and simulator random streams.

**Agents and inference.** Agents negotiate against a fixed environment-simulated counterpart: a parameterized stochastic kernel with a language-realization layer (§3.2). We compare three fixed-concession baselines (conceding 1%, 10%, and 30% of the remaining distance to reservation per `Offer`) to proprietary and open-weight LLM agents (Table 9) under a shared wrapper. LLMs are called via `OpenRouter` and use identical role-conditioned prompts, a structured per-round `JSON` interface, deterministic decoding (temperature $= 0$), and one rollout per seeded episode. When supported, reasoning effort is set to `xhigh` to give each model its strongest available configuration. The counterpart voice model is fixed to `GPT-5.2` and only renders committed kernel actions in language; it does not affect prices, acceptances, walk-aways, or outcomes.

**Reported metrics.** Agents output a message, economic action, and belief estimate over the counterpart's latent type. We report six primary metrics across four diagnostic axes (§2.3), and additionally %Oracle: surplus efficiency relative to a dynamic-programming oracle (Appendix D). While perfectly correlated with $CSE_\pi^+$, it makes the oracle gap explicit. Details are deferred to Appendix H.

**Evaluation surfaces.** Beyond the synthetic main suite, we evaluate two extensions that reuse the counterpart kernel, oracle, and metrics verbatim. The *data-grounded variant* (§3.3, §4.3) replaces synthetic price geometry and abstract product references with Amazon-catalog statistics (831 products, 14 categories) and exposes product context to the agent. The *commercial extension* (§4.4) wraps each episode in unit economics (COMMERCE MODE) and chains episodes into multi-period sessions with cash and belief carryover (BANKROLL MODE); construction details are in Appendix J. Both extensions reuse $SE_\pi^+$, $\text{AGR}_\pi^+$, and the latent-type belief diagnostics, and add surface-specific metrics (regret rate for commerce; terminal balance, survival rate, and an optional memory premium for bankroll).

[4]We focus on the counterpart-more-urgent direction because it directly probes exploitation of counterpart time pressure; Appendix H.4 reports the reverse direction as a directionality check.

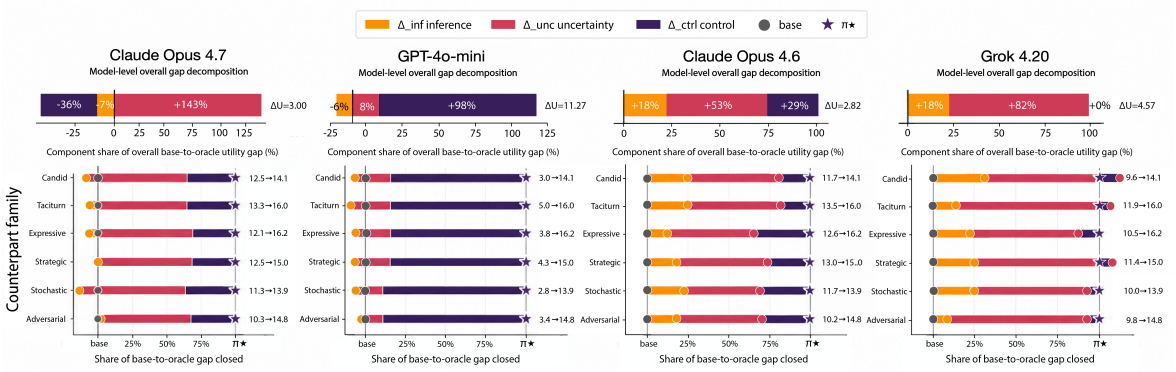



Figure 3: **Oracle gap decomposition.** Rows are normalized to each family's base-to-oracle utility gap, so the signed inference, uncertainty, and control components sum to 100%. Negative $\Delta_{\text{inf}}$: posterior injection hurt utility. Negative $\Delta_{\text{ctrl}}$: full-reveal LLM beat the discretized oracle (see §2.3 for details).

## 4.2 Synthetic Main Experiment

Table 2 reports aggregate performance for 13 LLM agents (Table 9) and 3 fixed baselines. Per-family and per-regime breakdowns appear in Table 16; full metrics in Tables 17–19. Surplus efficiency varies 3.7-fold across LLMs, from `GPT-4o-mini` to `Claude Opus 4.6`; the simplest fixed-concession baseline outperforms `GPT-4o-mini`, so LLMs do not uniformly clear hand-coded heuristics. We unpack these gaps in five findings: aggregate structure (F1), cue and information failures (F2–F3), per-agent bottleneck attribution (F4), and stable behavioral fingerprints (F5).

**Finding 1: High deal rate hides what matters.** Feasible agreement is nearly saturated: excluding `GPT-4o-mini`, all evaluated LLMs achieve $\text{AGR}^+_\pi \in [93.4\%, 99.9\%]$ (Table 2). Yet agreement does not imply value. Among agents with $\text{AGR}^+_\pi \geq 97\%$, $SE^+_\pi$ ranges from 0.522 (`Doubao-Seed-2.0-Pro`) to 0.694 (`Claude Opus 4.6`); `Doubao` attains the highest agreement rate (99.9%) but much lower conditional surplus than `Claude` ($CSE^+_\pi = 0.523$ vs. 0.699). Other axes reveal orthogonal failures: `GLM-5.1` has the best conditional surplus and oracle attainment but critical violations, while `Grok 4.20` has the lowest type-belief error (0.212) but only mid-tier surplus. Thus our metrics capture distinct capabilities that deal-rate alone would mask.

**Finding 2: Decoupling strategy from voice reveals a cue-use failure.** The counterpart families decouple language-facing cues from the economic kernel: CANDID/EXPRESSIVE expose informative cues; paired TACITURN/STRATEGIC mute them. We define the cue penalty $\alpha_{\text{cue}} = \overline{SE^+_\pi}(\text{cue-revealing}) - \overline{SE^+_\pi}(\text{cue-muted})$. For every LLM, $\alpha_{\text{cue}} < 0$ (Fig. 4): agents extract *less* surplus when informative cues are present. Penalties range from $-0.009$ (`GPT-5.4`) to $-0.063$ (`Claude Opus 4.6`, `Grok 4.20`), with 9 of 13 per-model 95% bootstrap CIs excluding zero and an across-model Wilcoxon test at $p < 10^{-3}$ (Appendix H.3.3). Stress families sharpen the picture: STOCHASTIC remains competitive despite corrupted prices and cues, while ADVERSARIAL is the lowest-surplus condition for most agents. The failure is not environmental noise but cue sensitivity used against the agent: warm cues induce over-concession, pressure cues trigger brittle behavior.

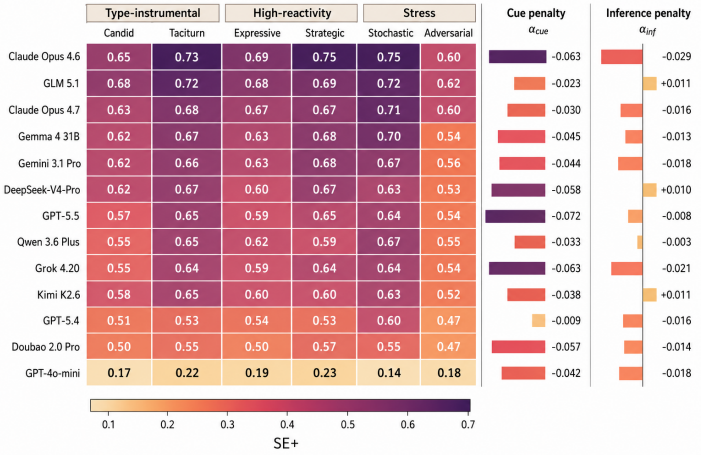



Figure 4: Per-family surplus efficiency, cue and inference penalty.

**Finding 3: Information does not reliably translate into surplus.** F2 shows that frontier LLMs lose value from counterpart cues. An information–action gap further appears across three diagnostics. First, posterior injection has weak and inconsistent returns: the *oracle-posterior* intervention supplies a Bayesian

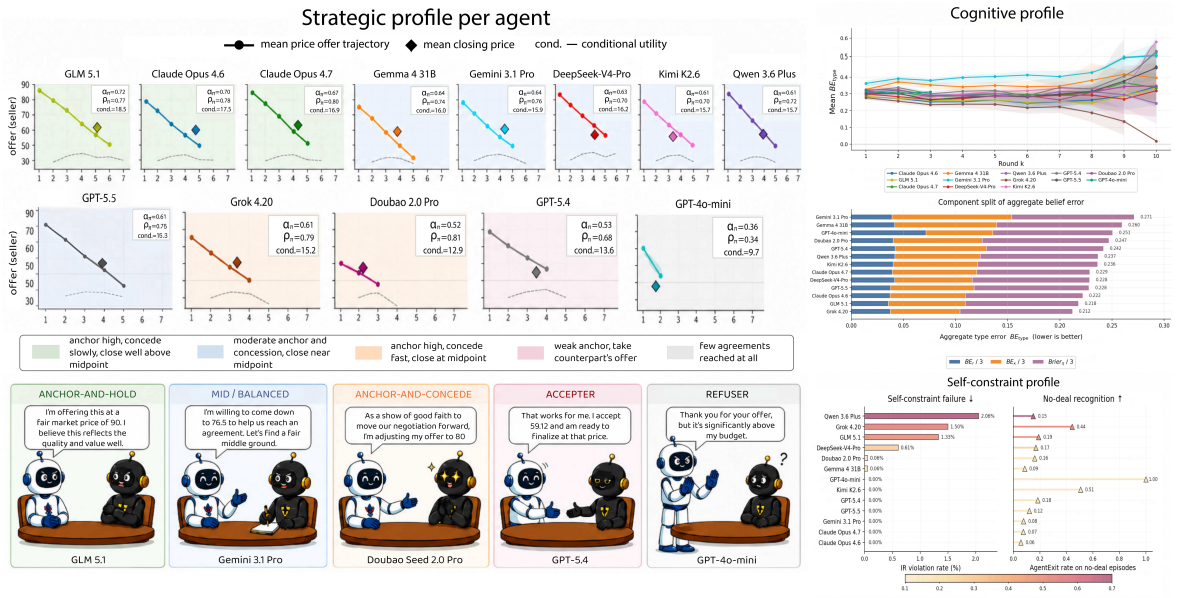



Figure 5: Behavioral profiles across three axes. **Strategic profile** (left) shows mean offer price trajectories in the (seller-opens, overlap) regime, where lines are clipped at mean closing round per-agent; diamonds mark mean closing price, and panel annotations report trajectory coefficient $\alpha_\pi$, closer rate $\rho_\pi$, and conditional utility (`cond.`). Background tints define five bargaining typologies: *anchor-and-hold*, *mid/balanced*, *anchor-and-concede*, *accepter*, and *refuser*. **Cognitive profile** (top right) reports mean aggregate belief error by round and its decomposition into reservation, urgency, and stance components. **Self-constraint profile** (bottom right) shows IR-violation rate (left) and $\mathrm{AgentExit}^-$ (§F; the share of no-deal episodes where the agent explicitly identifies infeasibility and rejects) on no-deal episodes (right). Lower IR-violation rate and higher $\mathrm{AgentExit}^-$ indicate cleaner no-deal recognition.

posterior $b_k \in \Delta(T_B)$ each round (§2.3, §E), but in the decomposition it closes only $1\%$ of the base-to-oracle gap for `Claude Opus 4.7` and is negative for `GPT-4o-mini` ($-6\%$; Fig.3). Second, beliefs do not improve online: for every frontier LLM, joint belief error is flat or *increases* over 10 rounds (Fig.5, top-right), with $\mathrm{Brier}_\eta$ dominating $BE_r$ and $BE_\kappa$. Third, family-level contrasts show the cost. Define $\alpha_{\mathrm{inf}} = \overline{SE^+_\pi}(\text{type-instrumental core}) - \overline{SE^+_\pi}(\text{high-reactivity core})$, where type-instrumental families (CANDID, TACITURN) reward latent-type inference more than high-reactivity families (EXPRESSIVE, STRATEGIC). A positive value indicates benefit from latent-type structure; instead, $\alpha_{\mathrm{inf}} < 0$ for 10 of 13 LLMs (Fig.4), with largest penalties for `Claude Opus 4.6` ($-0.029$) and `Grok 4.20` ($-0.021$). The inference penalty is rank-Wilcoxon-significant at the population level ($p = 0.007$) although the cruder sign test is only marginal ($p = 0.073$) and no individual CI excludes zero (Appendix H.3.3). Thus agents fail to convert both surface cues and payoff-relevant latent structure into better strategic action.

**Finding 4: Oracle interventions identify different agents fail for different reasons.** The oracle decomposition (Eq. 4) separates surplus gaps into inference, uncertainty, and control bottlenecks. `GPT-4o-mini` is control-limited: control accounts for $98\%$ of its much larger gap, while posterior and type revelation recover little surplus. `Claude Opus 4.7` and `Grok 4.20` are uncertainty-limited ($130\%$ and $82\%$), with small inference gains ($1\%$, $18\%$) and negligible residual control loss. `Claude Opus 4.6` is mixed across inference ($18\%$), uncertainty ($53\%$), and control ($29\%$). Oracle interventions thus point to distinct bottlenecks: smaller agents struggle to act on the revealed type, while stronger agents are bottlenecked by decisions under unresolved uncertainty.

**Finding 5: Stable bargaining fingerprints decouple concession from self-constraint.** Trace-level profiles decompose behavior into *concession dynamics* in feasible deals and *boundary monitoring* under infeasibility. In overlap episodes, trajectory shape, agent-closer rate $\rho_\pi$, and conditional utility separate five styles: anchor-and-hold, mid/balanced, anchor-and-concede, accepter, and refuser(Fig.5, left). In no-deal episodes, IR-violation rate and $\mathrm{AgentExit}^-$ form two independent axes of self-constraint, with agents spread along both (Fig.5, bottom-right) These axes are independent: `GLM-5.1` anchors strongly in overlap yet breaches reservation in no-deal, while `Doubao-Seed-2.0-Pro` compromises quickly in overlap yet holds cleanly under infeasibility. Within-agent variation across counterpart families is much smaller than between-agent variation and no agent crosses a typology boundary (Appendix H.5). Each model thus has a stable bargaining fingerprint, but strategic style and self-constraint are distinct capabilities that terminal metrics collapse.

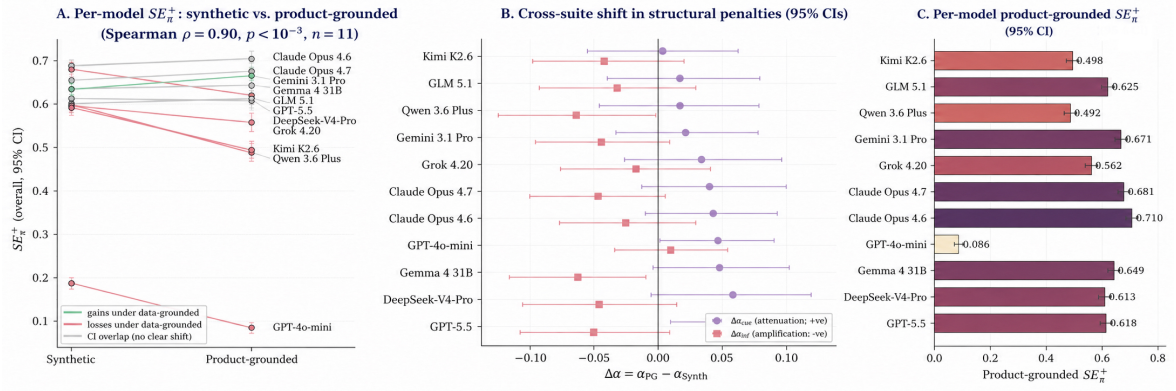



Figure 6: **Data-grounded variant summary** (eleven paired models). **A.** Per-model $SE_\pi^+$ slopegraph from the synthetic suite (left) to the data-grounded suite (right) with 95% bootstrap CIs. Lines are colored green where the data-grounded CI lies entirely above the synthetic CI (gains under data-grounded), pink where it lies entirely below (losses under data-grounded), and grey otherwise; rank order is largely preserved. **B.** Cross-suite shift in the structural penalties $\Delta\alpha = \alpha_{\mathrm{PG}} - \alpha_{\mathrm{Synth}}$ with 95% bootstrap CIs ($B$=2000), sorted by $\Delta\alpha_{\mathrm{cue}}$. All eleven models attenuate the cue-use penalty ($\Delta\alpha_{\mathrm{cue}} > 0$, paired Wilcoxon $p = 0.001$); ten of the eleven also amplify the inference penalty ($\Delta\alpha_{\mathrm{inf}} < 0$, $p = 0.002$), with `GPT-4o-mini` the lone exception. **C.** Per-model product-grounded $SE_\pi^+$ with 95% bootstrap CIs.

## 4.3 Data-Grounded Variant

We sweep a representative subset of the full model roster on the data-grounded instantiation (§H.2), holding the counterpart kernel, oracle, and metrics fixed and replacing only the price geometry and observable product context with Amazon-catalog statistics (100 episodes per regime, identical seeds across models). Geometry is non-trivially shifted: the public range becomes long-tailed and the relative ZOPA collapses by roughly an order of magnitude (§H.2), so agents must anchor on the public product reference rather than search uniformly over $[p_{\min}, p_{\max}]$. Figure 6 summarises the per-model cross-suite shift in $SE_\pi^+$ and the two structural penalties from F2–F3; full per-model tables, geometry comparisons, and bootstrap CIs are in Appendix H.2.3.

**Finding 6: Diagnostic ordering and bottlenecks survive a market-realistic geometry.** Across eleven paired models, the rank correlation between synthetic and data-grounded $SE_\pi^+$ is Spearman $\rho$=0.90 ($p < 10^{-3}$; Fig. 6A,C): `Claude Opus 4.6` remains on top and the bottom group is preserved, but the cross-suite shift is bimodal across the leaderboard, with upper-half models tending to gain and lower-half models tending to lose. Strong models exploit the public product anchor while weaker models flounder in the wider absolute price range, so the data-grounded suite acts as a difficulty multiplier that *amplifies* capability differences.

The two structural penalties from F2–F3 also replicate (Fig. 6B): $\alpha_{\mathrm{cue}}$ attenuates for all eleven models (median paired shift $+0.040$, paired Wilcoxon $p = 0.001$) as the product anchor partly insulates agents from cue over-reaction, while $\alpha_{\mathrm{inf}}$ amplifies for ten of eleven ($-0.045$, $p = 0.002$) as wider, skewed action ranges make latent-type inference more decisive; only `GPT-4o-mini` bucks the latter shift. No-deal recognition is also cleaner: $\mathrm{FAGR}_\pi^-$=0.000 for all eleven models. The agent-specific bottlenecks are therefore properties of the agents, not of the synthetic distribution; full per-model numbers, the geometry comparison, and the underlying Wilcoxon tables are in Appendix H.2.3.

## 4.4 Commercial Extension: Commerce & Bankroll

The negotiation-only diagnostics describe *how* an agent negotiates but not what those choices cost in dollars. The commercial extension wraps each episode in unit economics ($v$ resale value, $c$ fulfillment cost, $h$ fixed overhead, lot size $n \in \{1, \ldots, 50\}$, regime-conditioned outside option $o$). The merchant role earns $\Pi(p) = n(v - p - c) - h$ on agreement and $o$ on walk-away; the vendor role is symmetric. The headline single-episode metric is *regret rate*, the scale-invariant gap to the per-scenario best feasible profit summed over feasible scenarios: $\mathrm{Regret} = 1 - \sum_i \Pi_i / \sum_i \Pi_i^*$.

*Bankroll mode* chains $T$ commerce episodes into a single session, threading a cash ledger and per-supplier beliefs across periods. The cash balance evolves as

$$C_t = C_{t-1} + \Pi_t - (b + r \cdot R_t), \qquad C_0 = \text{starting capital}, \tag{10}$$

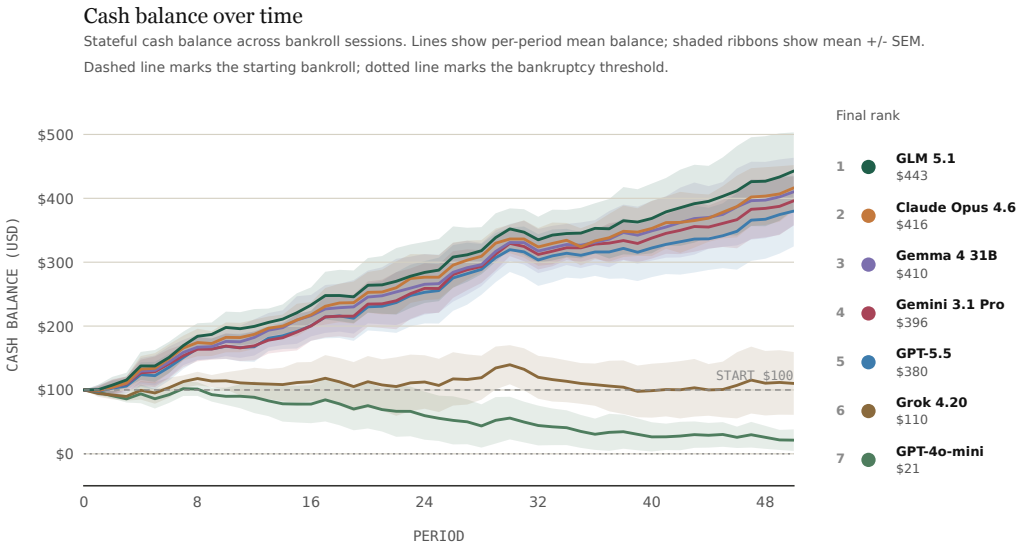



Figure 7: **Cash-balance trajectories under the bankroll chain.** Per-period mean cash balance for seven LLM merchants across stateful sessions; shaded ribbons show $\pm 1$ SEM and the right-edge ladder ranks agents by terminal balance. The dashed line marks the starting bankroll and the dotted line marks the bankruptcy threshold. Five LLMs compound to \$380–\$443 with full survival; `Grok 4.20` reaches \$110 (75% survival) and `GPT-4o-mini` ends near \$21 (50% survival). Full calibration in Appendix J.

where $\Pi_t$ is period $t$'s commerce profit and $(b, r)$ are fixed per-period and per-round operating costs over the $R_t$ rounds actually played. The session terminates at *hard ruin* the first time $C_t < \tau$ (default $\tau = 0$); remaining periods produce zero profit. The headline session-level metrics are *terminal balance* $C_T$ and *survival rate* $\Pr[T_{\text{ruin}} > T]$. Supplier-mode variants, the optional memory-premium diagnostic, and full defaults are in Appendix J.

**Finding 7: Diagnostic gaps translate into dollar-scale regret.** On a 192-scenario commerce sweep (merchant role, default regime mixture; Appendix J), the synthetic $SE_\pi^+$ ordering carries over but its magnitude becomes operational. `Claude Opus 4.6` earns \$68,592 against \$48,776 for the strongest fixed-concession baseline, a +\$19,816 swing on identical scenarios at the same walk-away rate, and leaves only 29% of feasible profit on the table, against 50% for the baseline and 80% for `GPT-4o-mini`. `GPT-4o-mini` is simultaneously over-cautious (66% walk-away) and miscalibrated (32.6% negative-profit closes); regret rate exposes this where average margin cannot, since margin is comparable across the bottom tier ($\approx 26\%$) while regret rates diverge by tens of points. Profit-summed-across-all- feasible-scenarios is therefore strictly more diagnostic than averaging over closed deals.

**Finding 8: Bankroll surfaces solvency failures invisible to single-episode metrics.** Chaining commerce episodes into a stateful session with per-period operating drag amplifies the diagnostic separation (Fig. 7). The five strongest LLMs (`GLM 5.1`, `Claude Opus 4.6`, `Gemma 4 31B`, `Gemini 3.1 Pro`, `GPT-5.5`) survive every session and compound to \$380–\$443 in terminal cash; `Grok 4.20` reaches only \$110 with one of four sessions ruining; `GPT-4o-mini` ends near \$21 with half its sessions bankrupt before the horizon. The terminal-balance spread between the strongest and weakest LLM is $\sim 21\times$, sharper than the $\sim 14\times$ spread in synthetic $SE_\pi^+$, because small per-period concession or walk-away errors compound across the chain into solvency failures that single-episode metrics cannot see. Calibration, per-model numbers, and reproduction scripts are in Appendix J.

**Additional Results.** Appendix experiments further stress-test the main findings. Difficulty stratification (Appendix G), prompt ablations (Appendix I.3), and extreme-reveal ablations (Appendix I.2) show that high agreement is not saturation: frontier agents close only a limited fraction of the oracle gap even when hidden environment information is exposed. Voice, role, opener, urgency, and runtime–cost analyses confirm that the observed failures are not artifacts of any single configuration, and trace-level typology reveals stable model-specific bargaining styles (Appendices H.3, H.5).

# 5 Conclusion

TERMS-BENCH reframes negotiation evaluation away from black-box LLM-vs.-LLM play toward agent-attributable diagnosis. By using the environment itself as verifier, its bilateral price-negotiation instantiation exposes structured failures hidden by outcome-only evaluation and grounds those diagnoses in economic bargaining primitives. Its goal is diagnostic: environment-side verifiability enables scalable, human-annotation-free capability isolation, while human studies and richer multi-issue settings remain important next steps for validating transfer beyond bilateral price negotiation.

# Acknowledgement

We thank Professor Ludwig Schmidt for his insightful feedback during the early development of this benchmark. We also thank Professor Alvin Roth for his valuable perspectives on the economics of bargaining theory and its relevance to our work. His foundational work on *Axiomatic Models of Bargaining* has been an important source of inspiration for the economic grounding of our framework.

E.Z. acknowledges support from the Stanford Graduate Fellowship (SGF) for Sciences and Engineering and a Jump Trading PhD Fellowship. F.Z. acknowledges support from SGF. A.P. and B.E. acknowledge support from the Knight-Hennessy scholarship. J.B. gratefully acknowledges support from DoD through the grants Air Force Office of Scientific Research under award number FA9550-20-1-0397 and ONR 1398311, also support from NSF via grants 2229012, 2312204, 2403007 is gratefully acknowledged.

# References

Sahar Abdelnabi, Amr Gomaa, Sarath Sivaprasad, Lea Schönherr, and Mario Fritz. Cooperation, competition, and maliciousness: LLM-stakeholders interactive negotiation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Lakshya A Agrawal, Shangyin Tan, Dilara Soylu, Noah Ziems, Rishi Khare, Krista Opsahl-Ong, Arnav Singhvi, Herumb Shandilya, Michael J Ryan, Meng Jiang, Christopher Potts, Koushik Sen, Alexandros G. Dimakis, Ion Stoica, Dan Klein, Matei Zaharia, and Omar Khattab. Gepa: Reflective prompt evolution can outperform reinforcement learning, 2026. URL https://arxiv.org/abs/2507.19457.

Anthropic. Introducing claude opus 4.6, February 2026a. URL https://www.anthropic.com/news/claude-opus-4-6.

Anthropic. Introducing claude opus 4.7. https://www.anthropic.com/news/claude-opus-4-7, 2026b.

Lawrence M. Ausubel, Peter Cramton, and Raymond J. Deneckere. Bargaining with incomplete information. In Robert J. Aumann and Sergiu Hart, editors, *Handbook of Game Theory with Economic Applications*, volume 3, pages 1897–1945. Elsevier, 2002. doi: 10.1016/S1574-0005(02)03013-8.

Tim Baarslag, Mark J. C. Hendrikx, Koen V. Hindriks, and Catholijn M. Jonker. Measuring the performance of online opponent models in automated bilateral negotiation. In *AI 2012: Advances in Artificial Intelligence*, pages 1–14. Springer, 2012. doi: 10.1007/978-3-642-35101-3_1.

Tim Baarslag, Katsuhide Fujita, Enrico H. Gerding, Koen V. Hindriks, Takayuki Ito, Nicholas R. Jennings, Catholijn M. Jonker, Sarit Kraus, Raz Lin, Valentin Robu, and Colin R. Williams. Evaluating practical negotiating agents: Results and analysis of the 2011 international competition. *Artificial Intelligence*, 198: 73–103, 2013. doi: 10.1016/j.artint.2012.09.004.

Tim Baarslag, Koen V. Hindriks, and Catholijn M. Jonker. Effective acceptance conditions in real-time automated negotiation. *Decision Support Systems*, 60:68–77, 2014. doi: 10.1016/j.dss.2013.05.021.

Tim Baarslag, Mark J. C. Hendrikx, Koen V. Hindriks, and Catholijn M. Jonker. Learning about the opponent in automated bilateral negotiation: A comprehensive survey of opponent modeling techniques. *Autonomous Agents and Multi-Agent Systems*, 30(5):849–898, 2016. doi: 10.1007/s10458-015-9309-1.

Linda Babcock and George Loewenstein. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives*, 11(1):109–126, 1997.

Max H Bazerman and Margaret A Neale. *Negotiating Rationally*. Free Press, New York, 1992. ISBN 9780029019863.

Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. How well can llms negotiate? negotiationarena platform and analysis, 2024. URL https://arxiv.org/abs/2402.05863.

ByteDance. Doubao-seed-2.0-pro. https://www.volcengine.com/product/doubao, 2026.

Aayan Chatterjee, Sydney Miller, and Nitya Parepally. AgreeMate: Teaching LLMs to haggle. *arXiv preprint arXiv:2412.18690*, 2024.

Kalyan Chatterjee and William Samuelson. Bargaining under incomplete information. *Operations Research*, 31 (5):835–851, 1983. URL https://www.jstor.org/stable/170889.

Kushal Chawla, Jaysa Ramirez, Rene Clever, Gale Lucas, Jonathan May, and Jonathan Gratch. CaSiNo: A corpus of campsite negotiation dialogues for automatic negotiation systems. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3167–3185, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.254. URL https://aclanthology.org/2021.naacl-main.254/.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. Evaluating large language models trained on code, 2021. URL https://arxiv.org/abs/2107.03374.

Myke C. Cohen, Zhe Su, Hsien-Te Kao, Daniel Nguyen, Spencer Lynch, Maarten Sap, and Svitlana Volkova. Exploring big five personality and ai capability effects in llm-simulated negotiation dialogues, 2025. URL https://arxiv.org/abs/2506.15928.

DeepSeek. Deepseek v4 pro. https://api-docs.deepseek.com/, 2026.

DeepSeek-AI. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Burak Dindaroglu and Seda Ertac. An empirical study of sequential offer bargaining during the festival of sacrifice. *Journal of Economic Psychology*, 101:102707, 2024. doi: 10.1016/j.joep.2024.102707.

Jinhao Duan, Renming Zhang, James Diffenderfer, Bhavya Kailkhura, Lichao Sun, Elias Stengel-Eskin, Mohit Bansal, Tianlong Chen, and Kaidi Xu. Gtbench: Uncovering the strategic reasoning limitations of llms via game-theoretic evaluations, 2024. URL https://arxiv.org/abs/2402.12348.

Yann Dubois, Xuechen Li, Rohan Taori, Tianyi Zhang, Ishaan Gulrajani, Jimmy Ba, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. AlpacaFarm: A simulation framework for methods that learn from human feedback. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Peyman Faratin, Carles Sierra, and Nick R. Jennings. Negotiation decision functions for autonomous agents. *Robotics and Autonomous Systems*, 24(3–4):159–182, 1998. doi: 10.1016/S0921-8890(98)00029-3.

Google. google/gemma-4-31B-it. https://huggingface.co/google/gemma-4-31B-it, 2026.

Google DeepMind. Gemini 3.1 pro preview. https://ai.google.dev/gemini-api/docs/models/gemini-3.1-pro-preview, 2026.

Matthew Grennan. Price discrimination and bargaining: Empirical evidence from medical devices. *American Economic Review*, 103(1):145–177, 2013. doi: 10.1257/aer.103.1.145.

Matthew Grennan. Bargaining ability and competitive advantage: Empirical evidence from medical devices. *Management Science*, 60(12):3011–3025, 2014. doi: 10.1287/mnsc.2014.2006.

Sergiu Hart. Axiomatic approaches to coalitional bargaining. In Reinhard Selten, editor, *Rational Interaction: Essays in Honor of John C. Harsanyi*, pages 305–320. Springer, 1992.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2333–2343, 2018.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. URL https://openreview.net/forum?id=7Bywt2mQsCe.

Silke Herold, Jonas Heller, Frank Rozemeijer, and Dominik Mahr. Brave new procurement deals: An experimental study of how generative artificial intelligence reshapes buyer–supplier negotiations. *Journal of Purchasing and Supply Management*, 31(4):101012, 2025. doi: 10.1016/j.pursup.2025.101012.

Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lizhou Jiang, Mingyu Jin, Liangyu Fan, Fei Sun, William Wang, Xiang Wang, and Yongfeng Zhang. Game-theoretic LLM: Agent workflow for negotiation games. *arXiv preprint arXiv:2411.05990*, 2024.

Yu Jen Huang and Rafik Hadfi. How personality traits influence negotiation outcomes? a simulation based on large language models. *arXiv preprint arXiv:2407.11549*, 2024.

Minyoung Hwang, Luca Weihs, Chanhee Park, Kimin Lee, Aniruddha Kembhavi, and Kiana Ehsani. Promptable behaviors: Personalizing multi-objective rewards from human preferences. *arXiv preprint arXiv:2312.09337*, 2023.

Ruipeng Jia, Yunyi Yang, Yongbo Gai, Kai Luo, Shihao Huang, Jianhe Lin, Xiaoxi Jiang, and Guanjun Jiang. Writing-zero: Bridge the gap between non-verifiable tasks and verifiable rewards, 2025. URL https://arxiv.org/abs/2506.00103.

Carlos E. Jimenez, John Yang, Alexander Wettig, Shunyu Yao, Kexin Pei, Ofir Press, and Karthik Narasimhan. Swe-bench: Can language models resolve real-world github issues?, 2024. URL https://arxiv.org/abs/2310.06770.

Bharadwaj Kadiyala, Robert Phillips, A. Serdar Şimşek, and Garrett van Ryzin. Predicting transaction outcomes under customized pricing with discretion: A structural estimation approach. *Production and Operations Management*, 32(6):1654–1673, 2023. doi: 10.1111/poms.13931.

Emin Karagözoğlu and Martin G. Kocher. Bargaining under time pressure from deadlines. *Experimental Economics*, 22(2):419–440, 2019. doi: 10.1007/s10683-018-9579-y.

Deuksin Kwon, Emily Weiss, Tara Kulshrestha, Kushal Chawla, Gale Lucas, and Jonathan Gratch. Are LLMs effective negotiators? systematic evaluation of the multifaceted capabilities of LLMs in negotiation dialogues. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 5391–5413, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-emnlp.310. URL https://aclanthology.org/2024.findings-emnlp.310/.

Nathan Lambert, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, Lester James V. Miranda, Alisa Liu, Nouha Dziri, Shane Lyu, Yuling Gu, Saumya Malik, Victoria Graf, Jena D. Hwang, Jiangjiang Yang, Ronan Le Bras, Oyvind Tafjord, Chris Wilhelm, Luca Soldaini, Noah A. Smith, Yizhong Wang, Pradeep Dasigi, and Hannaneh Hajishirzi. Tülu 3: Pushing frontiers in open language model post-training. *arXiv preprint arXiv:2411.15124*, 2024.

Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. Deal or no deal? end-to-end learning for negotiation dialogues. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2443–2453, 2017.

Xianyang Liu, Shangding Gu, and Dawn Song. Agenticpay: A multi-agent llm negotiation system for buyer-seller transactions, 2026. URL https://arxiv.org/abs/2602.06008.

Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, Shudan Zhang, Xiang Deng, Aohan Zeng, Zhengxiao Du, Chenhui Zhang, Sheng Shen, Tianjun Zhang, Yu Su, Huan Sun, Minlie Huang, Yuxiao Dong, and Jie Tang. AgentBench: Evaluating LLMs as agents. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.

Zhenzhong Ma. Exploring the relationships between the big five personality factors, conflict styles, and bargaining behaviors. In *SSRN Electronic Journal*, June 2005. doi: 10.2139/ssrn.735063. URL https://ssrn.com/abstract=735063. Available at SSRN.

Shaoguang Mao, Yuzhe Cai, Yan Xia, Wenshan Wu, Xun Wang, Fengyi Wang, Tao Ge, and Furu Wei. Alympics: Llm agents meet game theory – exploring strategic decision-making with ai agents, 2024. URL https://arxiv.org/abs/2311.03220.

Microsoft Azure and xAI. Grok 4.2 reasoning, April 2026. URL https://ai.azure.com/catalog/models/grok-4-20-reasoning. Model catalog entry for xAI Grok 4.2 Reasoning. Accessed: 2026-05-04.

Moonshot AI. Kimi k2.6. https://platform.kimi.ai/, 2026.

Roger B Myerson. Two-person bargaining problems with incomplete information. *Econometrica: Journal of the Econometric Society*, 52(2):461–487, 1984. doi: 10.2307/1911499. URL https://www.jstor.org/stable/1911499.

Roger B Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281, 1983.

John F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

Jihwan Oh, Murad Aghazada, Yooju Shin, Se-Young Yun, and Taehyeon Kim. Merit feedback elicits better bargaining in llm negotiators, 2026. URL https://arxiv.org/abs/2602.10467.

OpenAI. Gpt-4o mini: Advancing cost-efficient intelligence. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/, 2024.

OpenAI. Introducing GPT-5.5, 2026a. URL https://openai.com/index/introducing-gpt-5-5/.

OpenAI. Gpt-5.4. https://platform.openai.com/docs/models, 2026b.

Hannes M. Petrowsky, Lea Boecker, Yannik A. Escher, Marie-Lena Frech, Malte Friese, Adam D. Galinsky, Brian Gunia, Alice J. Lee, Michael Schaerer, Martin Schweinsberg, Meikel Soliman, Roderick Swaab, Eve S. Troll, Marcel Weber, and David D. Loschelder. The power and peril of first offers in negotiations: A conceptual, meta-analytic, and experimental synthesis. *Organizational Behavior and Human Decision Processes*, 191:104448, 2025. doi: 10.1016/j.obhdp.2025.104448.

Qwen Team. Qwen3.6-plus. https://qwen.ai/, 2026.

Howard Raiffa. *The Art and Science of Negotiation*. Harvard University Press, 1982.

Howard Raiffa, John Richardson, and David Metcalfe. *Negotiation Analysis: The Science and Art of Collaborative Decision Making*. Harvard University Press, 2002.

Alvin E. Roth. Individual rationality and nash's solution to the bargaining problem. *Mathematics of Operations Research*, 2(1):64–65, 1977.

Alvin E. Roth. *Axiomatic Models of Bargaining*, volume 170 of *Lecture Notes in Economics and Mathematical Systems*. Springer, 1979. doi: 10.1007/978-3-642-51570-5.

Thomas C Schelling. *The Strategy of Conflict*. Harvard University Press, 1960.

Johannes Schneider, Stefanie Haag, and Leona C. Kruse. Negotiating with LLMs: Prompt hacks, skill gaps, and reasoning deficits. *arXiv preprint arXiv:2312.03720*, 2024.

Martin Schweinsberg, Stefan Thau, and Madan M. Pillutla. Negotiation impasses: Types, causes, and resolutions. *Journal of Management*, 48(1):49–76, 2022. doi: 10.1177/01492063211021657.

Aldís Guðný Sigurðardóttir, Ali Hotait, and Tilman Eichstädt. Buyer and seller differences in business-to-business negotiations. *Negotiation Journal*, 35(2):297–331, 2019. doi: 10.1111/nejo.12289.

Alice F. Stuhlmacher, Treena L. Gillespie, and Matthew V. Champagne. The impact of time pressure in negotiation: A meta-analysis. *International Journal of Conflict Management*, 9(2):97–116, 1998.

Kian Siong Tey, Michael Schaerer, Nikhil Madan, and Roderick I. Swaab. The impact of concession patterns on negotiations: When and why decreasing concessions lead to a distributive disadvantage. *Organizational Behavior and Human Decision Processes*, 165:153–166, 2021. doi: 10.1016/j.obhdp.2021.05.003.

William Thomson. Bargaining and the theory of cooperative games: John nash and beyond. Technical Report Working Paper No. 554, University of Rochester, 2009.

William Thomson. On the axiomatic theory of bargaining: A survey of recent results. *Review of Economic Design*, 26(4):491–542, 2022. doi: 10.1007/s10058-022-00319-1.

Tian Xia, Zhiwei He, Tong Ren, Yibo Miao, Zhuosheng Zhang, Yang Yang, and Rui Wang. Measuring bargaining abilities of llms: A benchmark and a buyer-enhancement method, 2024. URL https://arxiv.org/abs/2402.15813.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2023.

Denis Yarats and Mike Lewis. Hierarchical text generation and planning for strategic dialogue. *arXiv preprint arXiv:1712.05846*, 2018.

Z.AI. Glm-5.1. https://docs.z.ai/, 2026.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Shuyan Zhou, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig. WebArena: A realistic web environment for building autonomous agents. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024.

APPENDIX

# Contents

# A Economic Basis of TERMS-BENCH

TERMS-BENCH is a general framework for constructing verifiable negotiation environments for the purpose of evaluating LLM agents. In this paper, we instantiate it in a stylized bilateral price-negotiation setting. The goal is not absolute realism, since real-world negotiations vary widely across contexts, but diagnostic control: to distill core bargaining primitives into a minimal environment where agent behavior can be attributed to specific capabilities and failures.

Although controlled, the instantiation is not arbitrary. It combines a canonical Bayesian bargaining backbone with practitioner-facing features such as bounded prices, explicit no-deal cases, urgency, language-mediated signaling, and data-grounded public context. We organize the literature grounding around five simulator components: (i) the Bayesian bargaining environment; (ii) the interaction protocol; (iii) the regime design; (iv) the counterpart-family mechanism; and (v) the evaluation metrics. This preserves the economic structure of bargaining while surfacing concrete decision problems for deployed agents, and leaves room for customization toward procurement, pricing, and marketplace applications.

**(i) Bayesian bargaining environment setup.** At the environment level, the bilateral instantiation follows the classical bargaining representation of a feasible set together with a disagreement outcome, as in Nash's bargaining problem and Roth's axiomatic treatment of bargaining [Nash, 1950, Roth, 1979, Thomson, 2009]. It then introduces incomplete information over payoff-relevant parameters, most centrally reservation values, consistent with the standard economics literature on bargaining under incomplete information [Chatterjee and Samuelson, 1983, Ausubel et al., 2002]. Hart's discussion of the axiomatic tradition is especially relevant here: when the bargaining problem is specified without an extensive form, procedural assumptions are "necessarily ... ad hoc," so solutions should rest on "general principles."[5] In this sense, the bilateral instantiation should be read as a controlled protocol layered on top of a standard bargaining scaffold, rather than as an attempt to recover a *uniquely correct* real-world bargaining procedure [Hart, 1992, Thomson, 2022].

**(ii) Interaction protocol design.** Private reservation values are a canonical abstraction in bargaining theory [Chatterjee and Samuelson, 1983, Ausubel et al., 2002], and they remain common in empirical pricing and customized-negotiation models, where transaction outcomes are explained using latent reserve prices, willingness-to-pay, and bargaining power [Grennan, 2013, Kadiyala et al., 2023]. Likewise, sequential offer exchange is a standard bargaining protocol in both economic theory and automated-negotiation research [Ausubel et al., 2002, Faratin et al., 1998, Baarslag et al., 2016], and it has also been documented in naturally occurring field settings [Dindaroglu and Ertac, 2024]. These ingredients make the interaction protocol more than a generic alternating-offer game: it captures the recurring setting in which an agent must act under private constraints, interpret counterpart behavior from sequential signals, and decide whether to continue, agree, or walk away.

**(iii) Regime design.** The regime design is likewise grounded in core distinctions from bargaining and negotiation research. The overlap and no-deal regimes formalize whether the bargaining problem admits a mutually acceptable agreement at all, which corresponds to the basic distinction between feasible agreement and disagreement as the individually rational fallback outcome in classical bargaining theory [Nash, 1950, Roth, 1977]. Modeling this distinction explicitly is important because negotiation research has long emphasized that impasse is not merely a failure to optimize price; it is often a substantively different outcome that must be analyzed in its own right [Schweinsberg et al., 2022]. The urgency-shift regime isolates a second economically meaningful axis: time pressure. A broad literature shows that deadlines and time pressure shape concession behavior, agreement likelihood, and the incidence or timing of disagreement [Stuhlmacher et al., 1998, Karagözoğlu and Kocher, 2019]. By holding reservation geometry fixed while shifting urgency, the benchmark separates adaptation to counterpart time pressure from simple variation in feasible surplus. This is precisely the kind of controlled design needed to diagnose whether an agent is responding to latent temporal pressure rather than to a trivially easier or harder price geometry.

**(iv) Counterpart-family mechanism and design.** The counterpart policy should be understood as a literature-grounded benchmark kernel, *not* as a structurally estimated or purely hand-engineered model of human behavior. Its components—opening behavior, concession dynamics, acceptance conditions, walk-away under time pressure, and latent-type inference from offers and messages—mirror dimensions that are widely studied in economics, organizational behavior, and automated negotiation [Faratin et al., 1998, Baarslag et al., 2014, 2016]. Prior work shows that first offers materially affect negotiated outcomes [Petrowsky et al., 2025], concession patterns

[5] Hart argues that when the extensive form of bargaining is not specified, such procedural assumptions are "necessarily ... ad hoc," so solutions should be based on "general principles" [Hart, 1992, p. 317].

shape beliefs about reservation values and later counteroffers [Tey et al., 2021], and deadlines or time pressure influence bargaining dynamics and disagreement patterns [Stuhlmacher et al., 1998, Karagözoğlu and Kocher, 2019]. The role of the counterpart families is therefore not to claim an exhaustive taxonomy of real negotiators, but to induce interpretable stress conditions over these well-studied dimensions. This makes it possible to test whether an agent is brittle to sparse evidence, over-reacts to cue style, mishandles time pressure, or fails under harder bargaining postures.

**(v) Evaluation metrics.** The metric design is motivated by the same principle. The disagreement outcome and individual rationality are constitutive of the bargaining problem itself [Nash, 1950, Roth, 1977], and under incomplete information, individual rationality binds any feasible bargaining mechanism [Myerson and Satterthwaite, 1983], so a negotiation benchmark should not collapse all performance into a single agreement-rate number. The negotiation-analysis and automated-negotiation literatures accordingly evaluate systems along multiple dimensions rather than a single scalar score [Raiffa et al., 2002, Baarslag et al., 2012, 2013, 2016]. Accordingly, the headline metrics are grouped around terminal value, agreement calibration, opponent modeling, and protocol compliance. Feasible terminal value captures surplus realization relative to the Myerson and Satterthwaite [1983] theoretical ceiling; agreement calibration reflects the long-standing finding that systematic miscalibration produces impasse even when a feasible agreement exists [Babcock and Loewenstein, 1997]; false agreement in no-deal regimes operationalizes individual rationality as a first-class violation; belief accuracy reflects opponent modeling as a prerequisite for efficient negotiation [Baarslag et al., 2016] and a distinct axis in recent LLM negotiation evaluation [Kwon et al., 2024]; and protocol compliance reflects the Schelling [1960] view that constraint adherence is itself strategic behavior, carried into modern LLM-agent evaluation [Abdelnabi et al., 2024].

Accordingly, the bilateral instantiation is intended to measure whether an agent can satisfy prerequisite competencies that recur in bilateral buyer–seller settings: preserving individual rationality, recognizing no-deal cases, adapting to urgency and concession behavior, responding to opening anchors, and using language cues without allowing language to override structured economic actions. This focus is consistent with empirical work showing that negotiated outcomes in business markets depend on latent willingness-to-pay, bargaining ability, and sequential interaction [Grennan, 2014, Dindaroglu and Ertac, 2024], that real B2B negotiations exhibit systematic buyer–seller differences in tactics and information use [Sigurðardóttir et al., 2019], and that AI negotiation style can affect discounts, speed, trust, and willingness for future interaction in buyer–supplier negotiations [Herold et al., 2025]. Taken together, these findings support the practitioner-facing relevance of this instantiation: while it is not a perfect simulator of deployed negotiations, it evaluates capabilities that are closely tied to real negotiated pricing and procurement workflows.

Because TERMS-BENCH fixes the counterpart kernel while hiding latent state from the evaluated agent, it supports failure attribution at a level that outcome-only negotiation arenas do not: poor performance can be traced to inference errors, concession-control failures, brittle cue use, or explicit policy violations rather than being absorbed into a single deal-rate or surplus number. In this sense, the bilateral price-negotiation setting serves as a controlled diagnostic instantiation of TERMS-BENCH.

# B Specification of the Bilateral Price-Negotiation Instantiation

This section consolidates the formal specification deferred from Section 2. It complements §2 by enumerating agent information state, the action and observation spaces, termination cases, and hard constraints in detail.

## B.1 Information State and Belief Dynamics

We expand on the information-state space and belief update deferred from §2.1. The information state $s_k = (h_k, x_A, b_k)$ lives in $\mathcal{S} := \mathcal{H} \times \mathcal{X}_A \times \mathcal{B}$, with $\mathcal{H}$ the space of public interaction histories, $\mathcal{X}_A$ the space of private side information available to the agent (e.g., market statistics or profile-level context), and $\mathcal{B} \subseteq \Delta(T_B)$ the space of internal belief states over opponent types. The belief is updated as $b_k = \Psi(h_k)$, where the update mechanism $\Psi$ is part of the agent policy and is not specified by $\mu$. The true counterpart type $t_B$ is not directly observable; the agent must infer it from the sequence of prices and language realizations.

The policy $\pi : \mathcal{S} \to \mathcal{A}$ thus jointly determines (i) how the agent performs latent-state inference over opponent types from observed signals, and (ii) how it selects strategic actions to optimize outcomes under the fixed bargaining environment and protocol. We do not restrict the implementation of $\pi$: it may be a prompted LLM, a rule-based agent, a learned policy, or a model-based planner.

### B.2 Action and Observation Spaces

**Action.** At each round $k$ the agent selects $a_k = (d_k, p_k, l_k)$, with $d_k \in \{\texttt{Offer}, \texttt{Accept}, \texttt{Reject}\}$, $p_k \in [p_{\min}, p_{\max}]$ active only when $d_k = \texttt{Offer}$, and $l_k \in \mathcal{L}$ a free-form natural-language message. `Accept` binds at the counterpart's last offered price; `Reject` terminates with disagreement $\perp$. The message $l_k$ lets agents justify offers, request information, signal constraints, or employ other communicative strategies. Messages may influence counterpart behavior through the cue layer of $\pi_B$ (Appendix C), but only the economic decision $d_k$ and price $p_k$ directly determine negotiation outcomes and constraint compliance.

**Observation.** At each round $k$, the evaluated agent observes

$$o_k = (p_k^B, l_k^B) \in \mathcal{O}, \tag{11}$$

where $p_k^B \in [p_{\min}, p_{\max}]$ is the counterpart's price offer and $l_k^B \in \mathcal{L}$ is the counterpart's natural-language message. The simulator-internal cue variables $(\tilde{s}_k, \tilde{c}_k) \in \{\texttt{positive}, \texttt{neutral}, \texttt{negative}\} \times \{\texttt{Concede}, \texttt{Hold}, \texttt{Pressure}\}$ are not part of $\mathcal{O}$ (see §2.2); they parameterize the realization of $l_k^B$ but are never themselves observed. The public interaction history up to round $k$ is $h_k = (o_1, a_1, \ldots, o_k)$, accumulating all observations and the agent's past actions. The agent's private context $x_A$ includes its assigned role, its own reservation price $r_A$, and any additional side information.

### B.3 Termination and Constraints

**Termination.** A negotiation episode ends in one of five scenarios:

1. The agent chooses `Accept`; the outcome is the counterpart's last offered price.
2. The agent chooses `Reject`; the outcome is disagreement $\perp$.
3. The counterpart accepts the agent's offer; the outcome is the agent's proposed price.
4. The counterpart terminally rejects (walk-away); the outcome is disagreement $\perp$.
5. The round limit $K$ is reached without agreement; the outcome is disagreement $\perp$.

**Constraints.** All `Offer` actions must satisfy: (i) price bounds $p_{\min} \le p_k \le p_{\max}$; (ii) monotonic concession: for buyer agents, $p_k \ge p_{k-1}$ where $k, k-1$ index the agent's own offers; for seller agents, $p_k \le p_{k-1}$; (iii) turn budget $k \le K$. Violations of (i) or of individual rationality (accepting or offering a price strictly worse than $r_A$) are recorded as critical violations in $\mathrm{CritViol}\%$. Monotonicity and turn-budget violations are recorded as secondary procedural diagnostics (Appendix F).

### B.4 Type Parameterization: Stance vs. Personality

Our type parameterization $t_i = (r_i, \kappa_i, \eta_i)$ excludes personality traits, despite their documented correlations with negotiation outcomes in recent LLM studies [Cohen et al., 2025]. Following Ma [2005] and Huang and Hadfi [2024], we instead model strategic stance $\eta_i \in \{\texttt{conciliatory}, \texttt{neutral}, \texttt{aggressive}\}$ as the primary behavioral driver, since conflict-management styles have been shown to mediate the relationship between personality and bargaining behavior. Incorporating personality as an orthogonal communication-level dimension (e.g., as a separate language-realization parameter that does not affect the economic kernel) remains a direction for future work.

## C Counterpart Behavior Families

Counterpart behavior in TERMS-BENCH's bilateral price-negotiation is governed by the fixed policy in Section 3.2 and the cue generation mechanism in Appendix C.5. Behavior families correspond to presets of this common simulator. They do not introduce separate acceptance, walk-away, or offer-generation formulas.

We use six main families. Four form a diagnostic core over two axes:

(i) **Economic reactivity.** How strongly does the counterpart's economic behavior depend on the agent's recent offer trajectory, relative to the counterpart's latent type $(r_B, \kappa_B, \eta_B)$?

(ii) **Cue reliability.** Do the language-facing sentiment and posture cues reliably reflect the counterpart's latent stance $\eta_B$, or are they collapsed to noncommittal states?

| Family | Economic preset | Cue channel | $\lambda_2(\texttt{C}, \texttt{N}, \texttt{A})$ | $\eta_B$ prior | $\bar{\sigma}_p$ | Role |
|---|---|---|---|---|---|---|
| Candid | type-instrumental | accurate | $(0.30, 0.50, 1.00)$ | uniform | low | core |
| Taciturn | type-instrumental | uninformative | $(0.30, 0.50, 1.00)$ | uniform | low | core |
| Expressive | high-reactivity | accurate | $(0.45, 0.90, 1.80)$ | uniform | mod. | core |
| Strategic | high-reactivity | uninformative | $(0.45, 0.90, 1.80)$ | uniform | mod. | core |
| Stochastic | moderate-reactivity | noisy/weak | $(0.35, 0.70, 1.40)$ | uniform | high | noise floor |
| Adversarial | hardball | pressuring | $(0.60, 1.40, 2.60)$ | aggressive-skewed | low | stress |

Table 3: Summary of counterpart behavior families. The first four families form the diagnostic core, crossing economic reactivity with cue reliability. The tuple $\lambda_2(\texttt{C}, \texttt{N}, \texttt{A})$ gives the concession-reciprocity coefficient for conciliatory, neutral, and aggressive stance types. The `Stochastic` family is a noise-floor condition with high price noise and weak/noisy cues, while `Adversarial` is a hardball stress condition with an aggressive-skewed stance prior and pressuring cues.

Two additional families, `Stochastic` and `Adversarial`, are stress conditions: `Stochastic` provides a noise-floor condition under noisy price and cue channels, while `Adversarial` provides a hardball condition with an aggressive-skewed stance prior and pressuring cues.

1. *Candid counterparts* [type-instrumental economics, accurate cues]. This family replaces the earlier `Truthful` family. The counterpart's economic behavior is low-noise and strongly type-conditioned: reservation value sets the feasible boundary, urgency changes acceptance and concession timing, and stance changes the payoff consequences of rigidity and concession. Sentiment and strategic cues are generated from the base cue model, so language is informative about latent stance. This family is designed so that correct inference of $(r_B, \kappa_B, \eta_B)$ is instrumentally valuable for surplus extraction.
2. *Taciturn counterparts* [type-instrumental economics, uninformative cues]. Economic behavior follows the same type-instrumental preset as `Candid`, but the cue channel is collapsed to neutral, noncommittal states. This isolates inference from economic behavior alone: an agent that degrades relative to `Candid` is relying heavily on linguistic or stylistic cues.
3. *Expressive counterparts* [high-reactivity economics, accurate cues]. The cue channel remains accurate, but economic behavior is more strongly history-reactive. Counter-offers and acceptance probabilities respond more to the agent's recent concession pattern and rigidity. This family tests whether agents can use reliable cues while avoiding confusion between latent type and state-dependent reactions to their own behavior.
4. *Strategic counterparts* [high-reactivity economics, uninformative cues]. Economic behavior is strongly history-reactive, and the cue channel is uninformative. The counterpart is linguistically guarded while adapting tactically through price and acceptance behavior. This is the hardest core family for opponent modeling because both the economic and language channels provide imperfect evidence.
5. *Adversarial counterparts* [hardball economics, pressuring cues]. This family is an explicit stress test rather than part of the core factorial. The stance prior is skewed toward aggressive counterparts, economic reactivity is high, concessionary behavior is strongly exploited, and rigidity is punished for aggressive types. The cue channel is biased toward negative sentiment and `Pressure`. This family tests whether agents can recognize and adapt to hardball negotiation without either over-conceding to pressure or holding firm until walk-away.
6. *Stochastic counterparts* [moderate-reactivity economics, noisy/weak cues]. This family degrades both the price and cue channels through noise rather than through deliberate strategic concealment. Economic behavior uses a moderate reactivity preset, but price noise is high, so offer trajectories are less diagnostic of the underlying concession rule. The cue channel remains weakly coupled to latent stance but is made noisy through elevated sentiment variance and strategic-cue temperature, rather than being collapsed to neutral states. This family provides a noise-floor test for belief calibration and surplus extraction: a robust agent should avoid over-interpreting noisy price movements or unstable linguistic cues.

The family identity and parameter configuration are hidden from the negotiating agent and logged only for evaluator analysis. We report performance by family to distinguish failures of cue use, economic-channel inference, type-vs-state decomposition, and robustness to hardball pressure.

**Implementation under the canonical simulator.** Table 3 describes the six main counterpart behavior families. All six families are instantiated by varying only parameter presets already present in the fixed counterpart policy of Section 3.2 and the cue-generation mechanism of Appendix C.5. In particular, all families share the same opening-role protocol, randomized opening-offer model, acceptance model, walk-away hazard, counter-offer rule, and cue-generation interface.

Family differences arise through two sets of presets. First, the economic preset controls stance-dependent response parameters, including $\rho_{\mathcal{F}}(\eta_B)$, $\xi_{\mathcal{F}}(\eta_B)$, and $\lambda_{2,\mathcal{F}}(\eta_B)$, together with the price-noise scale and the stance prior over $\eta_B$. Second, the cue preset controls whether the language-facing sentiment and strategic-posture cues are informative, uninformative, noisy, or pressuring. The `Candid` and `Taciturn` families share the same type-instrumental economic preset and differ only in cue reliability; the `Expressive` and `Strategic` families share the same high-reactivity economic preset and differ only in cue reliability. The `Stochastic` family provides a noise-floor condition by using moderate economic reactivity with high price noise and weak/noisy cues, while the `Adversarial` family uses a hardball economic preset with an aggressive-skewed stance prior and pressuring cues.

## C.1 Family-Specific Economic Presets

The acceptance and counter-offer equations in Section 3.2 use stance-dependent coefficients $\rho_{\mathcal{F}}(\eta_B)$, $\xi_{\mathcal{F}}(\eta_B)$, and $\lambda_{2,\mathcal{F}}(\eta_B)$. We order stance types as

$$(\texttt{C}, \texttt{N}, \texttt{A}) := (\texttt{conciliatory}, \texttt{neutral}, \texttt{aggressive}).$$

Table 4 gives the economic preset used by each counterpart family. These presets instantiate the same acceptance and counter-offer formulas for all families.

| **Preset** | $\rho(\texttt{C},\texttt{N},\texttt{A})$ | $\xi(\texttt{C},\texttt{N},\texttt{A})$ | $\lambda_2(\texttt{C},\texttt{N},\texttt{A})$ | **Used by** |
|---|---|---|---|---|
| Type-instrumental | $(0,\ -0.25,\ -0.75)$ | $(+0.40,\ 0,\ -0.50)$ | $(0.30,\ 0.50,\ 1.00)$ | `Candid`, `Taciturn` |
| High-reactivity | $(0,\ -0.75,\ -1.50)$ | $(+0.40,\ 0,\ -0.75)$ | $(0.45,\ 0.90,\ 1.80)$ | `Expressive`, `Strategic` |
| Moderate-stochastic | $(0,\ -0.50,\ -1.10)$ | $(+0.35,\ 0,\ -0.60)$ | $(0.35,\ 0.70,\ 1.40)$ | `Stochastic` |
| Hardball | $(-0.25,\ -1.25,\ -2.25)$ | $(0,\ -0.50,\ -1.20)$ | $(0.60,\ 1.40,\ 2.60)$ | `Adversarial` |

Table 4: Family-specific economic presets. Coefficients are ordered by stance type: conciliatory, neutral, aggressive. Conciliatory counterparts are most receptive to firm but feasible bargaining: they reward rigidity in the acceptance model, but do not directly reward fast concession. Aggressive counterparts, by contrast, penalize pure rigidity and exploit rapid concession, making measured movement preferable. Thus, the agent's optimal concession posture depends on correctly inferring latent stance: firm bargaining is distinctly more effective against conciliatory counterparts, while aggressive counterparts punish pure rigidity and exploit rapid concession.

The type-instrumental preset is designed so that latent stance is economically meaningful without making recent agent history dominate the latent-type signal. In this preset, conciliatory counterparts are receptive to firm but feasible offers, neutral counterparts behave as the middle case, and aggressive counterparts penalize pure rigidity while exploiting rapid concession. This makes stance inference useful: agents should *not* use the same concession posture against all counterpart types.

On the other hand, the high-reactivity preset increases the influence of the agent's recent offer trajectory while preserving the same stance semantics. It is harder than the type-instrumental preset because observed price dynamics are more strongly confounded by the agent's own behavior. The stochastic preset uses moderate economic reactivity but is paired with elevated price and cue noise. The hardball preset combines strong exploitation of fast concession with penalties for rigidity, especially under aggressive stance. The adversarial family additionally uses an aggressive-skewed stance prior:

$$\Pr(\eta_B = \texttt{conciliatory}) = 0.05,\ \ \Pr(\eta_B = \texttt{neutral}) = 0.15,\ \ \Pr(\eta_B = \texttt{aggressive}) = 0.80.$$

## C.2 Economic Response Details

This section provides additional details for the economic response model in Section 3.2. We use the same notations as in Section 3.2. The response model separates three counterpart outcomes after an agent offer: acceptance of the current offer, terminal walk-away, and continued bargaining through a counter-offer. All counterpart behavior families use the same response equations; families differ only through parameter presets specified in Appendix C.1.

### C.2.1 Acceptance

Recall that $\bar{\Delta}_k \geq 0$ means that the offer is individually rational for the counterpart. The acceptance gate $\mathbf{1}\{\bar{\Delta}_k \geq 0\}$ enforces this individual-rationality constraint: seller counterparts never accept prices below $r_B$, and buyer counterparts never accept prices above $r_B$.

We use the concave deadline clock

$$D_k := \frac{k}{K}, \qquad \widetilde{D}_k := \sqrt{D_k}, \qquad \widetilde{\bar{D}}_k := 1 - \widetilde{D}_k.$$

The square-root transformation spreads deadline pressure over the later part of the negotiation instead of concentrating it almost entirely in the final round.

Conditional on individual rationality, acceptance is stochastic:

$$a_k := \pi_B(d_k^B = \texttt{Accept} \mid p_k^A, t_B, h_{k-1}) = \mathbf{1}\{\bar{\Delta}_k \geq 0\}\sigma\big(g_\theta(p_k^A, t_B, k, h_{k-1})\big)\,,$$

where

$$g_\theta(p_k^A, t_B, k, h_{k-1}) = \alpha\bar{\Delta}_k + \beta\kappa_B - \gamma\widetilde{\bar{D}}_k + \rho_{\mathcal{F}}(\eta_B)\,\mathrm{ConcedeSpeed}_k + \xi_{\mathcal{F}}(\eta_B)\,\mathrm{Rigidity}_k.$$

Here $\mathcal{F}$ denotes the counterpart behavior family. The terms $\mathrm{ConcedeSpeed}_k$ and $\mathrm{Rigidity}_k$ are deterministic history features defined in Appendix C.3. The coefficients $\rho_{\mathcal{F}}(\eta_B)$ and $\xi_{\mathcal{F}}(\eta_B)$ are family- and stance-dependent. Negative values of $\rho_{\mathcal{F}}(\eta_B)$ mean that fast recent agent concession reduces the counterpart's willingness to accept, while positive values make concession more reciprocated. Positive values of $\xi_{\mathcal{F}}(\eta_B)$ make rigidity more effective; negative values make rigidity costly. In the type-instrumental families, this makes stance inference directly payoff-relevant: firmness, concession, and balanced movement have different consequences for conciliatory, neutral, and aggressive counterparts.

### C.2.2 Terminal Walk-Away

The walk-away branch is a reduced-form participation constraint. It prevents continuation from being costless when the agent repeatedly proposes prices that are outside the counterpart's individually rational region. The default hazard is deliberately sparse:

$$\omega_k = \mathbf{1}\{k \geq k_{\mathrm{walk}}\}\mathbf{1}\{\bar{\Delta}_k < 0\}\sigma\big(\phi_0 + \phi_\Delta[-\bar{\Delta}_k]_+ + \phi_T\tau_k^W\big)\,,$$

where

$$\tau_k^W = \min\left\{1, \max\left\{0, \frac{k - k_{\mathrm{walk}}}{K - k_{\mathrm{walk}}}\right\}\right\}.$$

Here $[-\bar{\Delta}_k]_+$ is the normalized reservation shortfall of the agent's current offer from the counterpart's perspective, and $\tau_k^W$ is a walk-away clock that equals zero when terminal exit first becomes available and approaches one near the round limit. In the default configuration we use

$$k_{\mathrm{walk}} = \left\lceil \frac{K}{2} \right\rceil, \qquad \phi_0 = -4.5, \qquad \phi_\Delta = 30.0, \qquad \phi_T = 1.5.$$

The hard gate $\mathbf{1}\{\bar{\Delta}_k < 0\}$ implies that walk-away does not alter ordinary feasible bargaining when the agent remains inside the counterpart's individually rational region. It is active primarily in no-deal episodes and in feasible episodes where the agent opens or persists outside the counterpart's reservation constraint. This makes the mechanism diagnostic rather than broadly punitive: a competent agent can still bargain firmly within the feasible region, but persistent non-individually-rational offers become increasingly risky late in the interaction.

**Why the default hazard omits urgency and stance.** We intentionally omit direct urgency and stance terms from the default walk-away hazard. Urgency already affects acceptance through $\beta\kappa_B$ and concession dynamics through $\lambda_1\kappa_B$. Stance already affects acceptance through $\rho_{\mathcal{F}}(\eta_B)$ and $\xi_{\mathcal{F}}(\eta_B)$, counter-offer dynamics through $\lambda_{2,\mathcal{F}}(\eta_B)$ and $\lambda_3, \lambda_4$, and cue generation through the family-specific cue model. Adding direct urgency or stance effects to walk-away would give the same latent variables additional behavioral channels and risk making them overly dominant. Keeping terminal exit type-neutral preserves the walk-away branch as a narrow participation-constraint mechanism.

### C.2.3 Termination Accounting

For $k < K$, the counterpart response probabilities are

$$\pi_B(\texttt{Accept}) = a_k, \qquad \pi_B(\texttt{Reject}) = (1 - a_k)\omega_k, \qquad \pi_B(\texttt{Offer}) = (1 - a_k)(1 - \omega_k).$$

At the terminal round, the remaining non-acceptance and non-walk-away mass is assigned to round-limit disagreement. In evaluation, we separately log the termination source:

$$\tau_{\mathrm{term}} \in \{\texttt{AgentAccept}, \texttt{CounterpartAccept}, \texttt{AgentReject}, \\ \texttt{CounterpartWalkAway}, \texttt{Timeout}, \texttt{AgreementViolation}\}.$$

This allows no-deal performance to be distinguished between disciplined agent exit, counterpart walk-away, timeout, and individually irrational agreement.

## C.3 History Feature Definitions

$\mathrm{ConcedeMagnitude}_k$, $\mathrm{ConcedeSpeed}_k$, and $\mathrm{Rigidity}_k$ are fixed scalar summaries of the agent's recent offer dynamics, computed deterministically from $h_{k-1}$. To make these features comparable across cheap and expensive products and across negotiation regimes, we normalize all offer changes by the public price range $R$, which is fixed within each episode and strictly positive by construction.

Let

$$s_A := \begin{cases} +1, & \text{if the agent role is buyer,} \\ -1, & \text{if the agent role is seller,} \end{cases}$$

so that $s_A(p_j^A - p_{j-1}^A) > 0$ always corresponds to a *concessionary* move by the agent, while $s_A(p_j^A - p_{j-1}^A) < 0$ corresponds to hardening or reversal. Let

$$\mathcal{J}_k := \{j : \max(2, k-3) \le j \le k-1\}$$

index the most recent rounds for which two consecutive agent offers are available.

- *Concession magnitude.* We define the normalized recent concession magnitude by

$$\mathrm{ConcedeMagnitude}_k = \begin{cases} \dfrac{1}{|\mathcal{J}_k|} \displaystyle\sum_{j \in \mathcal{J}_k} \dfrac{\max\{0,\ s_A(p_j^A - p_{j-1}^A)\}}{R}, & \text{if } |\mathcal{J}_k| \ge 1, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

  This feature measures how much the agent has recently conceded, on average, expressed as a fraction of the episode's public price range. Non-concessionary moves, including hardening or reversal, contribute zero rather than being counted as eagerness.

- *Concession speed.* We define the normalized directional concession speed by

$$\mathrm{ConcedeSpeed}_k = \begin{cases} \dfrac{1}{|\mathcal{J}_k|} \displaystyle\sum_{j \in \mathcal{J}_k} \dfrac{s_A(p_j^A - p_{j-1}^A)}{R}, & \text{if } |\mathcal{J}_k| \ge 1, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

  By construction, larger values indicate more concessionary recent behavior, values near zero indicate relatively stationary or mixed behavior, and negative values indicate recent hardening or reversal. This sign convention is role-invariant: positive $\mathrm{ConcedeSpeed}_k$ has the same behavioral meaning for buyers and sellers.

- *Rigidity.* We define the rigidity indicator by

$$\mathrm{Rigidity}_k = \begin{cases} 1, & \text{if } \dfrac{\max\{0,\ s_A(p_{k-1}^A - p_{k-2}^A)\}}{R} < \tau_{\mathrm{rigid}}, \\ 0, & \text{otherwise,} \end{cases} \qquad \tau_{\mathrm{rigid}} \in (0, 1), \tag{14}$$

  whenever the last two agent offers are available; otherwise $\mathrm{Rigidity}_k := 0$. Thus, $\mathrm{Rigidity}_k = 1$ indicates that the agent made only a minimal recent concession relative to the episode's price scale. In particular, holding firm or moving in a hardening direction both count as rigid behavior under this definition.

Unless otherwise stated, we set $\tau_{\mathrm{rigid}} = 0.1$.

**Boundary conditions with mixed opener roles.** History features are computed from realized public histories and are well-defined under both opener roles. If fewer than two agent offers have been observed, then

$$\mathrm{ConcedeMagnitude}_k = \mathrm{ConcedeSpeed}_k = \mathrm{Rigidity}_k = 0.$$

Thus, when the agent opens, the counterpart's first response to the agent's opening offer is evaluated without artificial history-reactivity. Once the agent has made at least two offers, the usual role-normalized definitions apply.

Similarly, counterpart-side concession features used in cue generation are set to zero until two counterpart offers have been observed. This covers both cases: the ordinary counterpart-opens case, where the first counterpart offer is the episode anchor, and the agent-opens case, where the first counterpart offer may occur only after the counterpart declines to accept the agent's opening offer.

## C.4 Opening Role and Opening-Offer Generation

At the start of each episode, the protocol specifies an opener role

$$\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}.$$

The opener role is an episode attribute of the interaction protocol, not a separate counterpart behavior family. In the main experimental suite, $\chi$ is assigned by blocked allocation so that agent-opens and counterpart-opens episodes are exactly balanced within each regime–family cell (Section 4.1). More general stochastic generators may instead draw $\chi$ from a user-specified distribution.

If $\chi = \texttt{CounterpartOpens}$, the counterpart produces the first price proposal. If $\chi = \texttt{AgentOpens}$, the evaluated agent produces the first price proposal, and `Accept` is unavailable because no counterpart offer has yet been observed. The counterpart then responds using the economic response model in Section 3.2. If it neither accepts nor walks away, its first price proposal is generated by the opening-offer model below, since no previous counterpart offer exists.

**Randomized counterpart opening harshness.** To avoid making the counterpart reservation value nearly invertible from its first price, we randomize the harshness of the counterpart's first offer. The episode-level harshness variable is

$$d_{0,e} \sim \mathcal{D}_{\text{open}}, \qquad \mathcal{D}_{\text{open}} = \text{Uniform}(d_{\min}, d_{\max}),$$

with default $d_{\min} = 0.20$, $d_{\max} = 0.80$. For controlled ablations, $\mathcal{D}_{\text{open}}$ may be replaced by stratum-specific intervals corresponding to soft, medium, or harsh counterpart anchors. The realized $d_{0,e}$ is hidden from the agent and logged for analysis. It is used whenever the counterpart generates its first offer, whether the counterpart opens the episode or makes its first offer after declining an agent opening. If no counterpart offer is ever generated, the variable is unused for that trajectory.

In difficulty grading, counterpart opening harshness is treated as an environment-side opening difficulty variable only when $\chi = \texttt{CounterpartOpens}$. When $\chi = \texttt{AgentOpens}$, the episode's first price is chosen by the evaluated agent and is therefore a policy decision rather than an instance property. If the counterpart later produces its first offer in an agent-opens episode, the realized anchor is logged as a response diagnostic but is not included in the environment difficulty score.

**Counterpart first-offer model.** When the counterpart must produce its first offer, we generate an initial counterpart price relative to its reservation value. Define the counterpart's directional slack

$$S_B^{\text{open}} := \begin{cases} p_{\max} - r_B, & \text{if the counterpart is a seller,} \\ r_B - p_{\min}, & \text{if the counterpart is a buyer,} \end{cases}$$

and direction

$$s_B := \begin{cases} +1, & \text{if the counterpart is a seller,} \\ -1, & \text{if the counterpart is a buyer.} \end{cases}$$

Thus $s_B(p - r_B) > 0$ means that price $p$ lies in the counterpart's favorable direction. The role-dependent feasible interval for the counterpart's first offer is

$$\mathcal{B}_B := \begin{cases} [r_B, p_{\max}], & \text{if the counterpart is a seller,} \\ [p_{\min}, r_B], & \text{if the counterpart is a buyer.} \end{cases}$$

The counterpart's first offer is

$$p_{\text{init}}^B = \Pi_{\mathcal{B}_B}\left(r_B + s_B d_{0,e} \phi(\kappa_B, \eta_B) S_B^{\text{open}} + \varepsilon_0\right), \qquad \varepsilon_0 \sim \mathcal{N}(0, \sigma_0^2). \tag{15}$$

Projection onto $\mathcal{B}_B$ ensures that the first counterpart offer respects both public price bounds and the counterpart's own reservation value.

The modulation factor is

$$\phi(\kappa_B, \eta_B) = \text{clip}\left(1 - \omega_\kappa \kappa_B + \omega_\eta \mathbf{1}\{\eta_B = \texttt{aggressive}\} - \omega'_\eta \mathbf{1}\{\eta_B = \texttt{conciliatory}\}, \phi_{\min}, \phi_{\max}\right). \tag{16}$$

More urgent counterparts open less aggressively, aggressive counterparts claim more favorable slack, and conciliatory counterparts open closer to reservation. Because $d_{0,e}$ is randomized and hidden, the counterpart's first offer remains informative but is not a point-identifying signal for $r_B$.

## C.5 Cue Generation and Language Realization

We give the complete specification of the hidden cue variables

$$(\tilde{s}_k, \tilde{c}_k) \in \{\texttt{positive}, \texttt{neutral}, \texttt{negative}\} \times \{\texttt{Concede}, \texttt{Hold}, \texttt{Pressure}\}$$

used by the environment-simulated counterpart. These cues are not revealed directly to the evaluated agent; instead, they parameterize the counterpart's natural-language message $m_k^B$ conditional on the already-committed economic action $(d_k^B, p_k^B)$.

**Counterpart concession magnitude.** Let $p_{k-1}^B$ and $p_k^B$ denote the counterpart's two most recent offers, when both exist. We define the normalized concession magnitude

$$C_k^B := \begin{cases} \min\left\{1, \dfrac{|p_k^B - p_{k-1}^B|}{|p_{k-1}^B - r_B| + \varepsilon_c}\right\}, & \text{if } d_k^B = \texttt{Offer} \text{ and a previous counterpart offer exists,} \\ 0, & \text{otherwise,} \end{cases}$$

where $\varepsilon_c > 0$ is a small constant used only for numerical stability. This definition is role-agnostic: for a seller, concessions correspond to lowering price; for a buyer, concessions correspond to increasing price.

**Deadline clock.** We define normalized round progress by

$$D_k := \frac{k}{K},$$

and use the concave deadline clock

$$\widetilde{D}_k := \sqrt{D_k}.$$

For equations written in terms of remaining time, we use

$$\widetilde{\widetilde{D}}_k := 1 - \widetilde{D}_k.$$

The square-root clock spreads deadline pressure over the later part of the interaction instead of concentrating it only at the final round. In the acceptance model, $\widetilde{\widetilde{D}}_k$ replaces the linear remaining-time term; in the strategic cue model, $\widetilde{D}_k$ replaces the linear deadline-proximity term.

**Boundary conditions under mixed opener roles.** Because either party may open the episode, boundary values are defined by the available public history rather than by a fixed round-one pattern. If the counterpart has made fewer than two offers, then no previous counterpart offer exists for computing $C_k^B$, and we set

$$C_k^B = 0.$$

If the agent has made fewer than two offers, then the agent-history features used by the acceptance and counter-offer models are set to their boundary values:

$$\text{ConcedeMagnitude}_k = 0, \qquad \text{ConcedeSpeed}_k = 0, \qquad \text{Rigidity}_k = 0.$$

Thus, when $\chi = \texttt{CounterpartOpens}$, the counterpart's opening message uses $C_k^B = 0$ because no previous counterpart offer exists. When $\chi = \texttt{AgentOpens}$, the counterpart's first response to the agent's opening offer is evaluated with zero counterpart-concession history and zero agent-history reactivity unless the relevant prior offers exist. These boundary values ensure that cue likelihoods, acceptance probabilities, and belief updates are well-defined under both opener roles without special-casing the protocol.

### C.5.1 Strategic Cue Generation

The coarse strategic cue $\tilde{c}_k$ captures the counterpart's current bargaining posture while remaining partially informative about its latent stance type $\eta_B$. We generate $\tilde{c}_k$ from a hybrid model that combines (i) a latent-type prior over postures with (ii) adjustments driven by realized economic action, current concession magnitude, and transformed deadline proximity.

For terminal-style actions, we use deterministic mappings:

$$\tilde{c}_k^{\text{base}} = \begin{cases} \texttt{Concede}, & \text{if } d_k^B = \texttt{Accept}, \\ \texttt{Pressure}, & \text{if } d_k^B = \texttt{Reject}. \end{cases}$$

Here $d_k^B = \texttt{Reject}$ denotes terminal counterpart walk-away.

For offer actions, define a latent-type bias vector over $\{\texttt{Concede}, \texttt{Hold}, \texttt{Pressure}\}$:

$$b(\eta_B) = \begin{cases} (b_C,\, 0,\, -b_C), & \eta_B = \texttt{conciliatory}, \\ (0,\, b_H,\, 0), & \eta_B = \texttt{neutral}, \\ (-b_P,\, 0,\, b_P), & \eta_B = \texttt{aggressive}, \end{cases}$$

where the coordinates are ordered as $(\texttt{Concede}, \texttt{Hold}, \texttt{Pressure})$. Given an offer action $d_k^B = \texttt{Offer}$, we define posture logits

$$\ell_k(\texttt{Concede}) = b_{\eta_B}^{(\texttt{Concede})} + \alpha_C(C_k^B - \tau_{\text{conc}}),$$
$$\ell_k(\texttt{Hold}) = b_{\eta_B}^{(\texttt{Hold})},$$
$$\ell_k(\texttt{Pressure}) = b_{\eta_B}^{(\texttt{Pressure})} + \alpha_P(\widetilde{D}_k - \tau_{\text{dead}}) - \beta_C C_k^B.$$

The base strategic cue is then sampled as

$$\tilde{c}_k^{\text{base}} \mid (d_k^B = \texttt{Offer}, \eta_B, C_k^B, \widetilde{D}_k) \sim \text{Categorical}\left(\text{softmax}(\ell_k)\right).$$

Thus, conciliatory types are biased toward `Concede`, aggressive types toward `Pressure`, and neutral types toward `Hold`, while realized concession and deadline proximity modulate these latent tendencies.

### C.5.2 Sentiment Cue Generation

The affective sentiment cue $\tilde{s}_k$ is sampled from a three-level latent score model whose mean depends on the counterpart's stance type $\eta_B \in \{\texttt{conciliatory}, \texttt{neutral}, \texttt{aggressive}\}$. Let

$$\mu(\eta_B) = \begin{cases} +\mu_s, & \eta_B = \texttt{conciliatory}, \\ 0, & \eta_B = \texttt{neutral}, \\ -\mu_s, & \eta_B = \texttt{aggressive}, \end{cases} \qquad z_k = \mu(\eta_B) + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, \sigma_s^2).$$

We then define

$$\tilde{s}_k^{\text{base}} = \begin{cases} \texttt{positive}, & z_k > \tau_s, \\ \texttt{neutral}, & |z_k| \le \tau_s, \\ \texttt{negative}, & z_k < -\tau_s, \end{cases}$$

where $\tau_s > 0$ is a fixed threshold.

Equivalently, conditional on $\eta_B$, the categorical probabilities are

$$\Pr(\tilde{s}_k^{\text{base}} = \texttt{positive} \mid \eta_B) = 1 - \Phi\left(\frac{\tau_s - \mu(\eta_B)}{\sigma_s}\right),$$
$$\Pr(\tilde{s}_k^{\text{base}} = \texttt{neutral} \mid \eta_B) = \Phi\left(\frac{\tau_s - \mu(\eta_B)}{\sigma_s}\right) - \Phi\left(\frac{-\tau_s - \mu(\eta_B)}{\sigma_s}\right),$$
$$\Pr(\tilde{s}_k^{\text{base}} = \texttt{negative} \mid \eta_B) = \Phi\left(\frac{-\tau_s - \mu(\eta_B)}{\sigma_s}\right),$$

where $\Phi$ is the standard normal CDF. Larger $\sigma_s^2$ weakens the coupling between sentiment and latent stance; smaller $\sigma_s^2$ makes tone more tightly tied to $\eta_B$.

### C.5.3 Family-Specific Cue Parameterization

The constructions above define base sentiment and strategic cues. The six counterpart behavior families are instantiated through simple family-specific cue settings applied to these base cues.

For `Candid` and `Expressive`, we use the base cue model directly:

$$\tilde{s}_k := \tilde{s}_k^{\text{base}}, \qquad \tilde{c}_k := \tilde{c}_k^{\text{base}}.$$

For `Taciturn` and `Strategic`, the cue channel is made uninformative by collapsing to noncommittal middle states:

$$\tilde{s}_k := \texttt{neutral}, \qquad \tilde{c}_k := \texttt{Hold}.$$

This removes direct linguistic evidence of latent stance while leaving the economic channel unchanged.

For `Stochastic`, cues are weakly informative but noisy. We sample sentiment from the same latent-score model with elevated noise $\sigma_{s,\text{stoch}}$ and sample strategic posture using a softened distribution:

$$\tilde{c}_k \sim \text{Categorical}\left(\text{softmax}\left(\frac{\ell_k}{T_{\text{stoch}}}\right)\right), \qquad T_{\text{stoch}} > 1.$$

Unless otherwise stated, we use

$$\sigma_{s,\text{stoch}} = 2.0, \qquad T_{\text{stoch}} = 2.5.$$

This preserves a noisy relationship between cues and latent stance, unlike the collapsed cue channel used by `Taciturn` and `Strategic`.

For `Adversarial`, the cue channel is pressuring:

$$\tilde{s}_k := \texttt{negative}, \qquad \tilde{c}_k := \texttt{Pressure}.$$

This produces a systematically hardball linguistic surface. Because the adversarial family also uses an aggressive-skewed stance prior and hardball economic preset, these cues are often directionally aligned with pressure, but they should not be interpreted as calibrated posterior probabilities over stance.

### C.5.4 Language Realization

Natural-language realizations are generated only after the economic kernel has already committed to $(d_k^B, p_k^B, \tilde{s}_k, \tilde{c}_k)$. The voice layer therefore cannot alter economic outcomes. The language model receives a rendering prompt containing:

- the counterpart role, buyer or seller;
- the fixed economic action $d_k^B$;
- the fixed price $p_k^B$ if $d_k^B = \texttt{Offer}$;
- the sentiment cue $\tilde{s}_k$;
- the strategic cue $\tilde{c}_k$; and
- a short summary of the public history $h_{k-1}$.

It is instructed to produce a brief message that is semantically consistent with the fixed action and price, while expressing the specified tone and posture. In particular:

- `Concede` encourages compromise-oriented phrasing;
- `Hold` encourages firm but non-escalatory phrasing;
- `Pressure` encourages urgency- or deadline-oriented phrasing.

The sentiment cue $\tilde{s}_k$ controls whether that phrasing is polite/constructive (`positive`), matter-of-fact (`neutral`), or tense (`negative`).

**Default simulator hyperparameters and calibration.** Unless otherwise stated, we use

$$\tau_{\text{conc}} = 0.10, \qquad \tau_{\text{dead}} = 0.80, \qquad \mu_s = 1.0, \qquad \tau_s = 0.5, \qquad \sigma_s = 0.75,$$

together with strategic-cue parameters

$$b_C = 1.0, \qquad b_H = 0.5, \qquad b_P = 1.0,$$

$$\alpha_C = 2.0, \qquad \alpha_P = 2.0, \qquad \beta_C = 1.0.$$

These values are chosen so that latent stance and interaction state exert comparable influence on emitted posture: under the base cue model, conciliatory types making visibly concessionary offers are more likely to emit `Concede`, aggressive types late in the negotiation with low concession are more likely to emit `Pressure`, and neutral types under moderate conditions tend to emit `Hold`.

### C.6 Default Parameter Summary

For readability, we compile the default hyperparameters for counterpart simulator in Table 5-6 and note that regime-specific task generation parameters used in experiments are provided in Table 10 in Section H.

We note that the shared counter-offer defaults are deliberately conservative. Under baseline urgency $\kappa_B = 0.5$, no recent agent concession, and before family-specific reciprocity effects, the shared baseline yields $\lambda_0 + \lambda_1\kappa_B = 0.12 + 0.28(0.5) = 0.26$. Thus, absent noise and additional stance-dependent effects, the counterpart retains approximately $(1 - 0.26)^2 = 0.55$ of its initial distance to reservation after two counter-offers and $(1 - 0.26)^3 = 0.41$ after three. This avoids near-saturation from overly fast concession while still allowing urgency and family-specific stance presets to modulate behavior. In particular, stance affects not only concession speed but also whether agent rigidity and recent concession are rewarded or exploited.

## D Oracle-Cue Bayes-Optimal Reference Policy

Because the counterpart policy $\pi_B$ and environment prior $\mu$ are fully specified and fixed, TERMS-BENCH admits computation of a Bayes-optimal benchmark policy $\pi^*$ via backward induction. The negotiation reduces to a partially observable Markov decision process in which the hidden state comprises the counterpart type $t_B = (r_B, \kappa_B, \eta_B)$ together with an episode-level opening-harshness nuisance variable $d_{0,e}$ (Appendix D.3.6). We solve this decision process on a discretized belief space to obtain the reference policy and the associated optimality gap.

In the benchmark's observation model, the agent observes $(p_k^B, m_k^B)$: the counterpart's price and a natural-language message. The sentiment cue $\tilde{s}_k$ and strategic cue $\tilde{c}_k$ are *latent* variables embedded in $m_k^B$ and must be inferred. To define a computable upper bound, we derive the reference policy under an oracle-cue assumption: the agent directly observes the counterpart's economic action $d_k^B \in \{\texttt{Offer}, \texttt{Accept}, \texttt{Reject}\}$, the sentiment cue $\tilde{s}_k$, and the strategic cue $\tilde{c}_k$, together with a counter-offer price $p_k^B$ when $d_k^B = \texttt{Offer}$. The three-way economic action set reflects the three-branch counterpart response model of Section 2.2.2, in which the counterpart may terminate negotiation through a walk-away action in addition to accepting or counter-offering. Any real LLM agent operates under the true observation model with imperfect cue extraction, so the oracle policy provides an upper bound on achievable performance. Throughout this appendix, we specialize to the linear-surplus case $g_b(x) = g_s(x) = x$, so that $u_{\text{buyer}}(p) = r_A - p$ and $u_{\text{seller}}(p) = p - r_A$.

The oracle additionally conditions on the opener role $\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}$, which is sampled once at the start of each episode and is common knowledge. The backward induction is executed over both initial conditions: under $\chi = \texttt{CounterpartOpens}$, the round-1 state already contains a counterpart opening offer $p_1^B$ and the agent's action set at round 1 is $\{\texttt{Accept}, \texttt{Reject}\} \cup \{\texttt{Offer}(p) : p \in \mathcal{P}\}$; under $\chi = \texttt{AgentOpens}$, the round-1 action set is $\{\texttt{Offer}(p) : p \in \mathcal{P}\}$ alone because no counterpart offer has yet been observed.

**Notational convention.** Throughout this appendix, $a_k(p_k^A, t_B, \psi_k)$ (with arguments shown) denotes the counterpart's *acceptance probability* at round $k$ conditional on the agent's offer $p_k^A$, while the plain symbol $a_k$ (without arguments, only in contexts like the belief update below) denotes the agent's round-$k$ action. Where the two could be confused within a single equation, we write the acceptance probability with its full argument list.

### D.1 Belief Discretization

We discretize the type space $\mathcal{T}_B$ into a finite grid:

- Reservation: $\mathcal{R} = \{r_{\min}, r_{\min} + \delta_r, \ldots, r_{\max}\}$ with spacing $\delta_r = 50$;
- Urgency: $\mathcal{K} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ (5 levels);
- Stance: $\mathcal{H} = \{\texttt{C}, \texttt{N}, \texttt{A}\}$ for conciliatory, neutral, and aggressive (3 categories).

This yields $N = |\mathcal{R}| \times 5 \times 3$ discrete types, where $|\mathcal{R}| = (r_{\max} - r_{\min})/\delta_r$. The belief state at round $k$ is a probability vector $b_k \in \Delta(\mathcal{R} \times \mathcal{K} \times \mathcal{H})$, initialized from the environment prior $\mu$.

**Family-specific initial belief.** For all families except ADVERSARIAL, the stance marginal of $\mu$ is uniform on $\mathcal{H}$ and the initial belief $b_0$ inherits this uniform marginal. For ADVERSARIAL, the stance marginal is skewed toward aggressive counterparts,

$$\Pr(\eta_B = \texttt{C}) = 0.05, \quad \Pr(\eta_B = \texttt{N}) = 0.15, \quad \Pr(\eta_B = \texttt{A}) = 0.80, \tag{17}$$

| Component | Parameter | Default | Interpretation |
|---|---|---|---|
| Acceptance model | $\alpha$ | 6.0 | Sensitivity of acceptance to normalized offer favorability $\bar{\Delta}_k$; larger values make acceptance more responsive to how favorable the current offer is for the counterpart. |
| Acceptance model | $\beta$ | 1.0 | Sensitivity of acceptance to counterpart urgency $\kappa_B$; larger values make urgent counterparts more willing to accept. |
| Acceptance model | $\gamma$ | 2.0 | Sensitivity to transformed remaining time $\widetilde{D}_k = 1 - \sqrt{k/K}$; larger values make early-round acceptance less likely while using a concave deadline clock to avoid concentrating deadline pressure only in the final round. |
| Acceptance model | $\rho_{\mathcal{F}}(\eta_B)$ | family-specific | Stance-dependent sensitivity to $\mathrm{ConcedeSpeed}_k$; values are specified by the family economic preset in Table 4. |
| Acceptance model | $\xi_{\mathcal{F}}(\eta_B)$ | family-specific | Stance-dependent sensitivity to $\mathrm{Rigidity}_k$; values are specified by the family economic preset in Table 4. |
| Walk-away model | $k_{\mathrm{walk}}$ | $\lceil K/2 \rceil$ | First agent-response round in which terminal counterpart exit is enabled. This gives the agent a grace period to infer infeasibility and reject before the counterpart can preemptively exit. |
| Walk-away model | $\phi_0$ | $-4.5$ | Intercept of the walk-away hazard. The negative value makes walk-away rare unless the current offer is outside the counterpart's reservation constraint. |
| Walk-away model | $\phi_\Delta$ | 30.0 | Sensitivity to normalized reservation shortfall $[-\bar{\Delta}_k]_+$. Larger values make strongly non-individually-rational offers more likely to trigger terminal exit. |
| Walk-away model | $\phi_T$ | 1.5 | Sensitivity to the walk-away clock $\tau_k^W$. Larger values make persistent infeasible offers more likely to trigger terminal exit as the round limit approaches. |
| Counter-offer model | $\lambda_0$ | 0.12 | Baseline latent concession tendency in the unconstrained concession score $\tilde{\lambda}_B(h_{k-1})$. This lower default avoids overly rapid convergence to the counterpart's reservation value. |
| Counter-offer model | $\lambda_1$ | 0.28 | Urgency sensitivity in the latent concession score. At baseline urgency $\kappa_B = 0.5$, a neutral counterpart with no agent concession has $\tilde{\lambda}_B \approx 0.26$. |
| Counter-offer model | $\lambda_{2,\mathcal{F}}(\eta_B)$ | family-specific | Stance-dependent reciprocity sensitivity to the agent's recent role-normalized concession magnitude. Larger values make the counterpart hold firmer against fast-conceding agents. Values are specified by the family economic preset in Table 4. |
| Counter-offer model | $\lambda_3,\ \lambda_4$ | 0.10, 0.10 | Stance adjustments in the concession score. $\lambda_3$ slows aggressive counterparts; $\lambda_4$ accelerates conciliatory counterparts. |
| Counter-offer model | $\bar{\sigma}_p^{\mathrm{low}},\ \bar{\sigma}_p^{\mathrm{mod}},\ \bar{\sigma}_p^{\mathrm{high}}$ | 0.01, 0.03, 0.08 | Normalized additive price-noise scale, where $\bar{\sigma}_p := \sigma_p/(p_{\max} - p_{\min})$. Actual offer noise is $\sigma_p = \bar{\sigma}_p(p_{\max} - p_{\min})$. |
| Counter-offer model | clipping of $\lambda_B$ | $[0, 1]$ | Effective concession rate is clipped to $[0, 1]$ so that the deterministic price update moves weakly toward reservation and never anti-concedes. |
| Counter-offer model | offer projection | $\mathcal{M}_B(k)$ | Subsequent counter-offers are projected onto the dynamic monotone feasible interval $\mathcal{M}_B(k) = [r_B, p_{k-1}^B]$ for seller counterparts and $\mathcal{M}_B(k) = [p_{k-1}^B, r_B]$ for buyer counterparts. This enforces both counterpart individual rationality and weak monotonicity; if noise would reverse the concession direction, the counterpart holds at its previous offer. |
| History features | $\tau_{\mathrm{rigid}}$ | 0.10 | Threshold for the rigidity indicator. If the agent's most recent role-normalized concession is below this level, the agent is treated as rigid. |

Table 5: Default economic-response and price-dynamics hyperparameters used in TERMS-BENCH's bilateral price-negotiation instantiation. Parameters are specified in normalized state space whenever possible. The revised counter-offer defaults slow baseline concession relative to earlier settings, while the walk-away hazard introduces terminal counterpart exit only for offers outside the counterpart's individually rational region.

and $b_0$ is set accordingly; the reservation and urgency marginals remain as in other families.

**Opening harshness as a nuisance variable.** Opening harshness $d_{0,e} \sim \mathrm{Uniform}(d_{\min}, d_{\max})$ (defaults $d_{\min} = 0.20$, $d_{\max} = 0.80$) is realized once per episode and influences only the counterpart's opening offer under $\chi = \texttt{CounterpartOpens}$, or the counterpart's round-1 response under $\chi = \texttt{AgentOpens}$ when that response is `Offer`. Because $d_{0,e}$ does not enter any subsequent-round dynamics, we do not expand the belief

| Component | Parameter | Default | Interpretation |
|---|---|---|---|
| Opening-offer model | $d_{0,e}$ | $\mathrm{Unif}(0.20, 0.80)$ | Episode-level opening harshness. Randomizing $d_{0,e}$ prevents the counterpart's first offer from being nearly invertible into its reservation value. Controlled ablations may use narrower stratum-specific intervals. |
| Opening-offer model | $\omega_\kappa$ | 0.30 | Urgency-discount coefficient in the opening-offer modulation factor $\phi(\kappa_B, \eta_B)$; larger values make more urgent counterparts open less aggressively. |
| Opening-offer model | $\omega_\eta$, $\omega'_\eta$ | 0.15, 0.15 | Stance modulation coefficients in $\phi(\kappa_B, \eta_B)$. $\omega_\eta$ makes aggressive counterparts claim more favorable slack; $\omega'_\eta$ makes conciliatory counterparts open closer to reservation. |
| Opening-offer model | $\phi_{\min}$, $\phi_{\max}$ | 0.5, 1.5 | Clipping range for the opening-offer modulation factor. Ensures the type/urgency adjustment remains positive and bounded. |
| Opening-offer model | $\bar{\sigma}_0$ | 0.02 | Normalized opening-offer noise scale, where $\sigma_0 = \bar{\sigma}_0(p_{\max} - p_{\min})$. |
| Strategic cue model | $\tau_{\text{dead}}$ | 0.80 | Threshold applied to transformed deadline proximity $\widetilde{D}_k = \sqrt{k/K}$ in the strategic-cue logits. |
| Strategic cue model | $\alpha_P$ | 2.0 | Sensitivity of the `Pressure` logit to transformed deadline proximity $\widetilde{D}_k$. |
| Strategic cue model | $b_C$, $b_H$, $b_P$ | 1.0, 0.5, 1.0 | Baseline latent-type bias toward `Concede`, `Hold`, and `Pressure`, respectively. These determine how strongly $\eta_B$ influences the strategic cue absent strong state signals. |
| Strategic cue model | $\alpha_C$, $\alpha_P$, $\beta_C$ | 2.0, 2.0, 1.0 | State-sensitivity coefficients. $\alpha_C$ increases the `Concede` logit after realized counterpart concession; $\alpha_P$ increases `Pressure` near the deadline; $\beta_C$ suppresses `Pressure` after visible concession. |
| Sentiment cue model | $\mu_s$, $\tau_s$, $\sigma_s$ | 1.0, 0.5, 0.75 | Sentiment-generation parameters. $\mu_s$ controls separation between conciliatory and aggressive sentiment means, $\tau_s$ discretizes latent sentiment, and $\sigma_s$ controls baseline sentiment noise. |
| Family-specific cue settings | Candid / Expressive | base cue model | Accurate cue channel: sentiment and strategic posture are sampled from the base cue model and remain informative about latent stance. |
| Family-specific cue settings | Taciturn / Strategic | $\tilde{s}_k$ = `neutral`, $\tilde{c}_k$ = `Hold` | Uninformative cue channel: sentiment and strategic posture are collapsed to noncommittal middle states. |
| Family-specific cue settings | $\sigma_{s,\text{stoch}}$, $T_{\text{stoch}}$ | 2.0, 2.5 | Noisy cue channel for the `Stochastic` family: elevated sentiment noise and strategic-cue temperature produce weak but non-collapsed cues. |
| Family-specific cue settings | Adversarial | $\tilde{s}_k$ = `negative`, $\tilde{c}_k$ = `Pressure` | Pressuring cue channel: adversarial counterparts use systematically aggressive sentiment and pressure-oriented posture. |

Table 6: Default opening-offer and cue-generation hyperparameters used in TERMS-BENCH's bilateral price-negotiation. Opening-offer parameters determine the counterpart's initial anchor, while cue-generation parameters control the informativeness and tone of the language-facing signal channel.

over $\mathcal{T}_B$ to include it; instead, we marginalize $d_{0,e}$ analytically in the round-1 likelihood (Appendix D.3.6). An alternative implementation expands $\mathcal{T}_B$ by a coarse $d_0$-grid (for example, five levels), increasing $N$ by a factor of five; we adopt the marginalization approach because $d_{0,e}$ is information-free after round 1 and further posterior refinement over it has no downstream value.

## D.2 Augmented Information State

The belief $b_k$ alone is not a sufficient statistic for the agent's decision problem. The counterpart model depends on additional known quantities:

1. **Opener role** $\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}$, assigned at episode start and constant across rounds.
2. **History-reactive features** $\phi_k := (\mathrm{ConcedeSpeed}_k, \mathrm{Rigidity}_k, \mathrm{ConcedeMagnitude}_k)$, which are deterministic functions of the agent's past offer sequence (Appendix C.3) and parameterize the counterpart's acceptance probability, walk-away hazard, and concession rate.
3. **Offer-history summary** $h_k^B := (p_k^B, p_{k-1}^B)$, which records the counterpart's current and previous offers. These are needed to evaluate acceptance utility $u_A(p_k^B)$, to compute the price-likelihood mean (28), to derive the counterpart's concession magnitude $C_k^B$ used in the strategic-cue model, and to specify the role-dependent monotone feasible interval $\mathcal{M}_B(k)$ onto which counter-offers are projected.

Additionally, the monotonic concession constraint and the update of $\phi_{k+1}$ require the agent's recent own-offer history. We store this as $\xi_k^A := (p_{k-3}^A, p_{k-2}^A, p_{k-1}^A)$, with entries set to $\varnothing$ before they exist. All components besides $b_k$ are deterministic given the public history. We define the augmented information state

$$\psi_k \;:=\; (b_k,\, \phi_k,\, \xi_k^A,\, h_k^B,\, \chi). \tag{18}$$

All value functions, $Q$-functions, likelihoods, and state transitions below are defined over $\psi_k$.

**Boundary values under mixed opener roles.** History features and the offer-history summary adopt boundary values whenever the requisite offers have not yet been realized. If the agent has made fewer than two own offers, then $\text{ConcedeSpeed}_k = \text{Rigidity}_k = \text{ConcedeMagnitude}_k = 0$. If the counterpart has produced fewer than two offers, the entries of $h_k^B$ that do not yet exist are set to $\varnothing$ and $C_k^B = 0$. Under $\chi = \texttt{AgentOpens}$, the counterpart has made no offer prior to the agent's round-1 action, so these boundary values additionally extend into round 2 whenever the counterpart responded at round 1 with `Offer` (yielding $p_1^B$) but no earlier counterpart offer exists. This is a natural extension of the round-1 boundary handling and does not require special-casing in the DP.

## D.3 Observation Likelihood

Under the oracle-cue assumption, the agent observes the counterpart's economic action $d_k^B \in \{\texttt{Offer}, \texttt{Accept}, \texttt{Reject}\}$, the sentiment cue $\tilde{s}_k$, and the strategic cue $\tilde{c}_k$, together with a counter-offer price $p_k^B$ when $d_k^B = \texttt{Offer}$. The observation likelihood decomposes according to the counterpart's sequential generative process: the economic action is drawn first; if $d_k^B = \texttt{Offer}$, the price is drawn next; then the sentiment cue is drawn independently from the stance type; and finally the strategic cue is drawn conditional on the realized economic action and price.

### D.3.1 Counterpart Action Distribution

Given the agent's most recent offer $p_k^A$, the counterpart first decides whether to accept. The acceptance probability is (Eq. 5–6):

$$a_k(p_k^A, t_B, \psi_k) \;:=\; \mathbf{1}\{\bar{\Delta}_k \geq 0\}\, \sigma\big(g_\theta(p_k^A, t_B, k, \phi_k)\big)\,, \tag{19}$$

where $\sigma(\cdot)$ is the logistic function and

$$g_\theta \;=\; \alpha\, \bar{\Delta}_k \;+\; \beta\, \kappa_B \;-\; \gamma\, \bar{\bar{D}}_k \;+\; \rho_F(\eta_B)\, \text{ConcedeSpeed}_k \;+\; \xi_F(\eta_B)\, \text{Rigidity}_k, \tag{20}$$

with the normalized quantities

$$\bar{\Delta}_k := \begin{cases} \dfrac{p_k^A - r_B}{p_{\max} - p_{\min}}, & \text{seller counterpart,} \\[2ex] \dfrac{r_B - p_k^A}{p_{\max} - p_{\min}}, & \text{buyer counterpart,} \end{cases} \qquad \bar{\bar{D}}_k \;:=\; 1 - \sqrt{k/K}.$$

The concave remaining-time term $\bar{\bar{D}}_k$ replaces the previous linear clock; the default deadline weight is correspondingly raised to $\gamma = 2.0$ to partially compensate for the concavity. The coefficients $\rho_F(\eta_B)$ and $\xi_F(\eta_B)$ are *stance- and family-dependent* (Appendix C.1), so the payoff consequences of conceding quickly and of holding firm differ across the three stance types within a fixed family.

Conditional on non-acceptance, the counterpart either terminates through a walk-away or produces a counter-offer. The walk-away hazard is

$$\omega_k(p_k^A, t_B, \psi_k) \;:=\; \mathbf{1}\{k \geq k_{\text{walk}}\}\, \mathbf{1}\{\bar{\Delta}_k < 0\}\, \sigma\big(\phi_0 + \phi_\Delta\, [-\bar{\Delta}_k]_+ + \phi_T\, \tau_k^W\big)\,, \tag{21}$$

where $\tau_k^W = \min\{1, \max\{0,\, (k - k_{\text{walk}})/(K - k_{\text{walk}})\}\}$ and the defaults are $k_{\text{walk}} = \lceil K/2 \rceil$, $\phi_0 = -4.5$, $\phi_\Delta = 30.0$, $\phi_T = 1.5$. The hazard is nonzero only after a grace period and only when the current offer is outside the counterpart's individually rational region.

For $k < K$, and writing $a_k \equiv a_k(p_k^A, t_B, \psi_k)$ and $\omega_k \equiv \omega_k(p_k^A, t_B, \psi_k)$ for brevity, the counterpart action probabilities are

$$P(d_k^B = \texttt{Accept} \mid t_B, \psi_k, p_k^A) = a_k, \tag{22}$$

$$P(d_k^B = \texttt{Reject} \mid t_B, \psi_k, p_k^A) = (1 - a_k)\, \omega_k, \tag{23}$$

$$P(d_k^B = \texttt{Offer} \mid t_B, \psi_k, p_k^A) = (1 - a_k)(1 - \omega_k). \tag{24}$$

At $k = K$, any remaining non-acceptance and non-walk-away mass results in round-limit disagreement.

**Hard support constraint from `Reject` observations.** Because $\omega_k$ carries the indicator $\mathbf{1}\{\bar{\Delta}_k < 0\}$, observing $d_k^B = \texttt{Reject}$ at any round $k \geq k_{\text{walk}}$ is a *hard constraint* on $t_B$: it can only occur for types whose reservation value $r_B$ makes the agent's current offer non-individually-rational. The Bayesian update following a `Reject` observation therefore assigns posterior mass zero to every type with $\bar{\Delta}_k \geq 0$, eliminating a contiguous region of the belief support.

### D.3.2 Counter-Offer Price Likelihood

If $d_k^B = \texttt{Offer}$, the counterpart's price is generated by Eq. (8)-(9):

$$p_k^B \;=\; \Pi_{\mathcal{M}_B(k)}\big[\, p_{k-1}^B - \lambda_B(\phi_k;\, t_B)\cdot(p_{k-1}^B - r_B) + \varepsilon_k \,\big], \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_p^2), \tag{25}$$

where the effective concession rate is

$$\lambda_B(\phi_k;\, t_B) \;=\; \text{clip}_{[0,1]}\Big[\lambda_0 + \lambda_1\,\kappa_B - \lambda_{2,F}(\eta_B)\,\text{ConcedeMagnitude}_k - \lambda_3\,\mathbf{1}\{\eta_B = \texttt{A}\} + \lambda_4\,\mathbf{1}\{\eta_B = \texttt{C}\}\Big], \tag{26}$$

with revised defaults $\lambda_0 = 0.12$, $\lambda_1 = 0.28$, and stance-dependent $\lambda_{2,F}(\eta_B)$ per the family preset. The projection is onto the role-dependent monotone feasible interval

$$\mathcal{M}_B(k) \;:=\; \begin{cases} [r_B,\, p_{k-1}^B], & \text{seller counterpart,} \\ [p_{k-1}^B,\, r_B], & \text{buyer counterpart,} \end{cases} \tag{27}$$

rather than onto the public bounds $[p_{\min}, p_{\max}]$. Define the deterministic pre-projection mean

$$\bar{p}_k^B(t_B, \psi_k) \;:=\; p_{k-1}^B - \lambda_B(\phi_k;\, t_B)\cdot(p_{k-1}^B - r_B), \tag{28}$$

so that the unprojected offer is $\tilde{p}_k^B \sim \mathcal{N}(\bar{p}_k^B,\, \sigma_p^2)$ and the observed price $p_k^B = \Pi_{\mathcal{M}_B(k)}(\tilde{p}_k^B)$. Writing $\mathcal{M}_B(k) = [a, b]$ with

$$a \;=\; \begin{cases} r_B, & \text{seller,} \\ p_{k-1}^B, & \text{buyer,} \end{cases} \qquad b \;=\; \begin{cases} p_{k-1}^B, & \text{seller,} \\ r_B, & \text{buyer,} \end{cases}$$

the price likelihood is a mixed distribution with point masses at $a$ and $b$:

$$f_{\text{price}}(p_k^B \mid t_B, \psi_k) \;=\; \begin{cases} \Phi\!\left(\dfrac{a - \bar{p}_k^B}{\sigma_p}\right), & p_k^B = a, \\[2ex] \dfrac{1}{\sigma_p}\,\phi\!\left(\dfrac{p_k^B - \bar{p}_k^B}{\sigma_p}\right), & a < p_k^B < b, \\[2ex] 1 - \Phi\!\left(\dfrac{b - \bar{p}_k^B}{\sigma_p}\right), & p_k^B = b, \end{cases} \tag{29}$$

where $\phi$ and $\Phi$ denote the standard normal pdf and cdf, respectively.

**Informativeness of the reservation-endpoint point mass.** The lower endpoint $a$ of $\mathcal{M}_B(k)$ is $r_B$ for a seller counterpart; symmetrically, the upper endpoint $b$ is $r_B$ for a buyer counterpart. Observing $p_k^B$ pinned at the reservation endpoint is therefore highly informative about the hidden reservation value: it indicates that the Gaussian draw was clamped against the counterpart's own reservation constraint, revealing that the counterpart has conceded to the boundary of its feasible region. The other endpoint of $\mathcal{M}_B(k)$ is the previous counterpart offer $p_{k-1}^B$, which is already known from $h_k^B$; a draw pinned there signals that noise attempted to reverse the concession direction and carries information about $\kappa_B$ and $\eta_B$ through $\lambda_B(\phi_k; t_B)$ but not directly about $r_B$. The oracle's posterior over $r_B$ should therefore tighten substantially whenever a counter-offer is pinned at the reservation endpoint, particularly in later rounds when multiple such pinnings accumulate.

### D.3.3 Sentiment Cue Likelihood

The sentiment cue $\tilde{s}_k$ depends only on the stance type $\eta_B$ and is conditionally independent of the realized price. From the latent-score model in Appendix C.5:

$$\begin{aligned} P(\tilde{s}_k = \texttt{pos} \mid \eta_B) &= 1 - \Phi\!\left(\frac{\tau_s - \mu(\eta_B)}{\sigma_s}\right), \\ P(\tilde{s}_k = \texttt{neu} \mid \eta_B) &= \Phi\!\left(\frac{\tau_s - \mu(\eta_B)}{\sigma_s}\right) - \Phi\!\left(\frac{-\tau_s - \mu(\eta_B)}{\sigma_s}\right), \\ P(\tilde{s}_k = \texttt{neg} \mid \eta_B) &= \Phi\!\left(\frac{-\tau_s - \mu(\eta_B)}{\sigma_s}\right), \end{aligned} \tag{30}$$

where $\mu(\texttt{C}) = +\mu_s$, $\mu(\texttt{N}) = 0$, $\mu(\texttt{A}) = -\mu_s$.

### D.3.4 Strategic Cue Likelihood

For terminal counterpart actions, the mapping is deterministic:

$$\tilde{c}_k \;=\; \begin{cases} \texttt{Concede}, & d_k^B = \texttt{Accept}, \\ \texttt{Pressure}, & d_k^B = \texttt{Reject}. \end{cases} \tag{31}$$

For offer actions, the cue is sampled from $\mathrm{Categorical}(\mathrm{softmax}(\ell_k))$ with the concave deadline clock $\tilde{D}_k := \sqrt{k/K}$:

$$\begin{aligned} \ell_k(\texttt{Concede}) &= b_{\eta_B}^{(\texttt{C})} + \alpha_C\,(C_k^B - \tau_{\mathrm{conc}}), \\ \ell_k(\texttt{Hold}) &= b_{\eta_B}^{(\texttt{H})}, \\ \ell_k(\texttt{Pressure}) &= b_{\eta_B}^{(\texttt{P})} + \alpha_P\,(\tilde{D}_k - \tau_{\mathrm{dead}}) - \beta_C\,C_k^B, \end{aligned} \tag{32}$$

where $b(\eta_B)$ are the stance-dependent bias vectors from Appendix C.5 and

$$C_k^B \;=\; \min\!\left(1,\; \frac{|p_k^B - p_{k-1}^B|}{|p_{k-1}^B - r_B| + \epsilon_c}\right) \tag{33}$$

is the realized concession magnitude, computed from the offer pair $(p_k^B, p_{k-1}^B) \in h_k^B$ when both offers exist; otherwise $C_k^B = 0$ by the boundary convention of Appendix D.2. We define the strategic-cue likelihood as a function of the stance type, deadline clock, and concession magnitude:

$$P_{\mathrm{strat}}(\tilde{c}_k = c \mid \eta_B, \tilde{D}_k, C_k^B) \;:=\; \frac{\exp\big(\ell_k(c;\,\eta_B, \tilde{D}_k, C_k^B)\big)}{\sum_{c'} \exp\big(\ell_k(c';\,\eta_B, \tilde{D}_k, C_k^B)\big)}. \tag{34}$$

In the generic round-$k$ case (rounds $k \geq 2$ with two counterpart offers available), $C_k^B$ is computed from $(p_k^B, p_{k-1}^B) \in h_k^B$, so we also write $P_{\mathrm{strat}}(\tilde{c}_k \mid t_B, h_k^B) \equiv P_{\mathrm{strat}}(\tilde{c}_k \mid \eta_B, \tilde{D}_k, C_k^B(h_k^B))$ as convenient shorthand. Family-specific overrides apply: STOCHASTIC replaces $\sigma_s$ by $\sigma_{s,\mathrm{stoch}} = 2.0$ in (30) and uses softmax temperature $T_{\mathrm{stoch}} = 2.5$ in (34); TACITURN and STRATEGIC collapse cues to $\tilde{s}_k = \texttt{neu}$, $\tilde{c}_k = \texttt{Hold}$; ADVERSARIAL collapses cues to $\tilde{s}_k = \texttt{neg}$, $\tilde{c}_k = \texttt{Pressure}$. The remaining families use the base cue model above.

### D.3.5 Full Observation Likelihood

The generative process is sequential: draw the economic action, then (if `Offer`) draw the price, then draw sentiment (independently), then draw the strategic cue (deterministic for `Accept`/`Reject`; conditional on the realized price for `Offer`). Using the shorthand $a_k \equiv a_k(p_k^A, t_B, \psi_k)$ and $\omega_k \equiv \omega_k(p_k^A, t_B, \psi_k)$ for the counterpart acceptance probability and walk-away hazard, the full likelihood for each observation type is:

**Counter-offer observation ($d_k^B = \texttt{Offer}$):**

$$P(o_k \mid t_B, \psi_k, p_k^A) \;=\; \underbrace{(1-a_k)(1-\omega_k)}_{\text{offer branch}} \cdot \underbrace{f_{\mathrm{price}}(p_k^B \mid t_B, \psi_k)}_{\text{price}} \cdot \underbrace{P(\tilde{s}_k \mid \eta_B)}_{\text{sentiment}} \cdot \underbrace{P_{\mathrm{strat}}(\tilde{c}_k \mid \eta_B, \tilde{D}_k, C_k^B)}_{\text{strategic cue}}. \tag{35}$$

**Acceptance observation ($d_k^B = \texttt{Accept}$):**

$$P(o_k \mid t_B, \psi_k, p_k^A) \;=\; a_k \;\cdot\; P(\tilde{s}_k \mid \eta_B) \;\cdot\; \mathbf{1}\{\tilde{c}_k = \texttt{Concede}\}. \tag{36}$$

**Walk-away observation ($d_k^B = \texttt{Reject}$):**

$$P(o_k \mid t_B, \psi_k, p_k^A) \;=\; (1-a_k)\,\omega_k \;\cdot\; P(\tilde{s}_k \mid \eta_B) \;\cdot\; \mathbf{1}\{\tilde{c}_k = \texttt{Pressure}\}. \tag{37}$$

Because $\omega_k$ contains $\mathbf{1}\{k \geq k_{\mathrm{walk}}\}\mathbf{1}\{\bar{\Delta}_k < 0\}$, (37) is zero for every type with $\bar{\Delta}_k \geq 0$, inducing the hard support constraint of Section D.3.1.

### D.3.6 Opening-Round Likelihood

The round-1 observation depends on the opener role. In both cases the counterpart's emissions at round 1 include cues, so the round-1 likelihood carries the same sentiment and strategic-cue factors as in later rounds; only the *price* factor is replaced by the opening-offer model when the counterpart produces its first offer, and the hidden opening harshness $d_{0,e}$ is analytically marginalized. Because the round-1 strategic cue uses the boundary value $C_1^B = 0$ (no previous counterpart offer), it depends only on $\eta_B$ and the deadline clock $\tilde{D}_1 = \sqrt{1/K}$, and we write it in the explicit form $P_{\text{strat}}(\tilde{c}_1 \mid \eta_B, \tilde{D}_1, C_1^B{=}0)$ to emphasize that $p_1^B$ itself does not enter the cue logit.

**Opening-offer price model.** When the counterpart produces its first offer $p_1^B$ (either as an opening under $\chi = \texttt{CounterpartOpens}$, or as a response to the agent's opening under $\chi = \texttt{AgentOpens}$), the price is generated by the opening-offer model of Appendix C.4:

$$p_1^B \;=\; \Pi_{\mathcal{B}_B}\big[r_B + s_B\, d_{0,e}\, \varphi(\kappa_B, \eta_B)\, \Delta_B^{\text{slack}} + \varepsilon_0\big], \quad \varepsilon_0 \sim \mathcal{N}(0,\, \sigma_0^2), \tag{38}$$

where $s_B = +1$ for a seller counterpart, $s_B = -1$ for a buyer; $\mathcal{B}_B = [r_B, p_{\max}]$ for a seller and $[p_{\min}, r_B]$ for a buyer; $\Delta_B^{\text{slack}}$ is the counterpart's directional slack; and $\varphi(\kappa_B, \eta_B)$ is the modulation factor. Conditional on $(t_B, d_{0,e})$, the unprojected opening is Gaussian with mean $\mu_0(t_B, d) := r_B + s_B\, d\, \varphi(\kappa_B, \eta_B)\, \Delta_B^{\text{slack}}$ and variance $\sigma_0^2$. Projection onto $\mathcal{B}_B = [a_0, b_0]$ produces a mixed distribution with point masses at the two endpoints:

$$f_{\text{open}}(p_1^B \mid t_B, d_{0,e}) \;=\; \begin{cases} \Phi\left(\dfrac{a_0 - \mu_0(t_B, d_{0,e})}{\sigma_0}\right), & p_1^B = a_0, \\[2ex] \dfrac{1}{\sigma_0}\,\phi\left(\dfrac{p_1^B - \mu_0(t_B, d_{0,e})}{\sigma_0}\right), & a_0 < p_1^B < b_0, \\[2ex] 1 - \Phi\left(\dfrac{b_0 - \mu_0(t_B, d_{0,e})}{\sigma_0}\right), & p_1^B = b_0, \end{cases} \tag{39}$$

with $(a_0, b_0) = (r_B, p_{\max})$ for a seller counterpart and $(p_{\min}, r_B)$ for a buyer. In both cases one of the point masses sits at $r_B$, so the opening likelihood is reservation-informative whenever the noise clamps the raw opening against the counterpart's reservation constraint. Because $d_{0,e}$ is hidden, the oracle uses the marginal likelihood

$$P_{\text{open}}(p_1^B \mid t_B) \;:=\; \frac{1}{d_{\max} - d_{\min}} \int_{d_{\min}}^{d_{\max}} f_{\text{open}}(p_1^B \mid t_B, d)\, \mathrm{d}d, \tag{40}$$

computed numerically on a fixed quadrature grid $\{d^{(1)}, \ldots, d^{(L)}\}$ (we use $L = 9$ composite-Simpson nodes).

**`CounterpartOpens`.** The counterpart opens with $p_1^B$ together with cues $(\tilde{s}_1, \tilde{c}_1)$, before the agent has acted. Because the sentiment and strategic-cue likelihoods are conditionally independent of $d_{0,e}$ given $(t_B, p_1^B)$, the full round-1 likelihood factorizes as

$$P(o_1 \mid t_B,\, \chi{=}\texttt{CounterpartOpens}) \;=\; \underbrace{P_{\text{open}}(p_1^B \mid t_B)}_{\text{opening price, } d_{0,e}\text{-marg.}} \cdot \underbrace{P(\tilde{s}_1 \mid \eta_B)}_{\text{sentiment}} \cdot \underbrace{P_{\text{strat}}(\tilde{c}_1 \mid \eta_B,\, \tilde{D}_1,\, C_1^B{=}0)}_{\text{strategic cue, boundary}}. \tag{41}$$

Family-specific cue overrides from Appendix C.5 apply. Because $d_{0,e}$ does not enter any subsequent-round dynamics, no posterior over it need be propagated; the round-1 belief update (42) uses (41) directly.

**`AgentOpens`.** No counterpart observation precedes the agent's round-1 action, so $b_1 = b_0$ and the agent chooses $a_1$ under the prior. The counterpart's response is emitted *within* round 1 (conditional on the agent's round-1 offer $p_1^A$, with history-reactive features at their zero boundary values) and is consumed by the $b_2$ update; thus $p_1^B$ under $\chi = \texttt{AgentOpens}$ denotes the counterpart's first and only round-1 offer, occurring after the agent's action rather than before it. The response follows the three-branch distribution (22)–(24), and the corresponding observation likelihoods are (36) for `Accept`, (37) for walk-away, and the following three-factor analogue of (35) for `Offer`, with the opening-offer price model replacing the counter-offer price model and $p_1^A$ conditioning the acceptance and walk-away probabilities:

$$\begin{aligned} P(o_1 \mid t_B, \psi_1, p_1^A) = {} & \big[1 - a_1(p_1^A, t_B, \psi_1)\big]\big[1 - \omega_1(p_1^A, t_B, \psi_1)\big] \\ & \cdot P_{\text{open}}(p_1^B \mid t_B)\, P(\tilde{s}_1 \mid \eta_B)\, P_{\text{strat}}(\tilde{c}_1 \mid \eta_B, \tilde{D}_1, C_1^B = 0). \end{aligned} \tag{55}$$

Here we condition on $p_1^A$ rather than on the full agent action $a_1$ to avoid the notational clash with the counterpart acceptance probability $a_1(\cdot)$; since $d_1 = \texttt{Offer}$ is forced under $\chi = \texttt{AgentOpens}$ at round 1, $p_1^A$ fully identifies the agent's action. Both $a_1(p_1^A, t_B, \psi_1)$ and $\omega_1(p_1^A, t_B, \psi_1)$ are evaluated with the zero-boundary agent history in $\psi_1$.

## D.4 Belief Update

Given prior belief $b_k$ and observation $o_{k+1}$ after the agent takes action $a_k$ with offer price $p_k^A$ (when $d_k = \texttt{Offer}$), the posterior follows from Bayes' rule:

$$b_{k+1}(t_B) \;=\; \frac{b_k(t_B) \;\cdot\; P(o_{k+1} \mid t_B, \psi_k, p_k^A)}{\displaystyle\sum_{t'_B \in \mathcal{T}_B} b_k(t'_B) \;\cdot\; P(o_{k+1} \mid t'_B, \psi_k, p_k^A)}, \tag{42}$$

where the observation likelihood is (35), (36), or (37) according to the counterpart's action in rounds $k \geq 2$. The round-1 case is handled separately:

- Under $\chi = \texttt{CounterpartOpens}$, $b_1 = \mathrm{BayesUpdate}(b_0, o_1)$ uses the full opening likelihood (41) (no agent action is required since the counterpart emits first). The agent then chooses its round-1 action from $b_1$.
- Under $\chi = \texttt{AgentOpens}$, $b_1 = b_0$ (no observation precedes the agent's action), and the first nontrivial update is $b_2 = \mathrm{BayesUpdate}(b_1, o_1, p_1^A)$, using (36) if $d_1^B = \texttt{Accept}$, (37) if $d_1^B = \texttt{Reject}$, or (55) if $d_1^B = \texttt{Offer}$.

Conditioning on $p_k^A$ is necessary because the counterpart's acceptance probability, walk-away hazard, and response distribution all depend on the agent's offered price. The components $\phi_k$, $\xi_k^A$, $h_k^B$, and $\chi$ of $\psi_k$ are deterministic given the public history and do not require Bayesian updating.

**Per-stance likelihood evaluation.** Because $\rho_F(\eta_B)$, $\xi_F(\eta_B)$, and $\lambda_{2,F}(\eta_B)$ all depend on stance, the acceptance score $g_\theta$, the walk-away hazard $\omega_k$, and the concession mean $\bar{p}_k^B$ must be evaluated separately for each of $\eta_B \in \{\texttt{C}, \texttt{N}, \texttt{A}\}$ when computing the per-type likelihoods. The belief support does not gain any new dimensions, but the per-$(r_B, \kappa_B)$ evaluation cost scales by the three stance branches. Because stance-dependent coefficients produce distinct acceptance probabilities and distinct concession rates for the same $(r_B, \kappa_B)$, the likelihood ratios across stances are sharper under the new formulation than under uniform coefficients, which should accelerate posterior concentration on $\eta_B$ in cue-informative families.

## D.5 Value Function and Bellman Equation

All value functions are defined over the augmented state $\psi_k = (b_k, \phi_k, \xi_k^A, h_k^B, \chi)$.

**Terminal condition.** At round $K$, if no agreement has been reached, the outcome is disagreement with utility 0:

$$V_{K+1}^*(\psi) \;=\; 0 \qquad \forall\, \psi. \tag{43}$$

**Action set.** The agent's action set depends on what it has observed:

$$\mathcal{A}_k(\psi_k) \;=\; \begin{cases} \{\texttt{Offer}(p) : p \in \mathcal{P}\}, & k = 1 \text{ and } \chi = \texttt{AgentOpens}, \\ \{\texttt{Accept}, \texttt{Reject}\} \cup \{\texttt{Offer}(p) : p \in \mathcal{P}\}, & \text{otherwise (a counterpart offer exists in } h_k^B), \end{cases} \tag{44}$$

where $\mathcal{P} \subseteq [p_{\min}, p_{\max}]$ is the discretized admissible price set. The value function satisfies

$$V_k^*(\psi_k) \;=\; \max_{a \,\in\, \mathcal{A}_k(\psi_k)} Q_k(\psi_k, a), \tag{45}$$

with $Q$-functions as follows.

**Accept.** The agent accepts the counterpart's current offer $p_k^B \in h_k^B$:

$$Q_k(\psi_k, \texttt{Accept}) \;=\; u_A(p_k^B). \tag{46}$$

Since $p_k^B$ is in $h_k^B$, no expectation over $t_B$ is needed. Individual rationality requires $u_A(p_k^B) \geq 0$.

**Reject.** The agent exits the negotiation:

$$Q_k(\psi_k, \texttt{Reject}) \;=\; 0. \tag{47}$$

**Offer.** The agent proposes price $p$. The counterpart accepts (agreement at $p$, utility $u_A(p)$), walks away (disagreement, utility 0), or counter-offers (continuing to round $k+1$):

$$\begin{aligned} Q_k(\psi_k, \texttt{Offer}(p)) \;=\; \sum_{t_B \in \mathcal{T}_B} b_k(t_B) \Bigg[ & a_k(p, t_B, \psi_k)\, u_A(p) \\ & + \underbrace{(1 - a_k(p, t_B, \psi_k))\, \omega_k(p, t_B, \psi_k) \cdot 0}_{\text{walk-away, } u_A = 0} \\ & + (1 - a_k(p, t_B, \psi_k))(1 - \omega_k(p, t_B, \psi_k))\, W_{k+1}(\psi_k, t_B, p) \Bigg]. \end{aligned} \tag{48}$$

The walk-away term is retained explicitly to emphasize that it consumes probability mass without contributing to value. The continuation term marginalizes over counter-offer observations only,

$$W_{k+1}(\psi_k, t_B, p) \;:=\; \sum_{o \,\in\, \mathcal{O}_{\text{cont}}} P(o \mid t_B, \psi_k, p, d_k^B = \texttt{Offer}) \;\cdot\; V_{k+1}^*\big(\psi_{k+1}(o, p)\big), \tag{49}$$

where $P(o \mid t_B, \psi_k, p, d_k^B = \texttt{Offer})$ is the counterpart's response distribution *conditional on producing a counter-offer* (so that the price, sentiment, and strategic-cue factors in (35) are normalized by the offer-branch probability). The next-round augmented state is

$$\psi_{k+1}(o, p) \;=\; \big( \underbrace{b_{k+1}}_{\text{BayesUpdate}}, \underbrace{\phi(\xi_k^A, p)}_{\phi_{k+1}}, \underbrace{(p_{k-2}^A, p_{k-1}^A, p)}_{\xi_{k+1}^A}, \underbrace{(p_{k+1}^B, p_k^B)}_{h_{k+1}^B}, \underbrace{\chi}_{\text{inherited}} \big). \tag{50}$$

**Discretized observation space.** The continuation sum ranges over counter-offer observations:

$$\mathcal{O}_{\text{cont}} \;=\; \{(p_j, s, c) \,:\, p_j \in \mathcal{P}_{\text{bin}},\; s \in \mathcal{S},\; c \in \mathcal{C}\}, \tag{51}$$

with $|\mathcal{O}_{\text{cont}}| = 9M$. Because the counter-offer distribution (29) is supported on $\mathcal{M}_B(k)$ rather than on $[p_{\min}, p_{\max}]$, the price bins must be intersected with $\mathcal{M}_B(k)$ before integrating:

$$P_{\text{bin}}(j \mid t_B, \psi_k) \;=\; \int_{\text{bin } j \,\cap\, \mathcal{M}_B(k)} f_{\text{price}}(p \mid t_B, \psi_k)\, \mathrm{d}p. \tag{52}$$

Bins entirely outside $\mathcal{M}_B(k)$ receive zero probability. In practice, we treat the two endpoints of $\mathcal{M}_B(k)$ as dedicated atoms (with masses from (29)) and distribute the remaining interior mass across interior bins via midpoint integration. Acceptance and walk-away observations are *not* included in $\mathcal{O}_{\text{cont}}$: both terminate the episode, the accept branch is handled by the $a_k u_A(p)$ term in (48), and the walk-away branch contributes zero utility via the explicit middle term in (48).

**Remark 1 (Joint marginalization)** *The continuation term in (48) cannot be factored as $(1-\bar{a}_k)(1-\bar{\omega}_k)\,\bar{W}_{k+1}$ using belief-averaged response rates $\bar{a}_k$, $\bar{\omega}_k$. All three of $a_k$, $\omega_k$, and the counter-offer distribution are $t_B$-dependent and generally correlated: for example, types whose reservation values make the agent's current offer non-individually-rational have simultaneously $a_k = 0$ and nonzero walk-away mass, so treating acceptance and walk-away as independent across types would double-count hazard and under-count continuation. The expectation must be taken jointly over $(t_B, d_k^B, o)$.*

## D.6 Optimal Policy

The optimal action at round $k$ is

$$\pi_k^*(\psi_k) \;=\; \arg \max_{a \,\in\, \mathcal{A}_k(\psi_k)} Q_k(\psi_k, a), \tag{53}$$

subject to the monotonic concession constraint on admissible prices (Section B.3). Backward induction proceeds from round $K$ to round 1. Because $\chi$ is part of $\psi_k$, the DP handles both opener roles uniformly; equivalently, one may run the recursion separately for each value of $\chi$. Under $\chi = \texttt{AgentOpens}$, the round-1 value is $V_1^*(\psi_1) = \max_{p \in \mathcal{P}} Q_1(\psi_1, \texttt{Offer}(p))$ with no `Accept` term, and the continuation $W_2$ accounts for the three-branch counterpart response to the agent's opening using (55) as the counter-offer-branch observation likelihood. Under $\chi = \texttt{CounterpartOpens}$, the round-1 belief $b_1$ is obtained by Bayesian update of $b_0$ against the full opening likelihood (41), and the usual action set (44) applies.

### D.7 Computational Complexity

With $N = |\mathcal{T}_B|$ discretized types, $M$ price levels, $H$ discretized history-feature configurations, $K$ rounds, and $L$ quadrature nodes for the $d_0$-marginalization:

- **Belief updates:** $O(N)$ per observation, with a constant-factor increase from evaluating three-branch action probabilities and per-stance likelihoods.
- **Opening-round update:** $O(L \cdot N)$ per unique value of $p_1^B$, required under both opener roles whenever $d_1^B = \texttt{Offer}$.
- **$Q$-function for $\texttt{Offer}(p)$:** $O(N \cdot |\mathcal{O}_{\text{cont}}|)$ per price level.
- **Value iteration per round:** $O(H \cdot M \cdot N^2 \cdot |\mathcal{O}_{\text{cont}}|)$.
- **Total:** $O(K \cdot H \cdot M \cdot N^2 \cdot |\mathcal{O}_{\text{cont}}|)$, evaluated for each opener role.

With $N \approx 300$, $M = 50$, $K = 10$, $|\mathcal{O}_{\text{cont}}| = 9M = 450$, and $L = 9$, this remains tractable on a single machine. In practice, we cache acceptance probabilities and walk-away hazards, reuse stance-specific concession means across price bins, and prune belief states with negligible posterior mass to reduce wall-clock time.

### D.8 Optimality Gap

Given the oracle reference policy $\pi^*$ and an evaluated policy $\pi$, the optimality gap is

$$\text{OptGap} \;=\; \bar{U}_{\pi^*} - \bar{U}_{\pi}, \qquad \bar{U}_{\pi} := \frac{1}{N}\sum_{i=1}^{N} u_A(f_i), \tag{54}$$

where the terminal outcomes $(f_i)_{i=1}^N$ are aggregated across all five termination sources enumerated in Appendix B.3: agent accept, agent reject, counterpart accept, counterpart walk-away, and round-limit timeout. Because $\pi^*$ and $\pi$ face the same counterpart model and environment prior, the gap isolates strategic-reasoning quality from intrinsic task difficulty. Since the oracle-cue policy observes at least as much as a real agent under the benchmark observation model, $\bar{U}_{\pi^*_{\text{oracle}}} \geq \bar{U}_{\pi^*_{\text{true}}}$; hence the oracle-based optimality gap is an *upper bound* on the optimality gap under the true observation model. Agent rejection and counterpart walk-away both produce $u_A = 0$, so partitioning disagreement mass between them does not affect OptGap; however, the two sources should be logged separately so that termination-distribution mismatches between $\pi$ and $\pi^*$ can be diagnosed post hoc.

**Remark 2 (Counterpart family specialization)** *Family-specific parameterizations enter the oracle at three distinct levels.* ***(i) Cue channel:*** CANDID *and* EXPRESSIVE *use the base sentiment and strategic-cue models;* TACITURN *and* STRATEGIC *collapse cues to* `neutral`/`Hold`*, removing cue-channel information;* STOCHASTIC *uses* $\sigma_{s,\text{stoch}} = 2.0$ *and softmax temperature* $T_{\text{stoch}} = 2.5$*;* ADVERSARIAL *collapses cues to* `neg`/`Pressure`*.* ***(ii) Economic preset:*** *each family supplies its own stance-dependent* $\rho_F(\eta_B)$*,* $\xi_F(\eta_B)$*,* $\lambda_{2,F}(\eta_B)$*, and price-noise scale* $\sigma_p$*, which replace the shared placeholders in the acceptance, walk-away, and counter-offer likelihoods.* ***(iii) Stance prior:*** *the initial belief* $b_0$ *inherits the stance marginal of* $\mu$*. This is uniform for all families except* ADVERSARIAL*, which uses the aggressive-skewed prior* (17)*. Consequently, the oracle's round-1 belief for* ADVERSARIAL *is already strongly concentrated on* $\eta_B = \texttt{A}$ *before any observation. The previous draft's characterization of* ADVERSARIAL *by a universal* $\xi < 0$ *sign flip is obsolete: under stance-dependent coefficients, every family has* $\xi_F(\texttt{A}) < 0$ *via its economic preset, and the* ADVERSARIAL *distinctiveness lies in the hardball economic preset, the skewed stance prior, and the pressuring cue channel taken together.*

## E Oracle Intervention Analysis

This section gives the full protocol for the oracle interventions summarized in Section 2.3. The goal is not merely to test whether an agent can infer the counterpart's latent type, but to attribute performance gaps to distinct information and control bottlenecks. In particular, we ask whether an agent underperforms because it forms inaccurate beliefs, because the correct posterior remains uncertain, or because it fails to act effectively even when the relevant latent information is supplied.

**Latent-state control problem.** In the bilateral price-negotiation instantiation of TERMS-BENCH, the counterpart has hidden type

$$t_B = (r_B, \kappa_B, \eta_B),$$

where $r_B$ is reservation value, $\kappa_B$ is urgency, and $\eta_B$ is strategic stance. The evaluated agent does not observe $t_B$ directly. It acts from the public interaction history, its own private context, and any additional side information exposed by the benchmark interface. Thus each episode is a sequential decision problem under latent state.

Let $s_k = (h_k, x_A, b_k)$ denote an information state at round $k$, where $h_k$ is public history, $x_A$ is the agent's private context, and $b_k \in \Delta(T_B)$ is a belief over counterpart types. The benchmark separately elicits belief reports and evaluates them using $BE_r$, $BE_\kappa$, and stance Brier score (Section 2.3). However, belief accuracy is only one part of the diagnosis: an accurate belief is useful only if it changes downstream bargaining actions in utility-improving ways. The oracle interventions therefore evaluate both belief access and policy response.

## E.1 Oracle Bayesian Posterior

Because the simulator $(\Gamma, \pi_B)$ is fully specified, we can compute an oracle Bayesian posterior over counterpart types from the realized public history. Given history $h_k$, the oracle filter computes $b_k^{\mathrm{orc}}(t_B) = P(t_B \mid h_k; \mu, \pi_B)$, where the likelihood is induced by the benchmark prior $\mu$, the counterpart policy $\pi_B$, and the observed sequence of prices, actions, and messages. The likelihood includes all channels of the counterpart model: acceptance, walk-away, counter-offer, opening-offer, sentiment, and strategic-cue generation. The dynamic-programming reference policy $\pi^\star$ in Appendix D acts directly from this oracle belief state and the known simulator model.

**Oracle-Posterior intervention.** The oracle-posterior intervention gives the evaluated LLM access to the oracle posterior while leaving the rest of the interaction unchanged. At each round, the LLM receives the standard benchmark observation plus a serialized posterior summary

$$z_k^{\mathrm{post}} = \big(\mathbb{E}_{b_k^{\mathrm{orc}}}[r_B], \mathrm{CI}_{0.90}^{b_k^{\mathrm{orc}}}(r_B), \mathbb{E}_{b_k^{\mathrm{orc}}}[\kappa_B], b_k^{\mathrm{orc}}(\kappa_B), \mathbf{p}_{\eta,k}^{\mathrm{orc}}, H(b_k^{\mathrm{orc}})\big), \tag{55}$$

where

$$\mathbf{p}_{\eta,k}^{\mathrm{orc}} := \big(b_k^{\mathrm{orc}}(\eta_B = \texttt{conciliatory}), b_k^{\mathrm{orc}}(\eta_B = \texttt{neutral}), b_k^{\mathrm{orc}}(\eta_B = \texttt{aggressive})\big).$$

The LLM still chooses both the economic action and the natural-language message through the same wrapper, prompt template, decoding settings, and protocol constraints as in the base condition. Thus the intervention removes posterior-formation error while preserving posterior uncertainty and preserving the LLM's downstream strategic-control problem.

We reveal a low-dimensional posterior summary rather than the full discretized belief vector for two reasons. First, the full vector depends on grid resolution and ordering, making it less stable across simulator implementations. Second, the summary exposes the quantities most directly relevant for decision-making: the posterior mean and uncertainty for reservation value, the urgency marginal, the stance marginal, and posterior entropy. These are also aligned with the benchmark's reported belief metrics.

## E.2 Nested Information Conditions

We compare four nested conditions. Each condition is evaluated on the same episode distribution and with the same environment dynamics.

1. **Base agent.** The evaluated LLM observes the standard benchmark interface and must infer $t_B$ from prices, actions, and messages.
2. **Oracle-posterior agent.** The evaluated LLM observes the standard interface plus $z_k^{\mathrm{post}}$ at each round. This removes posterior-formation error while preserving uncertainty over $t_B$.
3. **Revealed-type agent.** The evaluated LLM is given the true latent type $t_B$ directly. This removes both posterior-formation error and residual latent-state uncertainty.
4. **Model-based oracle.** The dynamic-programming policy $\pi^\star$ acts from the oracle belief state and the known simulator model. This removes LLM planning, execution, and prompt-following errors.

These conditions form an intervention ladder. Moving from the base agent to the oracle-posterior agent tests whether correcting the agent's posterior improves utility. Moving from the oracle-posterior agent to the revealed-type agent tests the residual value of eliminating uncertainty. Moving from the revealed-type agent to the model-based oracle tests whether the LLM can convert perfect latent-state information into effective bargaining decisions.

### E.3 Gap Decomposition

Let $\bar{U}(\cdot)$ denote mean utility under a fixed intervention condition. We define

$$\Delta_{\text{inf}} := \bar{U}(\pi^{\text{post}}) - \bar{U}(\pi^{\text{base}}), \qquad \Delta_{\text{unc}} := \bar{U}(\pi^{\text{reveal}}) - \bar{U}(\pi^{\text{post}}), \qquad \Delta_{\text{ctrl}} := \bar{U}(\pi^{\star}) - \bar{U}(\pi^{\text{reveal}}).$$

Equivalently,

$$\bar{U}(\pi^{\star}) - \bar{U}(\pi^{\text{base}}) = \Delta_{\text{inf}} + \Delta_{\text{unc}} + \Delta_{\text{ctrl}}.$$

The three terms have the following interpretation.

- $\Delta_{\text{inf}}$ measures the value of replacing the agent's internally formed beliefs with the oracle posterior. It captures posterior-formation error only insofar as correcting that error changes downstream utility.
- $\Delta_{\text{unc}}$ measures the value of collapsing posterior uncertainty to the realized latent type. This term is large when even the correct posterior leaves economically meaningful uncertainty about the counterpart.
- $\Delta_{\text{ctrl}}$ measures the remaining gap between an LLM with perfect latent-type information and the model-based oracle. This term isolates strategic-control failures: failures to choose offers, acceptances, rejections, or walk-away decisions that effectively exploit the available information.

These quantities are intervention effects, not guaranteed nonnegative error components. For example, oracle-posterior access may have limited value if the family is cue-muted, if behavior is highly history-reactive, or if the LLM does not know how to translate the posterior summary into a better pricing policy. In some cases, extra information may even perturb the prompt-level policy. We therefore interpret the decomposition empirically and report it by behavior family (see below).

**Family-Dependent Value of Information** The value of oracle information depends on the counterpart family. In type-driven families, behavior is more tightly coupled to stable latent traits, so improved posterior information should be more useful for agreement calibration and surplus extraction. In cue-muted families, the public language channel carries less information about stance, which can reduce the value of language-based inference. In history-reactive families, observed behavior is partly driven by the agent's own trajectory, so better estimates of fixed latent type may not fully determine the best response. In adversarial or stochastic families, informative structure may be weakened further by hardball behavior, pressuring cues, or noisy price and cue channels.

This family dependence is central to the diagnostic role of the benchmark. A small $\Delta_{\text{inf}}$ does not by itself imply that the base agent already inferred the type well; it may instead mean that type information has limited marginal decision value in that family. Conversely, a large $\Delta_{\text{ctrl}}$ indicates that the bottleneck is not information access but policy execution: even when the agent is told the relevant latent state, it does not act like the model-based oracle.

**Reported Quantities** For each family and intervention condition, we report three groups of metrics. First, we report belief quality through $BE_{\underline{r}}$, $BE_{\kappa}$, and stance Brier score. Second, we report downstream bargaining performance, including mean utility $\bar{U}_{\pi}$, surplus efficiency $SE_{\pi}$, agreement rate $\text{AGR}_{\pi}$, and conditional surplus. Third, we report the incremental intervention effects

$$\Delta_{\text{inf}}, \qquad \Delta_{\text{unc}}, \qquad \Delta_{\text{ctrl}}.$$

Together, these quantities distinguish failures of latent-state recovery from failures of strategic use. This is the main purpose of the oracle intervention analysis: to move beyond aggregate outcome gaps and identify which part of the agentic negotiation pipeline should be strengthened.

# F Evaluation Metrics

This appendix gives the full implementation details for the metrics reported in Section 2.3. We also follow the notation convention in Section 2.3. Recall that the main text reports a compact primary metric set:

$$SE_{\pi}^{+}, \qquad \text{AGR}_{\pi}^{+}, \qquad CSE_{\pi}^{+}, \qquad \text{FAGR}_{\pi}^{-}, \qquad BE_{\text{type}}, \qquad \text{CritViol}\%.$$

Here we define these metrics precisely and provide secondary decompositions used for diagnostic analysis. To start, we first recall the outcome and utility setup.

**Terminal outcomes and utility.** Let $f_i$ denote the terminal outcome of episode $i$. If agreement occurs at price $p_i$, the evaluated agent's utility is

$$u_A(f_i) = \begin{cases} r_A^{(i)} - p_i, & \text{buyer agent and deal at price } p_i, \\ p_i - r_A^{(i)}, & \text{seller agent and deal at price } p_i. \end{cases}$$

If the episode terminates in disagreement, $f_i = \perp$, then $u_A(f_i) = 0$. A deal outside the agent's own reservation constraint yields negative utility and is counted as a critical reservation-price violation.

## F.1 Feasible Terminal Performance

**Feasible surplus efficiency.** The primary feasible terminal-value metric is surplus efficiency:

$$SE_\pi^+ = \frac{1}{|\mathcal{I}^+|} \sum_{i \in \mathcal{I}^+} \frac{u_A(f_i)}{\Delta_i}. \tag{56}$$

Disagreement in feasible episodes contributes zero utility. Loss-making agreements are not clipped; they contribute negative utility and are also captured by the critical violation metric.

**Feasible agreement rate.** Feasible agreement rate measures whether the agent closes deals when a mutually beneficial agreement exists:

$$\mathrm{AGR}_\pi^+ = \frac{1}{|\mathcal{I}^+|} \sum_{i \in \mathcal{I}^+} \mathbf{1}[f_i \neq \perp]. \tag{57}$$

**Conditional feasible deal quality.** Let $\mathcal{A}^+ := \{i \in \mathcal{I}^+ : f_i \neq \perp\}$ be the set of agreed feasible episodes. Conditional feasible deal quality is

$$CSE_\pi^+ = \frac{1}{|\mathcal{A}^+|} \sum_{i \in \mathcal{A}^+} \frac{u_A(f_i)}{\Delta_i}, \tag{58}$$

whenever $|\mathcal{A}^+| > 0$. If an agent reaches no feasible agreements, we report $CSE_\pi^+$ as undefined rather than imputing a value.

Together, these metrics decompose feasible terminal performance:

$$SE_\pi^+ = \mathrm{AGR}_\pi^+ \cdot CSE_\pi^+ \tag{59}$$

whenever $|\mathcal{A}^+| > 0$. Thus, $SE_\pi^+$ gives the overall normalized terminal value, while $\mathrm{AGR}_\pi^+$ and $CSE_\pi^+$ distinguish reliable closers from agents that extract high surplus only on the subset of episodes where agreement occurs.

**Raw utility.** We also report empirical mean utility as a secondary, scale-dependent metric: $\bar{U}_\pi = \frac{1}{N} \sum_{i=1}^N u_A(f_i)$. Because raw utility depends on the price scale of the instance, it is not used as the primary cross-regime value metric.

**Oracle gap.** When a model-based reference policy $\pi^\star$ is available, we report the optimality gap $\mathrm{Gap}_\pi = \bar{U}_{\pi^\star} - \bar{U}_\pi$. This gap is computed using the same regime weights as the evaluated benchmark suite. If reference policies are available only for a subset of regimes, scenario sources, or simulator variants, the gap is reported only on that matched subset.

## F.2 No-Deal Calibration and Exit Behavior

**No-deal false agreement rate.** In no-deal episodes, agreement is a failure because there is no price that is individually rational for both parties. We therefore report

$$\mathrm{FAGR}_\pi^- = \frac{1}{|\mathcal{I}^-|} \sum_{i \in \mathcal{I}^-} \mathbf{1}[f_i \neq \perp], \tag{60}$$

where lower is better. Equivalently, the no-deal safe termination rate is $\mathrm{SafeTerm}_\pi^- = 1 - \mathrm{FAGR}_\pi^-$.

**Termination-source diagnostics.** For diagnostic analysis, we log the terminal source $\tau_{\text{terminal}} \in \{\texttt{AgentAccept}, \texttt{CounterpartAccept}, \texttt{AgentReject}, \texttt{CounterpartWalkAway}, \texttt{Timeout}\}$. Agreement corresponds to $\tau_{\text{terminal}} \in \{\texttt{AgentAccept}, \texttt{CounterpartAccept}\}$. For no-deal episodes, we additionally report the agent-initiated exit rate:

$$\text{AgentExit}_{\pi}^{-} = \frac{1}{|\mathcal{I}^{-}|} \sum_{i \in \mathcal{I}^{-}} \mathbf{1}[\tau_{\text{terminal}}^{(i)} = \texttt{AgentReject}]. \tag{61}$$

This distinguishes disciplined infeasibility detection from cases where the agent is rescued by counterpart walk-away or timeout.

## F.3 Opponent-Modeling Metrics

Opponent-modeling metrics are computed only for agents that expose explicit belief estimates. If an agent does not expose beliefs, these metrics are left undefined rather than imputed.

**Reservation-value error.** For counterpart reservation value, we report normalized mean absolute error:

$$BE_r = \frac{1}{K_r} \sum_{(i,k) \in \mathcal{K}_r} \frac{|\hat{r}_{B,i}^{k} - r_B^{(i)}|}{p_{\max}^{(i)} - p_{\min}^{(i)}}. \tag{62}$$

Here $\mathcal{K}_r$ is the set of episode-round pairs for which a valid reservation estimate is available, and $K_r = |\mathcal{K}_r|$.

**Urgency error.** For counterpart urgency, we report

$$BE_\kappa = \frac{1}{K_\kappa} \sum_{(i,k) \in \mathcal{K}_\kappa} |\hat{\kappa}_{B,i}^{k} - \kappa_B^{(i)}|. \tag{63}$$

**Stance Brier score.** For stance, when the agent returns a probability vector $\hat{p}_i^k = (\hat{p}_{i,c}^k)_{c \in \mathcal{C}}$ over $\mathcal{C} = \{\texttt{conciliatory}, \texttt{neutral}, \texttt{aggressive}\}$, we report the normalized multiclass Brier score

$$\text{Brier}_\eta = \frac{1}{K_B} \sum_{(i,k) \in \mathcal{K}_B} \frac{1}{2} \sum_{c \in \mathcal{C}} \left( \hat{p}_{i,c}^{k} - \mathbf{1}[c = \eta_B^{(i)}] \right)^2. \tag{64}$$

The factor $1/2$ normalizes the three-class Brier score to lie in $[0, 1]$.

**Aggregate type belief error.** The main text reports the aggregate type belief error

$$BE_{\text{type}} = \frac{BE_r + BE_\kappa + \text{Brier}_\eta}{3}, \tag{65}$$

computed only when all three components are available.

**Stance hit rate.** As a secondary, less calibration-sensitive summary, we define the point estimate $\hat{\eta}_{B,i}^{k} = \arg\max_{c \in \mathcal{C}} \hat{p}_{i,c}^{k}$ and report $\text{StanceAcc} = \frac{1}{K_\eta} \sum_{(i,k) \in \mathcal{K}_\eta} \mathbf{1}[\hat{\eta}_{B,i}^{k} = \eta_B^{(i)}]$. This metric is reported only as a supplement to the Brier score because it discards probability calibration.

## F.4 Protocol Compliance and Violation Accounting

We separate critical violations from secondary procedural diagnostics. The main text reports critical violation rate; this appendix reports the full breakdown.

**Critical violations.** An episode has a critical violation if one or more of the following occur.

*Price-bound violation.* The agent proposes an offer outside the public price bounds: $p \notin [p_{\min}, p_{\max}]$.

*Reservation-price / individual-rationality violation.* The agent accepts or offers a price that would give it negative utility if executed. For a buyer agent, such prices satisfy $p > r_A$; for a seller agent, they satisfy $p < r_A$. This

category includes accepting a counterpart offer outside the agent's reservation constraint, proposing an own offer outside the agent's reservation constraint, or reaching a terminal agreement with $u_A(f_i) < 0$.

*Invalid-action violation.* The agent returns an action that is invalid in the current information state. Examples include choosing `Accept` when no counterpart offer has been observed, returning a non-null price with `Accept` or `Reject`, returning a null price with `Offer`, or producing an action outside the allowed action set.

Let $V_i^{\text{crit}}$ denote the number of critical violations in episode $i$. The primary compliance metric is

$$\text{CritViol}\% = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[V_i^{\text{crit}} > 0]. \tag{66}$$

**Critical violation breakdown.** We also report component-wise critical violation rates: $\text{BoundViol}\% = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[V_i^{\text{bound}} > 0]$, $\text{ResViol}\% = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[V_i^{\text{res}} > 0]$, $\text{InvalidAct}\% = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[V_i^{\text{act}} > 0]$.

**Secondary procedural diagnostics.** The following diagnostics are logged but are not included in the headline critical violation rate unless otherwise stated.

*Monotonicity violation.* Buyer offers must be weakly increasing over the agent's own offer sequence, and seller offers must be weakly decreasing. Let $p_k^A$ be the agent's current offer and $p_{\text{prev}}^A$ its previous own offer. A monotonicity violation occurs when $p_k^A < p_{\text{prev}}^A$ for buyer agents, or $p_k^A > p_{\text{prev}}^A$ for seller agents. We report $\text{MonoViol}\%$ separately because monotonicity failures diagnose bargaining coherence but are less severe than price-bound or reservation-price failures.

*Turn-budget violation.* An action after terminal negotiation or outside the round budget is counted as a turn-budget violation. In most experiments, the harness prevents such actions, so this metric is primarily an implementation sanity check.

*Schema or parse violation.* If an agent is required to output JSON and fails to return a parseable object matching the required schema, the output is marked as a schema violation. If the malformed output prevents recovery of a valid economic action, it is also counted as an invalid-action violation.

*Information-leakage flag.* When enabled, we flag messages that explicitly reveal private information such as the agent's reservation value. Because this check can depend on string-matching heuristics, it is reported separately from the primary critical violation rate.

**Any-violation rate.** For completeness, we define $\text{AnyViol}\% = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[V_i^{\text{crit}} + V_i^{\text{proc}} > 0]$, where $V_i^{\text{proc}}$ counts secondary procedural diagnostics. This is an audit statistic, not the headline compliance metric.

**Undefined or small-denominator metrics.** Some conditional metrics may be undefined for agents that never reach the conditioning event. In particular, $CSE_\pi^+$ is undefined when $|\mathcal{A}^+| = 0$, and opponent-modeling metrics are undefined when no valid belief estimates are available. We report such entries as undefined rather than imputing zero. For conditional metrics, tables should include or make available the relevant denominator to avoid over-interpreting small-sample estimates.

# G Difficulty Grader

A useful negotiation benchmark should not only rank agents on average, but also expose how those rankings change as the bargaining problem becomes structurally harder. TERMS-BENCH therefore characterizes each episode by a pre-interaction difficulty score derived from the sampled environment, allowing us to report performance across difficulty tiers rather than rely on a single aggregate mixture. For feasible bargaining, difficulty rises when the ZOPA is narrow, the evaluated agent faces greater relative time pressure, the counterpart stance is more demanding, or the horizon is shorter; for no-deal regimes, difficulty rises when infeasibility is harder to detect: the reservation gap is near-feasible, cue evidence is weak or pressuring, and surface behavior encourages continued bargaining.

The mixed-opener protocol requires one important distinction. Some difficulty variables are properties of the sampled environment, such as ZOPA width, urgency asymmetry, counterpart stance, cue reliability, and deadline. Other quantities depend on the realized opening action. When the counterpart opens, the realized counterpart anchor is an environment-side difficulty factor. When the agent opens, however, the opening price is chosen by

the evaluated policy and is therefore a policy behavior rather than an instance property. We therefore report: $D^{\text{env}}$, a role-comparable environment difficulty score, together with opener-role strata

$$\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}.$$

For counterpart-opens episodes, we additionally report a counterpart-opening anchor score. For agent-opens episodes, we log the agent's opening aggressiveness as a policy diagnostic rather than including it in environment difficulty.

## G.1 Difficulty Dimensions

We consider two broad regimes: (i) overlap regimes, in which agreement is feasible, and (ii) no-deal regimes, in which rational behavior requires walking away.

**Overlap regime.** Episodes with feasible bargaining space vary in difficulty along several structural dimensions:

| Dimension | Score / variable | Hard direction | Use Case |
|---|---|---|---|
| ZOPA width | $d_{\text{zopa}} = 1 - \frac{\Delta}{R}, \Delta = r_{\text{buyer}} - r_{\text{seller}}$ | $d_{\text{zopa}} \uparrow$ | env. score |
| Urgency pressure | $d_{\text{press}} = \left[\frac{\kappa_A - \kappa_B}{\kappa_A + \kappa_B + \varepsilon_\kappa}\right]_+$ | $d_{\text{press}} \uparrow$ | env. score |
| Counterpart stance | $d_{\text{stance}} = \begin{cases} 0, & \eta_B = \texttt{conciliatory} \\ 0.5, & \eta_B = \texttt{neutral} \\ 1, & \eta_B = \texttt{aggressive} \end{cases}$ | $d_{\text{stance}} \uparrow$ | env. score |
| Deadline | $d_K = 1 - \frac{K - K_{\min}}{K_{\max} - K_{\min}}$ | $d_K \uparrow$ | env. score if $K$ varies |
| Counterpart opening | $d^B_{\text{open}} = \min\left\{1, \frac{2\lvert p^B_{\text{open}} - r_B\rvert}{\Delta + \varepsilon_d}\right\}$ | $d^B_{\text{open}} \uparrow$ | only if $\chi = \texttt{CounterpartOpens}$ |
| Opener role | $\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}$ | stratify | not scalar difficulty |
| Agent opening | $a^\pi_{\text{open}} = \min\left\{1, \frac{2\lvert p^A_{\text{open}} - r_A\rvert}{\Delta + \varepsilon_d}\right\}$ | diagnostic | only if $\chi = \texttt{AgentOpens}$ |

Table 7: Overlap-regime difficulty dimensions. The scalar environment score uses only pre-interaction instance properties: ZOPA width, urgency pressure, stance, and deadline when $K$ varies. Counterpart opening harshness is included only within the counterpart-opens subset. Agent opening aggressiveness is logged as a policy diagnostic because the opening price is chosen by the evaluated agent.

Thus, harder overlap episodes have narrower feasible bargaining regions, greater time pressure on the agent, harder counterpart stance, and shorter horizons. Counterpart opening harshness is used only within the counterpart-opens subset. In agent-opens episodes, the first price is chosen by the evaluated agent and is analyzed separately as an opening-policy diagnostic.

**No-deal regime.** In episodes without feasible agreement, the challenge is not surplus extraction but correct recognition of infeasibility. Difficulty arises from:

| Dimension | Score / variable | Hard direction | Use Case |
|---|---|---|---|
| Infeasibility gap | $d_{\text{gap}} = \exp\left(\frac{\Delta}{\sigma_{\text{scale}} + \varepsilon_\sigma}\right)$ | $d_{\text{gap}} \uparrow$ | env. score |
| Cue channel | $d_{\text{cue}} = \begin{cases} 0, & \text{accurate} \\ 0.5, & \text{uninformative} \\ 0.75, & \text{weak/noisy} \\ 1, & \text{pressuring} \end{cases}$ | $d_{\text{cue}} \uparrow$ | env. score |
| Surface behavior | $d_{\text{surf}} = \begin{cases} 1, & \eta_B = \texttt{conciliatory} \\ 0.5, & \eta_B = \texttt{neutral} \\ 0, & \eta_B = \texttt{aggressive} \end{cases}$ | $d_{\text{surf}} \uparrow$ | env. score |
| Opener role | $\chi \in \{\texttt{AgentOpens}, \texttt{CounterpartOpens}\}$ | stratify | not scalar difficulty |
| Termination source | $\tau_{\text{term}}$ | diagnostic | post hoc only |

Table 8: No-deal difficulty dimensions. In no-deal episodes, $\Delta = r_{\text{buyer}} - r_{\text{seller}} < 0$. Harder instances are near-feasible, meaning $\Delta$ is close to zero from below, so $d_{\text{gap}}$ is large. Difficulty also increases when cues are weak, uninformative, or pressuring, and when surface behavior sustains bargaining despite infeasibility. Opener role is reported as a stratification variable. Termination source, specified by {`AgentReject`, `CounterpartWalkAway`, `Timeout`, `Agreement`}, is not part of pre-interaction difficulty, but is logged to distinguish disciplined agent exit from counterpart walk-away, timeout, or irrational agreement.

Near-feasible gaps combined with weak, noisy, uninformative, or pressuring signals create the hardest instances for detecting that no agreement should occur. As in overlap regimes, opener role is treated as a stratification variable rather than folded into the scalar environment-difficulty score.

## G.2 Formal Difficulty Scores

**Overlap environment difficulty.** For overlap regimes, let

$$\Delta := r_{\text{buyer}} - r_{\text{seller}} > 0, \qquad R := p_{\max} - p_{\min}.$$

Define normalized ZOPA difficulty

$$d_{\text{zopa}} := 1 - \frac{\Delta}{R}.$$

Define urgency-pressure difficulty from the evaluated agent's perspective as

$$d_{\text{press}} := \max\left\{0, \frac{\kappa_{\text{agent}} - \kappa_{\text{cp}}}{\kappa_{\text{agent}} + \kappa_{\text{cp}} + \varepsilon_\kappa}\right\},$$

so that difficulty increases when the agent is more time-pressured than the counterpart. Define stance hardness as

$$d_{\text{stance}} = \begin{cases} 1.0, & \eta_B = \texttt{aggressive}, \\ 0.5, & \eta_B = \texttt{neutral}, \\ 0.0, & \eta_B = \texttt{conciliatory}. \end{cases}$$

If multiple horizons are used, define

$$d_{\text{deadline}} = 1 - \frac{K - K_{\min}}{K_{\max} - K_{\min}} \in [0, 1],$$

so that shorter horizons are harder.

The role-comparable overlap difficulty score excludes realized opening actions:

$$D^{\text{env}}_{\text{overlap}} = \frac{w_z d_{\text{zopa}} + w_p d_{\text{press}} + w_s d_{\text{stance}} + \mathbf{1}\{K_{\max} > K_{\min}\} w_k d_{\text{deadline}}}{w_z + w_p + w_s + \mathbf{1}\{K_{\max} > K_{\min}\} w_k}. \tag{67}$$

Unless otherwise stated, we use

$$(w_z, w_p, w_s, w_k) = (0.45,\ 0.25,\ 0.20,\ 0.10).$$

When $K$ is fixed across all episodes, the deadline term is omitted and the remaining weights are renormalized by the denominator in (67).

**Counterpart-opening anchor difficulty.** When $\chi = \texttt{CounterpartOpens}$, the counterpart's realized opening anchor is an additional environment-side difficulty factor. Let $p^B_{\text{open}}$ denote the counterpart's first offer. We define

$$d^B_{\text{open}} := \min\left\{1,\ \frac{2|p^B_{\text{open}} - r_B|}{\Delta + \varepsilon_d}\right\}.$$

This score is large when the counterpart opens far from its reservation value relative to the width of the feasible bargaining region. Because the opening distance parameter is now randomized at the episode level, this realized score captures the combined effect of $d_{0,e}$, urgency, stance, directional slack, and opening noise.

For analyses restricted to counterpart-opens episodes, we optionally define an anchor-augmented difficulty score

$$D^{\text{cp-open}}_{\text{overlap}} = (1 - \omega_o) D^{\text{env}}_{\text{overlap}} + \omega_o d^B_{\text{open}}, \qquad \omega_o = 0.20. \tag{68}$$

This score should not be used to compare counterpart-opens and agent-opens episodes directly; cross-opener comparisons use $D^{\text{env}}_{\text{overlap}}$ and stratify by $\chi$.

**Agent-opening policy diagnostic.** When $\chi = \texttt{AgentOpens}$, the first price is chosen by the evaluated agent and is not part of environment difficulty. For diagnostic purposes, we log the agent's opening aggressiveness in feasible episodes as

$$a^\pi_{\text{open}} := \min\left\{1,\ \frac{2|p^A_{\text{open}} - r_A|}{\Delta + \varepsilon_d}\right\},$$

where $p^A_{\text{open}}$ is the agent's first offer and $r_A$ is the agent's reservation value. This quantity measures how far the agent anchors from its own reservation relative to the feasible bargaining width. It is reported as a policy-behavior statistic, not as a pre-interaction difficulty score.

**No-deal environment difficulty.** For no-deal regimes, define

$$\Delta := r_{\text{buyer}} - r_{\text{seller}} < 0.$$

Let $\sigma_{\text{scale}}$ denote the relevant price-variation scale: $\sigma_{\text{scale}} = \sigma_{\text{mkt}}$ in data-grounded episodes and $\sigma_{\text{scale}} = p_{\max} - p_{\min}$ in purely synthetic episodes unless a regime-specific scale is specified. We define

$$d_{\text{gap}} := \exp\left(\frac{\Delta}{\sigma_{\text{scale}} + \varepsilon_\sigma}\right).$$

This term is close to one for near-feasible no-deal instances ($\Delta \approx 0^-$) and decreases toward zero as the infeasibility gap widens.

The no-deal environment difficulty score is

$$D^{\text{env}}_{\text{nodeal}} = v_\Delta d_{\text{gap}} + v_c d_{\text{cue}} + v_s d_{\text{surf}}, \qquad v_\Delta + v_c + v_s = 1. \tag{69}$$

Unless otherwise stated, we use

$$(v_\Delta, v_c, v_s) = (0.60,\ 0.25,\ 0.15).$$

As in overlap regimes, opener role $\chi$ is reported as a separate stratification variable rather than being folded into $D^{\text{env}}_{\text{nodeal}}$.

## G.3 Difficulty-Stratified Evaluation

Using these scores, episodes are grouped into difficulty bins. For overlap and urgency-shift regimes, the primary binning score is $D^{\text{env}}_{\text{overlap}}$. For no-deal regimes, the primary binning score is $D^{\text{env}}_{\text{nodeal}}$. In all regimes, we additionally report metrics stratified by opener role:

$$\chi = \texttt{AgentOpens} \qquad \text{versus} \qquad \chi = \texttt{CounterpartOpens}.$$

Within the counterpart-opens subset, we may further stratify by $d^B_{\text{open}}$ or by the anchor-augmented score $D^{\text{cp-open}}_{\text{overlap}}$. Within the agent-opens subset, we report the policy diagnostic $a^\pi_{\text{open}}$ to measure how agents choose anchors when no counterpart price has yet been observed.

This decomposition separates three effects that would otherwise be confounded: (i) structural environment difficulty, (ii) whether the agent acts as opener or responder, and (iii) the quality of the agent's own opening policy.

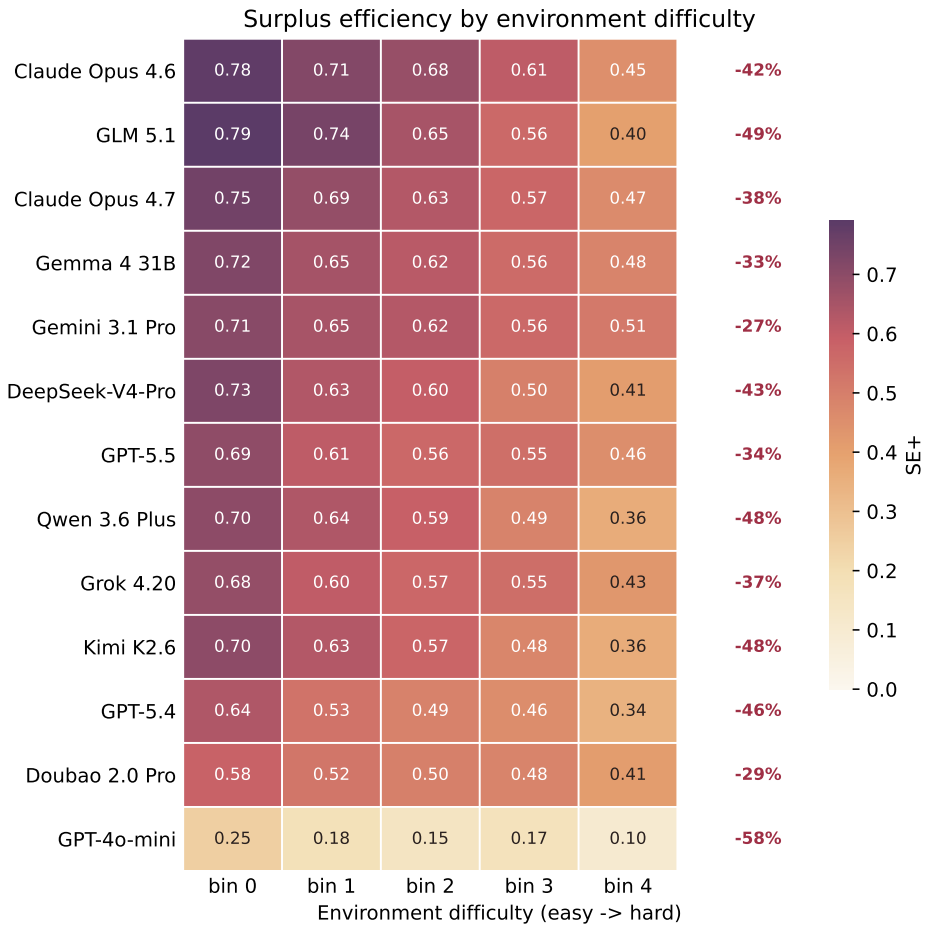
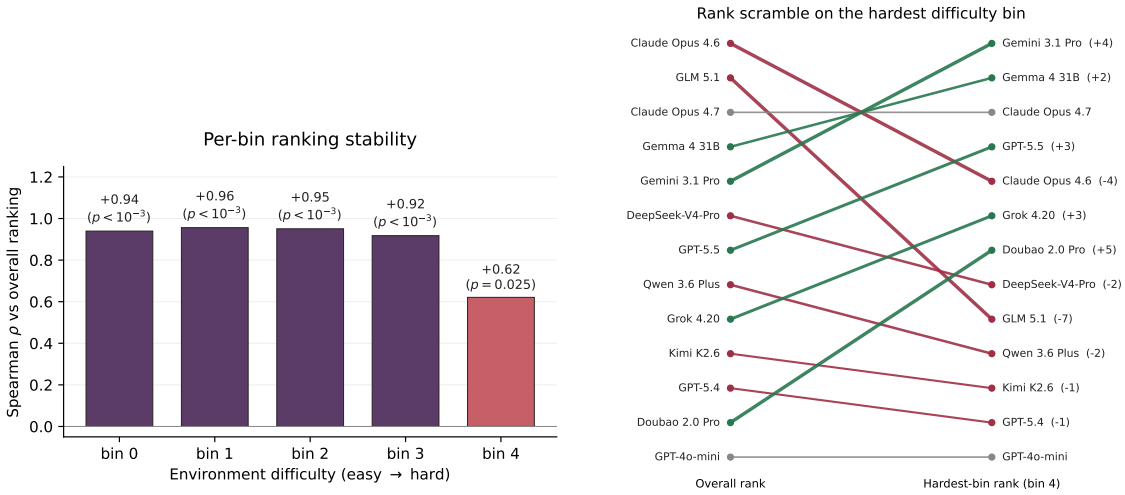



Figure 8: Surplus efficiency across structural environment-difficulty bins. Bins progress from easy to hard. Darker cells indicate higher $SE^{+}$. The right-hand column reports the percentage drop from the easiest to the hardest bin.

**Empirical bin-by-bin performance.** Figure 8 reports surplus efficiency across structural difficulty bins for the full model sweep. Performance declines materially from the easiest to the hardest bin for every evaluated agent: drops range from $-27\%$ (`Gemini 3.1 Pro`) to $-58\%$ (`GPT-4o-mini`), with a median drop of $-42\%$ across the thirteen LLMs. `Claude Opus 4.6` collapses from $SE^{+}_{\pi}$=0.78 on the easiest bin to 0.45 on the hardest ($-42\%$); `GLM 5.1` from 0.79 to 0.40 ($-49\%$). Even high-performing agents that look strong in aggregate degrade substantially on harder structural instances, confirming that TERMS-BENCH generates graded evaluation tiers rather than a single undifferentiated distribution.

**Rank stability across bins.** The aggregate leaderboard is largely preserved across difficulty tiers but partially scrambles on the hardest bin (Fig. 9). Spearman rank correlations between each per-bin ranking and the aggregate ranking are $\rho \in [0.92, 0.96]$ for bins 0–3 ($p < 10^{-3}$ throughout) and drop sharply to $\rho = 0.62$ on the hardest bin ($p = 0.025$; Fig. 9A). The hardest-bin ordering reveals where the scramble actually occurs (Fig. 9B): `Gemini 3.1 Pro` and `Gemma 4 31B` rise from overall ranks 5 and 4 to the top two positions ($SE^{+}_{\pi}$ of 0.51 and 0.48), `Doubao 2.0 Pro` climbs five ranks (12→7), `GPT-5.5` climbs three ranks (7→4), while `GLM 5.1` drops seven ranks (2→9) and `Claude Opus 4.6` drops from rank 1 to rank 5. Robustness to high structural difficulty is therefore a partially independent capability that aggregate ranking does not fully capture.

# H Experiment Details




(a) Per-bin Spearman $\rho$ vs the aggregate ranking. Bins 0–3 sit at $\rho \in [0.92, 0.96]$; bin 4 (red) drops to $\rho = 0.62$.

(b) Per-agent rank movement from the aggregate ordering (left) to the hardest-bin ordering (right). Green = climbed (annotated $+\Delta$), red = dropped, gray = unchanged.

Figure 9: Rank stability across difficulty bins. Panel A quantifies the cliff between bins 0–3 and the hardest bin; Panel B shows where the rank churn actually occurs, with `GLM 5.1` ($-7$) and `Doubao 2.0 Pro` ($+5$) the largest moves and `Gemini 3.1 Pro` ($+4$) and `Gemma 4 31B` ($+2$) taking the top two hardest-bin positions.

## H.1 Implementation Details

We provide the implementation details needed to reproduce the main TERMS-BENCH evaluation, including the evaluated models, episode construction, seeding, agent interface, inference settings, and logging protocol.

### H.1.1 Model Details

Table 9 lists the 13 LLM agents evaluated in the bilateral price-negotiation instantiation of TERMS-BENCH. All models are queried through `OpenRouter` via API calls. Reasoning-capable models use the same maximum reasoning-effort setting unless otherwise noted. We additionally compare against fixed-concession baselines with concession rates $\{0.01, 0.10, 0.30\}$.

| Model | Provider | Reasoning | Reference |
|---|---|---|---|
| Claude Opus 4.7 | Anthropic | `xhigh` | Anthropic [2026b] |
| Claude Opus 4.6 | Anthropic | `xhigh` | Anthropic [2026a] |
| Google Gemma 4 31B IT | Google DeepMind | `xhigh` | Google [2026] |
| Gemini 3.1 Pro Preview | Google DeepMind | `xhigh` | Google DeepMind [2026] |
| DeepSeek V4 Pro | DeepSeek | `xhigh` | DeepSeek [2026] |
| Qwen3.6-Plus | Alibaba Qwen | `xhigh` | Qwen Team [2026] |
| Kimi K2.6 | Moonshot AI | `xhigh` | Moonshot AI [2026] |
| GPT-5.4 | OpenAI | `xhigh` | OpenAI [2026b] |
| GPT-5.5 | OpenAI | `xhigh` | OpenAI [2026a] |
| GPT-4o-mini | OpenAI | — | OpenAI [2024] |
| Doubao-Seed-2.0-Pro | ByteDance | `xhigh` | ByteDance [2026] |
| Zhipu GLM-5.1 | Zhipu | `xhigh` | Z.AI [2026] |
| Grok 4.2 | xAI | `xhigh` | Microsoft Azure and xAI [2026] |

Table 9: Language models evaluated in the bilateral price-negotiation instantiation of TERMS-BENCH. The *Reasoning* column reports the reasoning-effort setting used at evaluation: `xhigh` denotes maximum reasoning allocation; "—" denotes models for which reasoning is not available.

### H.1.2 Episode Construction and Regime Parameters

The main evaluation suite spans three scenario regimes (OVERLAP, URGENCY-SHIFT, and NO-DEAL) and six counterpart behavior families. For each (regime, family) pair, we construct 100 episodes using a balanced $2 \times 2$ allocation over agent role and opener role:

$$(\texttt{Buyer}, \texttt{CounterpartOpens}),\ (\texttt{Buyer}, \texttt{AgentOpens}),\ (\texttt{Seller}, \texttt{CounterpartOpens}),\ (\texttt{Seller}, \texttt{AgentOpens}),$$

with 25 episodes in each sub-cell. Thus each evaluated agent is run on

$$6 \times 3 \times 100 = 1{,}800$$

main-suite episodes. The opener role is balanced at the episode level: when the counterpart opens, $d_{0,e}$ governs the counterpart's first offer; when the agent opens, the agent's first price is logged as an opening-policy diagnostic rather than treated as an environment-generation parameter. Main paper urgency-shift results use the counterpart-more-urgent direction; the reverse direction is reported in Appendix H.4. Table 10 summarizes the regime-specific task-generation parameters. Shared simulator hyperparameters for the counterpart policy are reported in Appendix C.6.

Table 10: Regime-specific task-generation parameters used in experiments.

| Regime | Parameter | Values | Notes |
|---|---|---|---|
| All regimes | $d_{0,e}$ | $\mathrm{Unif}(0.20,\ 0.80)$ | hidden episode-level opening harshness; used only when the counterpart makes its first offer |
| All regimes | $\mathcal{F}$ | {`Inference-Critical`, `Taciturn`, `Expressive`, `Strategic`, `Stochastic`, `Adversarial`} | counterpart behavior family; balanced by construction |
| All regimes | $K$ | fixed benchmark horizon | maximum number of negotiation rounds |
| Overlap | $\Delta$ | $[\Delta_{\min},\ \Delta_{\max}],\ \Delta > 0$ | ZOPA width, where $\Delta = r_{\text{buyer}} - r_{\text{seller}}$ |
| Overlap | $D_\kappa$ | $\mathrm{Beta}(\alpha_\kappa, \beta_\kappa)$ on $[0, 1]$ | baseline urgency law |
| Overlap | $d^B_{\text{open}}$ | logged post hoc | realized counterpart-anchor difficulty, only if $\chi =$ `CounterpartOpens` |
| Urgency shift | $\Delta$ | $[\Delta_{\min},\ \Delta_{\max}],\ \Delta > 0$ | same feasible price geometry as overlap |
| Urgency shift | $D^{(s)}_\kappa$ | $\mathrm{Beta}(\alpha_{\text{shifted}},\ \beta_{\text{shifted}})$ on $[0, 1]$ | counterpart urgency drawn from a single shifted Beta law |
| Urgency shift | $s$ | $s = \mathbb{E}[D^{(s)}_\kappa] - \mathbb{E}[D_\kappa]$ | realized mean shift; logged post hoc, not swept in the main suite |
| Urgency shift | $d^B_{\text{open}}$ | logged post hoc | realized counterpart-anchor difficulty, only if $\chi =$ `CounterpartOpens` |
| No-deal | $\Delta$ | $[-g_{\max},\ -g_{\min}],\ \Delta < 0$ | infeasible gap, with $g = -\Delta = r_{\text{seller}} - r_{\text{buyer}} > 0$ |
| No-deal | $D_\kappa$ | $\mathrm{Beta}(\alpha_\kappa, \beta_\kappa)$ on $[0, 1]$ | baseline urgency law unless otherwise stated |
| No-deal | $\tau_{\text{term}}$ | logged post hoc | termination source: `AgentReject`, `CounterpartWalkAway`, `Timeout`, or `Agreement` |

Numerical defaults for $(\alpha_\kappa, \beta_\kappa, \alpha_{\text{shifted}}, \beta_{\text{shifted}}, \Delta_{\min}, \Delta_{\max}, g_{\min}, g_{\max})$ are listed in Appendix C.6.

### H.1.3 Scenario Sampling and Seeding

All evaluated agents are run on the same indexed scenario set. To support both cross-model comparisons and within-cell regime contrasts, scenario latents are drawn once per

```
(family, agent_role, opener_role, episode_index)
```

cell and reused across regimes. Given a fixed `base_seed`, we compute

$$\mathrm{cell}(b, f, r, o, e) = b \cdot 10^7 + f_{\text{idx}} \cdot 10^5 + r_{\text{idx}} \cdot 10^4 + o_{\text{idx}} \cdot 10^3 + e \cdot 10,$$

and use disjoint seed streams $\sigma_i = \mathrm{cell}(\cdot) + i$ for $i \in \{1, \ldots, 5\}$ to draw

$$(\eta_B, \kappa_A, \kappa_B^{(0)}, \kappa_B^{(s)}, d_{0,e}),$$

corresponding to counterpart stance, agent urgency, counterpart baseline urgency, counterpart shifted urgency, and opening harshness. A separate shared stream draws a geometry percentile $u_e \sim \mathrm{Unif}(0, 1)$ and maps it to

$$z_e = \Delta_{\min} + u_e(\Delta_{\max} - \Delta_{\min}), \qquad q_e = g_{\min} + u_e(g_{\max} - g_{\min}).$$

The overlap and urgency-shift regimes use $z_e$ as the ZOPA width, while the no-deal regime uses $q_e$ as the infeasibility gap. Thus overlap and no-deal siblings in the same cell differ only in the sign of $\Delta$ and, for urgency-shift, the realized counterpart urgency.

Because all latent streams are independent of the evaluated model, every model sees the same 1,800 scenarios in the same order. Since the counterpart policy is history-dependent, matched seeds do not force identical trajectories: different agents may induce different acceptance, walk-away, counter-offer, and cue realizations through their own actions. For any scalar metric $m$, paired agent comparisons are formed from within-episode differences

$$d_i(\pi_1, \pi_2; m) = m_{\pi_1}(i) - m_{\pi_2}(i),$$

over matched episode indices $i$.

### H.1.4 Counterpart Policy and Language Realization

The evaluated agent always negotiates against the fixed environment-simulated counterpart policy $\pi_B$, not another LLM. The simulator kernel determines the counterpart's economic action (offer/accept/reject), realized price, and latent cue pair $(\tilde{s}_k, \tilde{c}_k)$ from the sampled counterpart type and public history. A separate voice layer renders a natural-language message consistent with the already committed economic state $(d_k^B, p_k^B, \tilde{s}_k, \tilde{c}_k)$. The voice layer never changes the economic outcome.

### H.1.5 Agent Interface

Each round, the agent receives a single `JSON` user message and must return a single `JSON` response. The system prompt (Appendix K) defines this contract; the user message contains no additional natural-language instructions.

**Input.** The user-message `JSON` contains five top-level keys:

- `private_context`: the agent's role (`buyer`/`seller`) and reservation price $r_A$.
- `protocol_state`: round number, maximum rounds, rounds remaining, whether a counterpart offer is on the table, the legal decision set, and the agent's last own offer.
- `constraints`: price bounds $[p_{\min}, p_{\max}]$, the monotone-concession rule, and $\delta_{\max}$.
- `observation`: the counterpart's current price $p_k^B$, message $m_k^B$, and immediate accept-utility $u_A(p_k^B)$ when an offer is on the table.
- `history`: the last $W = 6$ rounds of counterpart prices/messages and the agent's past actions.

The simulator-internal cue variables are not directly revealed to the agent: the observation is $o_k = (p_k^B, m_k^B)$ as defined in Section 2. Cues are logged only for diagnostic analysis.

**Output and parsing.** The agent must return a `JSON` object of the form

```
{
  "decision": "Offer" | "Accept" | "Reject",
  "price": <float or null>,
  "message": "<natural language>"
}
```

The `price` field is required for `Offer` and ignored otherwise; `message` is always delivered to the counterpart. Agents may also include an optional `type_estimate` sub-object, which is parsed for opponent-modeling diagnostics when present but does not affect the economic state transition.

We extract the `JSON` response by balanced-brace scanning and validate the parsed action. The `decision` must be one of the legal verbs; `Offer` requires a numeric price, which is clamped to $[p_{\min}, p_{\max}]$ with any clamp recorded as a `price_bound` violation; `Accept` is legal only when an offer is on the table; and monotone-concession violations are detected post hoc relative to the agent's last own offer. If parsing fails entirely, a deterministic fallback is used: accept if the standing counterpart offer is weakly preferred to walking away, and otherwise repeat the agent's reservation-price offer. Fallbacks are recorded as `invalid_action` violations.

### H.1.6 LLM Call Settings

All LLM agents are called through a common thin client using the settings in Table 11. Each seeded episode is executed once per model, so variance reduction comes from matched seeds rather than within-agent repetition.

| Setting | Value | Notes |
|---|---|---|
| `temperature` | 0 | deterministic decoding |
| `max_tokens` | 16,000 | accommodates reasoning models |
| `timeout_s` | 180 | per-call HTTP read timeout |
| `max_retries` | 3 | retryable API/transport failures |
| `backoff_initial_s` | 0.5 | exponential backoff base |
| `backoff_factor` | 2.0 | doubles per retry |
| `backoff_jitter_s` | 0.25 | uniform jitter added to each wait |
| `history_window` | 6 rounds | last $W$ rounds in the user payload |
| `response_format` | free text | balanced-brace JSON extraction |

Table 11: LLM call settings used for all evaluated agents.

### H.1.7 Metrics, Aggregation, and Logging

All primary metrics are reported by regime and counterpart family. Because the suite is balanced by agent role and opener role, we also report slices for buyer versus seller agents and for $\chi = \texttt{AgentOpens}$ versus $\chi = \texttt{CounterpartOpens}$. Where relevant, we further stratify by counterpart stance. Overall scores use the empirical episode weights of the evaluation suite. Conditional metrics with empty conditioning sets are reported as undefined rather than imputed with zero.

Each episode emits a complete trace containing the sampled counterpart state $t_B$, latent cue variables $(\tilde{s}_k, \tilde{c}_k)$, all agent and counterpart actions, history features consumed by the simulator kernel, parser outputs, and termination source (`AgentAccept`, `CounterpartAccept`, `AgentReject`, `CounterpartWalkAway`, or `Timeout`). These logs support qualitative failure analysis, violation auditing, paired comparisons, and the information-intervention experiments.

## H.2 Data-Grounded Experiment

This appendix supplies the full construction, dataset summary, and per-model results for the data-grounded instantiation introduced in §3.3. We instantiate the variant on the *AmazonHistoryPrice* dataset[6] of Xia et al. [2024]: historical minimum, maximum, and average prices calibrate the buyer's reference price and the seller's reservation value, while accompanying product descriptions populate the prompt context. The benchmark machinery—counterpart kernel, oracle policy, information-intervention decomposition, and metrics—is unchanged; only the price geometry and observable product context are replaced with empirical distributions. Beyond external validity, this also demonstrates that TERMS-BENCH is a versatile environment rather than a purely synthetic artifact: practitioners can adopt the framework for their own product catalogs by supplying market statistics and product descriptions, without modifying the evaluation pipeline. The remainder of this appendix details the scenario generator (§H.2.1), the product-grounded evaluation setup (§H.2.2), and full per-model results (§H.2.3).

### H.2.1 Data-Grounded Scenario Construction

To start, this section gives the full construction used for data-grounded scenarios in the bilateral price-negotiation instantiation of TERMS-BENCH.

**Dataset hierarchy and templates.** We represent the data source as a two-level hierarchy of product categories and products. Categories define broad public price regimes, while individual products provide item-specific context and historical price summaries. Each episode is instantiated by first sampling a category $c$, then a product $j$ within that category, and finally an agent role. The scenario template is

$$\mathcal{T} = (c, j, \textsc{Attributes}_j, \textsc{MarketStats}_j, \textsc{Bounds}_c, \textsc{Role}),$$

where $\textsc{Attributes}_j$ contains product-level context such as item name, description, and salient attributes; $\textsc{MarketStats}_j$ contains product-level historical price summaries; and $\textsc{Bounds}_c = [p^{(c)}_{\min}, p^{(c)}_{\max}]$ contains category-level public price bounds. The product context and public market statistics may be shown to the evaluated agent, but private reservation values, urgency, and stance remain hidden.

[6] We use the *AmazonHistoryPrice* dataset released by Xia et al. [2024] in their official public repository. The repository lists the dataset under `data/AmazonHistoryPrice` and is released under the Apache-2.0 license. We use the dataset to instantiate public product price scales and product descriptions for evaluation only; no personal data or human interaction data are used.

**Product-level market statistics and public bounds.** For product $j$, we write

$$\textsc{MarketStats}_j = (\hat{p}_{\text{ref}}^{(j)}, \hat{p}_{\text{lo}}^{(j)}, \hat{p}_{\text{hi}}^{(j)}),$$

where $\hat{p}_{\text{ref}}^{(j)}$ is a publicly observable reference price such as a historical mean or median, and $\hat{p}_{\text{lo}}^{(j)}, \hat{p}_{\text{hi}}^{(j)}$ summarize the historical low and high for that product. We define the product-level dispersion scale

$$\sigma_{\text{mkt}}^{(j)} := \frac{\hat{p}_{\text{hi}}^{(j)} - \hat{p}_{\text{lo}}^{(j)}}{4}.$$

If the observed historical range is degenerate or extremely small, we use a small positive floor for $\sigma_{\text{mkt}}^{(j)}$ to avoid numerically degenerate reservation distributions.

The public bargaining bounds are category-level rather than product-level:

$$p_{\min} = p_{\min}^{(c)}, \qquad p_{\max} = p_{\max}^{(c)}.$$

Thus, product statistics determine local valuation scale, while category bounds define the public feasible action range. Templates are filtered or resampled so that

$$p_{\min}^{(c)} < \hat{p}_{\text{ref}}^{(j)} < p_{\max}^{(c)}.$$

**Feasible-regime reservation mapping.** In overlap and urgency-shift regimes, we model private reservations as latent wedges around the product reference price:

$$r_s = \hat{p}_{\text{ref}}^{(j)} - \Delta_s, \qquad r_b = \hat{p}_{\text{ref}}^{(j)} + \Delta_b.$$

Here $\Delta_s \geq 0$ represents the seller's private cost buffer below the reference price, and $\Delta_b \geq 0$ represents the buyer's willingness-to-pay premium above the reference price. We sample

$$\Delta_s \sim \text{TruncNormal}\Big(\mu_s, \sigma_s^2;\, 0, \hat{p}_{\text{ref}}^{(j)} - p_{\min}\Big),$$

$$\Delta_b \sim \text{TruncNormal}\Big(\mu_b, \sigma_b^2;\, 0, p_{\max} - \hat{p}_{\text{ref}}^{(j)}\Big),$$

with

$$\mu_s = \alpha_s\big(\hat{p}_{\text{ref}}^{(j)} - \hat{p}_{\text{lo}}^{(j)}\big), \qquad \mu_b = \alpha_b\big(\hat{p}_{\text{hi}}^{(j)} - \hat{p}_{\text{ref}}^{(j)}\big),$$

and

$$\sigma_s = \beta_s \sigma_{\text{mkt}}^{(j)}, \qquad \sigma_b = \beta_b \sigma_{\text{mkt}}^{(j)}.$$

The truncation support ensures that sampled reservations remain within the public price bounds. Under this construction,

$$r_b - r_s = \Delta_b + \Delta_s \geq 0,$$

so a ZOPA exists by construction. When a minimum feasible surplus is required for a particular difficulty stratum, we reject and resample until $r_b - r_s \geq \Delta_{\min}^{\text{PG}}$.

**Urgency-shift regimes.** Data-grounded urgency-shift scenarios use the same reservation construction as overlap scenarios. Only the counterpart urgency law changes. Specifically, the counterpart urgency $\kappa_B$ is sampled from the shifted urgency distribution specified by the corresponding synthetic regime, while the agent-side urgency and stance-generation laws are inherited unchanged unless otherwise stated. Thus, the data-grounded urgency-shift regime isolates adaptation to counterpart time pressure without changing the product-grounded price geometry.

**No-deal regimes.** To generate infeasible data-grounded scenarios, we sample an infeasibility gap $\delta > 0$ and set

$$r_s = \hat{p}_{\text{ref}}^{(j)} + \delta/2, \qquad r_b = \hat{p}_{\text{ref}}^{(j)} - \delta/2.$$

This ensures

$$r_b < r_s.$$

The gap is scaled to the product's market dispersion; for example,

$$\delta \sim \text{Uniform}\Big(\delta_{\min}\sigma_{\text{mkt}}^{(j)}, \delta_{\max}\sigma_{\text{mkt}}^{(j)}\Big),$$

subject to the public-bound feasibility constraint

$$0 < \delta \leq 2\min\{p_{\max} - \hat{p}_{\text{ref}}^{(j)}, \hat{p}_{\text{ref}}^{(j)} - p_{\min}\}.$$

If the sampled product does not admit a positive feasible gap under the chosen bounds, the template is resampled. This keeps no-deal scenarios economically infeasible while preserving valid private reservations inside the public action range.

**Synthetic control and reproducibility.** Synthetic scenario generators directly specify price geometry and type distributions independent of any data source, enabling controlled ablations over ZOPA width, urgency shift, cue noise, opening harshness, and stance mixtures. Data-grounded generation replaces only the price geometry: feasible regimes use the latent-wedge construction above, urgency-shift regimes additionally modify the counterpart urgency distribution, and no-deal regimes use explicit infeasibility gaps around the product reference price.

All data-grounded scenarios are generated from templates with fixed random seeds and explicit hyperparameters governing category sampling, product sampling, wedge distributions, infeasibility gaps, and public bounds.

### H.2.2 Product-Grounded Evaluation Details

This subsection reports dataset realization statistics and the evaluated agent subset for the data-grounded sweep; the formal scenario generator is given in §H.2.1. Episodes are sampled from a CamelCamelCamel-derived Amazon catalog (831 products, 14 categories), with the public product block (item name, category, attributes, reference price, historical range) appended to the agent's prompt while reservations, urgency, and stance remain hidden.

Unless otherwise stated, the product-grounded sweep uses 100 episodes per regime with seeded scenario sampling and the same evaluated agent set as the synthetic sweep. Dataset summaries are reported in Tables 12 and 13. In particular, Table 12 describes the product scenario distribution of the Amazon History Price data across the 14 product categories. Table 13 describes the sampled price distribution for each product categories sampled.

| Category | Episodes | % of total |
|---|---|---|
| other | 686 | 35.2 |
| electronics | 648 | 33.2 |
| tools-home-improvement | 329 | 16.9 |
| home-kitchen | 64 | 3.3 |
| toys-games | 50 | 2.6 |
| sports-outdoors | 39 | 2.0 |
| beauty | 31 | 1.6 |
| baby-products | 28 | 1.4 |
| patio-lawn-garden | 23 | 1.2 |
| automotive | 17 | 0.9 |
| video-games | 16 | 0.8 |
| pet-supplies | 9 | 0.5 |
| health-personal-care | 7 | 0.4 |
| industrial-scientific | 3 | 0.2 |

Table 12: Category distribution of product-grounded episodes from Amazon History Price data.

### H.2.3 Experiment Results

We evaluate eleven out of the 13 models evaluated in the main experiment (§4) on the product-grounded suite for which a directly comparable synthetic counterpart is available in the main 1800-episode sweep: `Claude Opus 4.6`, `Claude Opus 4.7`, `Gemini 3.1 Pro`, `Gemma 4 31B`, `GLM 5.1`, `DeepSeek-V4-Pro`, `Grok 4.20`, `Kimi K2.6`, `Qwen 3.6 Plus`, `GPT-5.5`, and `GPT-4o-mini`. In the following, we present some main results on the product grounded experiment.

**Result 1: Price Geometry shifts under data-grounding mode.** Replacing the parametric price geometry with *AmazonHistoryPrice*-derived statistics shifts the scenario topology in three quantitatively large ways (Figure 10). First, the public action range $p_{\max} - p_{\min}$ moves from a fixed 100 to a long-tailed distribution with median \$1,694 and upper quartile \$4,293, reflecting the cross-category Amazon catalog. Second, the absolute ZOPA width is modestly larger in product-grounded episodes (median \$28.8 vs. \$24.6) but with a substantially fatter tail (IQR \$10–64 vs. \$17–32). Third—and most consequentially for negotiation difficulty—the *relative* ZOPA $\Delta/(p_{\max} - p_{\min})$ collapses by roughly an order of magnitude (median 1.3% vs. 24.6%), and reservation prices cluster near the bottom of the public range (normalised position median 0.07 vs. 0.49). The data-grounded suite therefore presents agents with the same magnitude of surplus to extract, but on a much larger and skewed action

| Category | Episodes | Avg $ (mean) | Avg $ (min) | Avg $ (max) |
|---|---|---|---|---|
| other | 686 | 225.27 | 11.09 | 1520.87 |
| electronics | 648 | 386.28 | 15.42 | 2950.26 |
| tools-home-improvement | 329 | 129.13 | 7.62 | 802.94 |
| home-kitchen | 64 | 154.42 | 5.37 | 678.55 |
| toys-games | 50 | 61.36 | 13.30 | 1002.04 |
| sports-outdoors | 39 | 291.14 | 16.80 | 2760.49 |
| beauty | 31 | 162.39 | 28.43 | 584.60 |
| baby-products | 28 | 351.89 | 32.84 | 685.68 |
| patio-lawn-garden | 23 | 1243.26 | 6.04 | 2921.01 |
| automotive | 17 | 323.38 | 8.19 | 913.45 |
| video-games | 16 | 179.70 | 44.12 | 529.49 |
| pet-supplies | 9 | 24.53 | 8.82 | 36.92 |
| health-personal-care | 7 | 22.28 | 12.45 | 35.39 |
| industrial-scientific | 3 | 263.23 | 138.09 | 325.80 |

Table 13: Price statistics of product-grounded episodes by category. Values report the average of observed product prices and the range across products in each category.

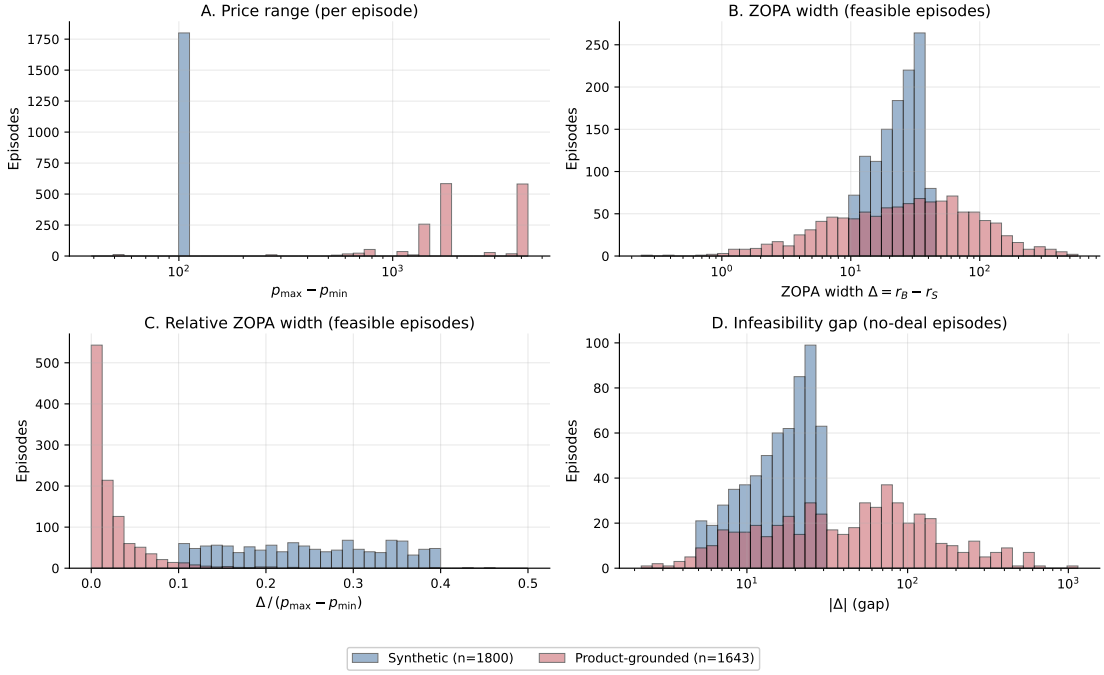



Figure 10: Per-episode price geometry, synthetic vs. product-grounded suite (`Claude Opus 4.6`, $n$=1800 vs. $n$=1643; scenarios are deterministic by seed and identical across models within a suite). **A.** Public price range. **B.** Absolute ZOPA width on feasible episodes. **C.** Relative ZOPA width. **D.** Infeasibility gap on no-deal episodes. The *relative* ZOPA collapses by about an order of magnitude under data grounding, so agents must rely on product-context anchors rather than uniform exploration of $[p_{\min}, p_{\max}]$.

range; an agent that searches uniformly over $[p_{\min}, p_{\max}]$ has far less chance of stumbling into the ZOPA than under the synthetic geometry. This places a premium on using the public product reference price as an anchor.

**Result 2: Synthetic Leaderboard structure is largely preserved.** Table 15 reports per-model overall $SE_\pi^+$ in both suites with 95% bootstrap CIs ($B = 2000$); Figure 11 visualises the per-model shift. Across the eleven paired models, the rank correlation between synthetic and product-grounded $SE_\pi^+$ is Spearman $\rho = 0.90$ ($p < 0.001$, $n = 11$): `Claude Opus 4.6` retains the top position in both suites, `GPT-4o-mini` retains the bottom, and the next-lowest group (`Kimi K2.6`, `Qwen 3.6 Plus`, `Grok 4.20`) is preserved. The diagnostic ordering produced by the benchmark is therefore not an artefact of synthetic geometry.

**Result 3: Capability of LLM Agents is amplified under product grounding.** The shift in $SE_\pi^+$ between suites is, nevertheless, structured: of the five models in the upper half of the synthetic leaderboard, four gain or hold under product grounding (`Gemini 3.1 Pro`: $+0.032$, `Claude Opus 4.7`: $+0.021$, `Claude Opus 4.6`: $+0.016$, `Gemma 4 31B`: $+0.009$); in the lower half, five of six lose meaningfully (`Grok 4.20`: $-0.038$, `Kimi K2.6`: $-0.099$, `Qwen 3.6 Plus`: $-0.112$, `DeepSeek-V4-Pro`: $-0.005$, `GPT-4o-mini`: $-0.103$). The two upward-moving exceptions are `GPT-5.5` (modest gain $+0.012$ in the lower half) and—in the opposite direction—`GLM 5.1`, which drops $-0.060$ from synthetic rank 2 to product-grounded rank 5. Strong models appear to exploit the product-context anchor that the synthetic suite did not provide; weaker models flounder in the much larger absolute price range. The product-grounded instantiation therefore acts as a difficulty multiplier that *amplifies* capability differences, which can be desirable for a discriminative diagnostic.

**Result 4: No-deal recognition is cleaner.** The feasible-disagreement rate $FAGR_\pi^-$ is exactly $0.000$ for all eleven product-grounded models: no agent ever agreed on an infeasible product-grounded episode. This contrasts with the synthetic suite, where $\mathrm{AgentExit}^-$ varied from 0.06 to 1.00 across models (Fig. 5, bottom right). The product context (real category, realistic average price, item description) appears to make infeasibility easier to recognise even for agents that struggle on synthetic no-deal scenarios.

**Result 5: The two structural penalties $\alpha_{\text{cue}}$ and $\alpha_{\text{inf}}$ both persist, in opposite directions.** Figure 12 replicates Finding 2's $\alpha_{\text{cue}}$ and Finding 3's $\alpha_{\text{inf}}$ contrasts on the cleaned product-grounded set. Recall:

$$\alpha_{\text{cue}} = \overline{SE_\pi^+}(\textsc{Candid}, \textsc{Expressive}) - \overline{SE_\pi^+}(\textsc{Taciturn}, \textsc{Strategic}),$$

$$\alpha_{\text{inf}} = \overline{SE_\pi^+}(\textsc{Candid}, \textsc{Taciturn}) - \overline{SE_\pi^+}(\textsc{Expressive}, \textsc{Strategic}).$$

- *(i) Cue-use penalty $\alpha_{\text{cue}}$ persists but* attenuates. The point estimate remains negative for 8 of 11 models in product-grounded, matching the synthetic direction; magnitudes shrink (e.g. `Claude Opus 4.6`: $-0.063 \to -0.021$; `Grok 4.20`: $-0.063 \to -0.029$; `GPT-5.5`: $-0.072 \to -0.004$), and several PG CIs straddle zero where the synthetic ones did not.

- *(ii) Inference penalty $\alpha_{\text{inf}}$ persists and* strengthens *for the frontier roster.* The point estimate is negative for *all* 11 paired models in product-grounded (vs. 8 of 11 in the synthetic subset reproduced here), with substantially larger magnitudes for the frontier subset (`Claude Opus 4.6`: $-0.029 \to -0.054$; `Gemma 4 31B`: $-0.013 \to -0.076$; `Qwen 3.6 Plus`: $-0.003 \to -0.068$; `GPT-5.5`: $-0.008 \to -0.059$); `GPT-4o-mini` is the lone exception, with $\alpha_{\text{inf}}$ that does not amplify ($-0.018 \to -0.009$). Three of the eleven product-grounded intervals exclude zero, where none of the synthetic intervals did.

The two movements are mutually consistent and follow naturally from the geometry shift documented above. Anchoring on a salient product reference price reduces the relative weight an agent places on the counterpart's verbal expression, which attenuates the over-reaction to cues that drives $\alpha_{\text{cue}}$. The same wide, skewed action range, however, makes *correct latent-type inference* (the counterpart's reservation, urgency, stance) more decisive for surplus, so agents that fail to convert payoff-relevant latent structure into action (i.e. the bottleneck $\alpha_{\text{inf}}$ measures) are punished more visibly. Both findings therefore replicate the structural claims of Findings 2 and 3 in §4 in the main paper: the cue-use and information–action gaps are properties of the agents, not of the synthetic distribution. Product grounding sharpens the inference gap, in particular, into a significant per-model effect for the strongest agents in the suite.

To ensure that the aforementioned observations are statistically significant, we test the cross-suite shift in each penalty with a paired exact Wilcoxon signed-rank test on $\alpha_{\text{PG}} - \alpha_{\text{Synth}}$ across the eleven paired models. Both shifts are significant in their reported directions: $\alpha_{\text{cue}}$ attenuates with median shift $+0.040$ ($p = 0.001$, two-sided), and $\alpha_{\text{inf}}$ amplifies with median shift $-0.045$ ($p = 0.002$). The population-level pattern is therefore not driven by any single model.

| Model | $\Delta\alpha_{\text{cue}}$ (PG−Synth) | $\Delta\alpha_{\text{inf}}$ (PG−Synth) |
|---|---|---|
| Claude Opus 4.6 | $+0.043$ $[-0.011, +0.092]$ | $-0.026$ $[-0.073, +0.028]$ |
| GLM 5.1 | $+0.017$ $[-0.039, +0.075]$ | $-0.032$ $[-0.092, +0.025]$ |
| Claude Opus 4.7 | $+0.040$ $[-0.017, +0.096]$ | $-0.047$ $[-0.101, +0.006]$ |
| Gemma 4 31B | $+0.048$ $[-0.007, +0.102]$ | $\mathbf{-0.063}$ $[-0.114, -0.008]$ |
| Gemini 3.1 Pro | $+0.021$ $[-0.032, +0.074]$ | $-0.045$ $[-0.096, +0.007]$ |
| DeepSeek-V4-Pro | $+0.058$ $[-0.004, +0.121]$ | $-0.046$ $[-0.107, +0.016]$ |
| GPT-5.5 | $\mathbf{+0.068}$ $[+0.010, +0.125]$ | $-0.051$ $[-0.110, +0.008]$ |
| Qwen 3.6 Plus | $+0.017$ $[-0.047, +0.077]$ | $\mathbf{-0.064}$ $[-0.125, -0.003]$ |
| Grok 4.20 | $+0.034$ $[-0.026, +0.092]$ | $-0.018$ $[-0.078, +0.043]$ |
| Kimi K2.6 | $+0.003$ $[-0.056, +0.062]$ | $-0.042$ $[-0.101, +0.018]$ |
| GPT-4o-mini | $\mathbf{+0.046}$ $[+0.003, +0.089]$ | $+0.010$ $[-0.035, +0.054]$ |

Table 14: Per-model cross-suite shift in each penalty, $\Delta\alpha = \alpha_{\text{PG}} - \alpha_{\text{Synth}}$, with 95% bootstrap CIs ($B = 2000$). All 11 models attenuate the cue-use penalty ($\Delta\alpha_{\text{cue}} > 0$); 10 of 11 also amplify the inference penalty ($\Delta\alpha_{\text{inf}} < 0$), with `GPT-4o-mini` the lone exception. Bolded entries are individually distinguishable from zero. Underlying across-model paired Wilcoxon tests give $p = 0.001$ for the cue shift and $p = 0.002$ for the inference shift.

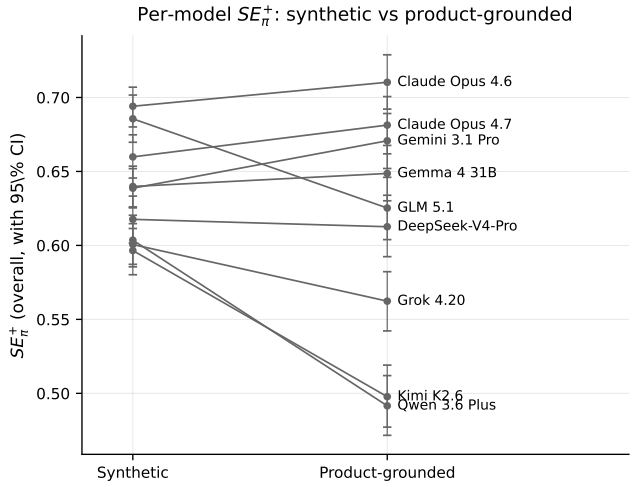



Figure 11: Per-model overall $SE_\pi^+$ in synthetic vs. product-grounded suites with 95% bootstrap CIs. Rank order is largely preserved (Spearman $\rho = 0.90$); the shift is structured: models in the upper half of the synthetic leaderboard tend to gain or hold (exception: `GLM 5.1`), while most lower-half models lose (exception: `GPT-5.5`).

| Model | Synthetic $SE_\pi^+$ | Product-grounded $SE_\pi^+$ | $\Delta$ (PG−Synth) |
|---|---|---|---|
| Claude Opus 4.6 | 0.694 [0.681, 0.708] | 0.710 [0.691, 0.728] | $+0.016$ $[-0.006, +0.039]$ |
| GLM 5.1 | 0.686 [0.670, 0.702] | 0.625 [0.604, 0.645] | $\mathbf{-0.060}$ $[-0.085, -0.033]$ |
| Claude Opus 4.7 | 0.660 [0.646, 0.674] | 0.681 [0.662, 0.700] | $+0.021$ $[-0.003, +0.045]$ |
| Gemma 4 31B | 0.640 [0.626, 0.653] | 0.649 [0.629, 0.666] | $+0.009$ $[-0.014, +0.031]$ |
| Gemini 3.1 Pro | 0.639 [0.626, 0.652] | 0.671 [0.651, 0.689] | $\mathbf{+0.032}$ $[+0.010, +0.054]$ |
| DeepSeek-V4-Pro | 0.618 [0.601, 0.634] | 0.613 [0.593, 0.633] | $-0.005$ $[-0.031, +0.020]$ |
| GPT-5.5 | 0.606 [0.592, 0.620] | 0.618 [0.597, 0.638] | $+0.012$ $[-0.013, +0.036]$ |
| Qwen 3.6 Plus | 0.604 [0.587, 0.619] | 0.492 [0.470, 0.511] | $\mathbf{-0.112}$ $[-0.139, -0.087]$ |
| Grok 4.20 | 0.601 [0.586, 0.615] | 0.562 [0.542, 0.583] | $\mathbf{-0.038}$ $[-0.063, -0.013]$ |
| Kimi K2.6 | 0.597 [0.581, 0.611] | 0.498 [0.476, 0.518] | $\mathbf{-0.099}$ $[-0.126, -0.073]$ |
| GPT-4o-mini | 0.189 [0.176, 0.203] | 0.086 [0.074, 0.099] | $\mathbf{-0.103}$ $[-0.121, -0.085]$ |

Table 15: Per-model overall $SE_\pi^+$ in the synthetic vs. product-grounded suites, with 95% bootstrap CIs ($B = 2000$) on per-episode means. $\Delta$ uses an independent two-sample bootstrap (off-diagonal sample sizes differ); bolded entries have CIs that exclude zero.

## H.3 Deferred Experiment Results

We present the deferred experiment results to supplement main experiment presented in §4. Additional experiments and analysis are given in the following two sections: Appendix H.4 and Appendix H.5.

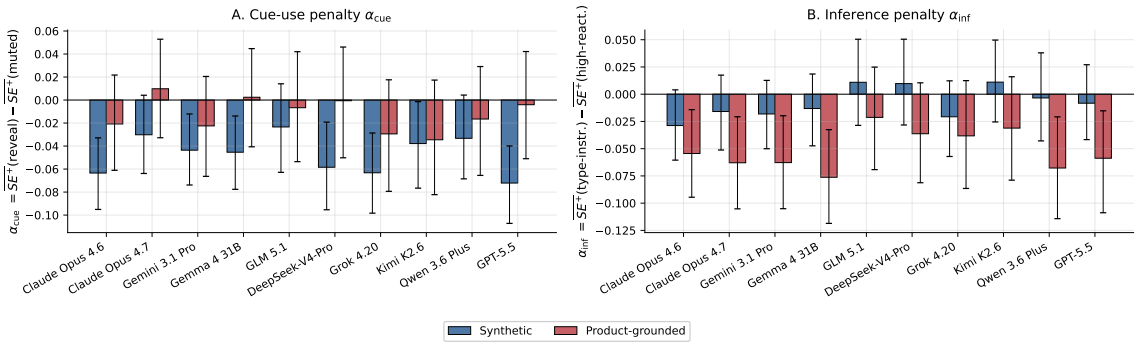



Figure 12: Persistence of the two structural penalties under product grounding (95% bootstrap CIs, $B = 2000$). **A.** $\alpha_{\text{cue}}$ remains negative for 8/11 models but attenuates: the salient product anchor partly insulates agents from verbal-cue over-reaction. **B.** $\alpha_{\text{inf}}$ becomes negative for *all* 11 paired models, and amplifies for 10 of 11 (GPT-4o-mini is the exception), with three PG intervals excluding zero—wide, skewed action ranges make correct latent-type inference more decisive for surplus, so the information–action gap from Finding 3 is sharpened.

### H.3.1 Full-Sweep Results

Tables 17–19 report the full aggregate results, decomposing performance into terminal economic outcomes, opponent-modeling diagnostics, and protocol-compliance statistics. Table 16 further breaks down the full-sweep experiments by regime and counterpart family, covering the three regimes and six counterpart families for all evaluated LLM agents and the three algorithmic baselines.

### H.3.2 Runtime and Inference Cost Comparison

We additionally compare the practical inference profile of the evaluated LLM agents along two axes: interaction runtime, measured as the mean number of negotiation rounds completed per episode, and estimated inference cost per episode. Runtime varies substantially across agents: `GLM 5.1`, `Claude Opus 4.7`, and `Claude Opus 4.6` sustain the longest negotiations on average, while `GPT-4o-mini` terminates much earlier. This runtime measure should be interpreted primarily behaviorally, and only indirectly as a driver of wall-clock latency: it captures how long each agent keeps the bargaining process alive before agreement, rejection, walk-away, or timeout, but does not itself measure elapsed seconds.

Inference cost shows a different ordering. Claude Opus models are the most expensive under the `OpenRouter` pricing assumptions, followed by `GPT-5.4` and `Gemini 3.1 Pro`. Several models with relatively long interaction horizons, such as `GLM 5.1` and `Kimi K2.6`, remain substantially cheaper per episode because their token prices are lower. Conversely, `GPT-4o-mini` is both short-horizon and very low-cost, but also much weaker in benchmark performance.

All cost and runtime measurements here reflect `OpenRouter`-routed inference rather than direct provider inference. `OpenRouter` may route requests through lower-cost backend providers or cloud compute paths, so these estimates should be treated as practical `OpenRouter` deployment statistics rather than canonical native-provider latency or pricing.

### H.3.3 Statistical significance of $\alpha_{\text{cue}}$ and $\alpha_{\text{inf}}$ on the synthetic suite

We accompany the point estimates in Fig. 4 with 95% percentile bootstrap CIs and across-model significance tests for the two structural penalties used in Findings 2 and 3. For each model we resample per-episode $SE^+_\pi$ values within each family pair ($B = 2000$) and form a two-sample bootstrap CI on the contrast

$$\alpha_{\text{cue}} = \overline{SE^+}(\textsc{Candid}, \textsc{Expressive}) - \overline{SE^+}(\textsc{Taciturn}, \textsc{Strategic}),$$

$$\alpha_{\text{inf}} = \overline{SE^+}(\textsc{Candid}, \textsc{Taciturn}) - \overline{SE^+}(\textsc{Expressive}, \textsc{Strategic}),$$

where negative values indicate the cue-use penalty (F2) and the inference penalty (F3) respectively. Across the 13 evaluated LLMs we also report a sign test against $\Pr(\alpha < 0) = 0.5$ and an exact one-sided paired Wilcoxon signed-rank test against $\text{median}(\alpha) = 0$.

**Results.** $\alpha_{\text{cue}}$ is negative for all 13 models (sign-test $p = 0.0001$); 9 of 13 individual CIs strictly exclude zero, and the exact Wilcoxon test against the negative direction reaches the precision floor of the exact null distribution ($p_{<0} = 0.0001$). The four models whose CIs cross zero (`GLM 5.1`, `Claude Opus 4.7`, `Qwen 3.6`

Table 16: Per-family and per-regime breakdown in the bilateral price-negotiation instantiation of TERMS-BENCH with 95% CIs ($n$=100 episodes per cell). In Panels A–B, entries are $SE_\pi^+$. In Panel C, entries are critical-violation rates (%). Bold: best $SE_\pi^+$ within each feasible regime–family column. Red: nonzero protocol violation rate.

**Panel A: Overlap regime**

| Agent | Adv. | Expr. | Candid | Stoch. | Strat. | Tac. |
|---|---|---|---|---|---|---|
| Claude 4.7 | .578±.050 | .648±.048 | .614±.048 | .694±.050 | .654±.049 | .675±.051 |
| Grok 4 | .519±.056 | .567±.049 | .540±.051 | .639±.048 | .610±.048 | .652±.046 |
| Gemini 3.1 | .564±.050 | .617±.045 | .619±.046 | .676±.042 | **.690**±.040 | .669±.048 |
| DeepSeek V4 | .516±.061 | .609±.053 | .574±.056 | .590±.061 | .668±.056 | .665±.055 |
| Qwen3.6 | .508±.055 | .579±.059 | .530±.056 | .642±.051 | .560±.059 | .642±.057 |
| Kimi K2.6 | .501±.055 | .557±.054 | .558±.055 | .584±.059 | .601±.051 | .654±.053 |
| GPT-5.4 | .455±.054 | .530±.056 | .505±.055 | .596±.053 | .526±.059 | .517±.063 |
| Doubao 2.0 | .445±.048 | .492±.047 | .505±.048 | .521±.049 | .556±.045 | .568±.045 |
| GLM-5.1 | **.590**±.062 | **.662**±.057 | **.663**±.057 | **.725**±.054 | .690±.058 | **.678**±.057 |
| GPT-4o-mini | .171±.046 | .167±.042 | .190±.046 | .137±.040 | .234±.049 | .230±.050 |
| Fixed 30% | .400±.050 | .371±.045 | .375±.053 | .389±.054 | .387±.051 | .377±.050 |
| Fixed 10% | .246±.050 | .264±.036 | .223±.039 | .351±.054 | .301±.046 | .311±.050 |
| Fixed 1% | .191±.038 | .253±.035 | .213±.037 | .320±.055 | .275±.043 | .282±.044 |

**Panel B: Urgency-shift regime**

| Agent | Adv. | Expr. | Candid | Stoch. | Strat. | Tac. |
|---|---|---|---|---|---|---|
| Claude 4.7 | .613±.048 | .692±.048 | .643±.047 | **.718**±.053 | .695±.048 | .694±.050 |
| Grok 4 | .570±.046 | .618±.052 | .563±.047 | .650±.054 | .661±.048 | .618±.047 |
| Gemini 3.1 | .566±.045 | .649±.043 | .621±.045 | .672±.051 | .673±.043 | .648±.046 |
| DeepSeek V4 | .540±.056 | .585±.057 | .668±.049 | .660±.055 | .671±.052 | .665±.053 |
| Qwen3.6 | .589±.049 | .657±.050 | .572±.050 | .693±.056 | .615±.057 | .655±.048 |
| Kimi K2.6 | .540±.056 | .641±.053 | .596±.046 | .678±.055 | .607±.051 | .642±.053 |
| GPT-5.4 | .477±.051 | .551±.058 | .515±.052 | .613±.056 | .541±.052 | .550±.062 |
| Doubao 2.0 | .499±.047 | .512±.044 | .485±.043 | .584±.053 | .575±.046 | .523±.045 |
| GLM-5.1 | **.650**±.047 | **.705**±.055 | **.698**±.047 | .711±.063 | **.696**±.055 | **.758**±.047 |
| GPT-4o-mini | .187±.051 | .223±.046 | .149±.041 | .150±.038 | .227±.047 | .208±.050 |
| Fixed 30% | .438±.054 | .369±.048 | .379±.050 | .419±.052 | .351±.046 | .387±.051 |
| Fixed 10% | .255±.047 | .308±.045 | .253±.039 | .370±.049 | .297±.042 | .296±.042 |
| Fixed 1% | .234±.038 | .297±.042 | .253±.039 | .367±.049 | .291±.038 | .298±.042 |

**Panel C: No-deal regime — critical-violation rate (%)**

| Agent | Adv. | Expr. | Candid | Stoch. | Strat. | Tac. |
|---|---|---|---|---|---|---|
| Claude 4.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Grok 4 | 4.0±3.8 | 7.0±5.0 | 5.0±4.3 | 3.0±3.3 | 3.0±3.3 | 5.0±4.3 |
| Gemini 3.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| DeepSeek V4 | 3.0±3.3 | 2.0±2.7 | 1.0±2.0 | 2.0±2.7 | 0.0 | 3.0±3.3 |
| Qwen3.6 | 6.0±4.7 | 10.0±5.9 | 3.0±3.3 | 7.0±5.0 | 7.0±5.0 | 4.0±3.8 |
| Kimi K2.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| GPT-5.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Doubao 2.0 | 0.0 | 0.0 | 1.0±2.0 | 0.0 | 0.0 | 0.0 |
| GLM-5.1 | 5.0±4.3 | 5.0±4.3 | 5.0±4.3 | 2.0±2.7 | 2.0±2.7 | 5.0±4.3 |
| GPT-4o-mini | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Fixed 30% | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Fixed 10% | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Fixed 1% | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

`Plus`, `GPT-5.4`) all carry the same sign as the population. The cue-use penalty in F2 is therefore robust at both the per-model and the across-model level.

| Agent | $SE_\pi^+ \uparrow$ | $\text{AGR}_\pi^+ \uparrow$ | $CSE_\pi^+ \uparrow$ | $\text{FAGR}_\pi^- \downarrow$ | $\bar{u} \uparrow$ | $u^\star$ | Gap $\downarrow$ | Oracle $\uparrow$ | $\text{AGR}_{\text{all}} \uparrow$ | $\mathbb{E}[u \mid \text{deal}] \uparrow$ | Safe$^-$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Claude Opus 4.7 | $0.660_{\pm 0.014}$ | $98.2_{\pm 0.8}$ | $0.672_{\pm 0.014}$ | 0.00 | $11.09_{\pm 0.50}$ | $15.00_{\pm 0.49}$ | $3.90_{\pm 0.36}$ | $74.1_{\pm 2.3}$ | $65.4_{\pm 2.2}$ | $16.95_{\pm 0.50}$ | 100.0 |
| Grok 4 | $0.601_{\pm 0.015}$ | $99.1_{\pm 0.5}$ | $0.606_{\pm 0.014}$ | 0.00 | $10.03_{\pm 0.46}$ | $15.00_{\pm 0.49}$ | $4.96_{\pm 0.36}$ | $67.1_{\pm 2.1}$ | $66.1_{\pm 2.2}$ | $15.19_{\pm 0.48}$ | 100.0 |
| Gemini-3.1-Pro | $0.639_{\pm 0.013}$ | $99.7_{\pm 0.3}$ | $0.641_{\pm 0.013}$ | 0.00 | $10.58_{\pm 0.47}$ | $15.00_{\pm 0.49}$ | $4.42_{\pm 0.35}$ | $70.7_{\pm 2.1}$ | $66.4_{\pm 2.2}$ | $15.92_{\pm 0.47}$ | 100.0 |
| DeepSeek-V4-Pro | $0.618_{\pm 0.016}$ | $97.5_{\pm 0.9}$ | $0.633_{\pm 0.016}$ | 0.00 | $10.53_{\pm 0.50}$ | $15.00_{\pm 0.49}$ | $4.47_{\pm 0.39}$ | $70.3_{\pm 2.4}$ | $65.0_{\pm 2.2}$ | $16.19_{\pm 0.54}$ | 100.0 |
| Qwen3.6-Plus | $0.604_{\pm 0.016}$ | $98.2_{\pm 0.7}$ | $0.614_{\pm 0.015}$ | $0.17_{\pm 0.33}$ | $10.31_{\pm 0.49}$ | $15.00_{\pm 0.49}$ | $4.69_{\pm 0.39}$ | $68.9_{\pm 2.4}$ | $65.6_{\pm 2.2}$ | $15.73_{\pm 0.53}$ | $99.8_{\pm 0.3}$ |
| Kimi-K2.6 | $0.597_{\pm 0.016}$ | $97.1_{\pm 1.0}$ | $0.614_{\pm 0.015}$ | 0.00 | $10.17_{\pm 0.48}$ | $15.00_{\pm 0.49}$ | $4.83_{\pm 0.39}$ | $67.9_{\pm 2.3}$ | $64.7_{\pm 2.2}$ | $15.71_{\pm 0.52}$ | 100.0 |
| GPT-5.4 | $0.531_{\pm 0.016}$ | $99.4_{\pm 0.4}$ | $0.535_{\pm 0.016}$ | 0.00 | $9.04_{\pm 0.46}$ | $15.00_{\pm 0.49}$ | $5.95_{\pm 0.40}$ | $60.4_{\pm 2.3}$ | $66.3_{\pm 2.2}$ | $13.64_{\pm 0.53}$ | 100.0 |
| Doubao-Seed-2.0-Pro | $0.522_{\pm 0.014}$ | $99.9_{\pm 0.2}$ | $0.523_{\pm 0.014}$ | 0.00 | $8.61_{\pm 0.40}$ | $15.00_{\pm 0.49}$ | $6.38_{\pm 0.35}$ | $57.6_{\pm 1.9}$ | $66.6_{\pm 2.2}$ | $12.93_{\pm 0.42}$ | 100.0 |
| GLM-5.1 | $0.686_{\pm 0.016}$ | $95.1_{\pm 1.2}$ | $0.721_{\pm 0.014}$ | 0.00 | $11.70_{\pm 0.53}$ | $15.00_{\pm 0.49}$ | $3.30_{\pm 0.39}$ | $78.2_{\pm 2.5}$ | $63.4_{\pm 2.2}$ | $18.45_{\pm 0.53}$ | 100.0 |
| GPT-4o-mini | $0.189_{\pm 0.013}$ | $52.2_{\pm 2.8}$ | $0.363_{\pm 0.016}$ | 0.00 | $3.37_{\pm 0.27}$ | $15.00_{\pm 0.49}$ | $11.63_{\pm 0.46}$ | $22.4_{\pm 1.7}$ | $34.8_{\pm 2.2}$ | $9.69_{\pm 0.49}$ | 100.0 |
| Fixed 30% | $0.387_{\pm 0.015}$ | $99.9_{\pm 0.2}$ | $0.387_{\pm 0.015}$ | 0.00 | $6.50_{\pm 0.36}$ | $15.00_{\pm 0.49}$ | $8.49_{\pm 0.41}$ | $43.4_{\pm 2.0}$ | $66.6_{\pm 2.2}$ | $9.76_{\pm 0.44}$ | 100.0 |
| Fixed 10% | $0.290_{\pm 0.013}$ | $94.5_{\pm 1.3}$ | $0.307_{\pm 0.013}$ | 0.00 | $5.08_{\pm 0.32}$ | $15.00_{\pm 0.49}$ | $9.92_{\pm 0.43}$ | $33.9_{\pm 1.8}$ | $63.0_{\pm 2.2}$ | $8.06_{\pm 0.42}$ | 100.0 |
| Fixed 1% | $0.273_{\pm 0.012}$ | $92.2_{\pm 1.5}$ | $0.296_{\pm 0.013}$ | 0.00 | $4.77_{\pm 0.30}$ | $15.00_{\pm 0.49}$ | $10.23_{\pm 0.42}$ | $31.8_{\pm 1.7}$ | $61.5_{\pm 2.2}$ | $7.75_{\pm 0.39}$ | 100.0 |

Table 17: Aggregate terminal and oracle-reference performance in the bilateral price-negotiation instantiation of TERMS-BENCH with 95% CIs ($n$=1,800 episodes per agent). CIs are normal-approximation half-widths for means and binomial half-widths for proportions. $u^\star$ is the per-episode optimal utility under a full-information oracle (cell-mean), averaged across all episodes (with $u^\star$=0 on no-deal); *Gap* $= u^\star - \bar{u}$; *Oracle* $= 100\bar{u}/u^\star$. Percent-valued columns reported in pp.

| Agent | $BE_{\text{type}} \downarrow$ | $BE_r \downarrow$ | $BE_\kappa \downarrow$ | $\text{Brier}_\eta \downarrow$ | Stance acc. $\uparrow$ | $BE_{\text{joint}}^{\ell_2} \downarrow$ | Type–stance mism. $\downarrow$ |
|---|---|---|---|---|---|---|---|
| Claude Opus 4.7 | $0.229_{\pm 0.006}$ | $0.118_{\pm 0.014}$ | $0.243_{\pm 0.015}$ | $0.325_{\pm 0.023}$ | $45.8_{\pm 4.3}$ | $0.296_{\pm 0.012}$ | $0.301_{\pm 0.012}$ |
| Grok 4 | $0.212_{\pm 0.006}$ | $0.112_{\pm 0.009}$ | $0.200_{\pm 0.008}$ | $0.324_{\pm 0.019}$ | $44.5_{\pm 4.1}$ | $0.251_{\pm 0.010}$ | $0.289_{\pm 0.012}$ |
| Gemini-3.1-Pro | $0.271_{\pm 0.012}$ | $0.117_{\pm 0.007}$ | $0.344_{\pm 0.020}$ | $0.352_{\pm 0.033}$ | $45.8_{\pm 4.7}$ | $0.386_{\pm 0.017}$ | $0.335_{\pm 0.016}$ |
| DeepSeek-V4-Pro | $0.228_{\pm 0.005}$ | $0.125_{\pm 0.013}$ | $0.223_{\pm 0.010}$ | $0.337_{\pm 0.017}$ | $42.9_{\pm 3.6}$ | $0.283_{\pm 0.009}$ | $0.306_{\pm 0.011}$ |
| Qwen3.6-Plus | $0.237_{\pm 0.006}$ | $0.125_{\pm 0.015}$ | $0.246_{\pm 0.014}$ | $0.339_{\pm 0.017}$ | $43.3_{\pm 3.3}$ | $0.304_{\pm 0.009}$ | $0.313_{\pm 0.012}$ |
| Kimi-K2.6 | $0.236_{\pm 0.009}$ | $0.120_{\pm 0.012}$ | $0.243_{\pm 0.013}$ | $0.345_{\pm 0.030}$ | $43.3_{\pm 5.4}$ | $0.298_{\pm 0.010}$ | $0.310_{\pm 0.017}$ |
| GPT-5.4 | $0.242_{\pm 0.009}$ | $0.126_{\pm 0.018}$ | $0.262_{\pm 0.017}$ | $0.337_{\pm 0.031}$ | $44.0_{\pm 5.7}$ | $0.318_{\pm 0.014}$ | $0.316_{\pm 0.017}$ |
| Doubao-Seed-2.0-Pro | $0.247_{\pm 0.009}$ | $0.121_{\pm 0.013}$ | $0.256_{\pm 0.011}$ | $0.364_{\pm 0.028}$ | $42.0_{\pm 4.7}$ | $0.307_{\pm 0.010}$ | $0.319_{\pm 0.016}$ |
| GLM-5.1 | $0.218_{\pm 0.008}$ | $0.106_{\pm 0.007}$ | $0.221_{\pm 0.011}$ | $0.326_{\pm 0.027}$ | $44.8_{\pm 5.8}$ | $0.268_{\pm 0.009}$ | $0.293_{\pm 0.018}$ |
| GPT-4o-mini | $0.251_{\pm 0.005}$ | $0.215_{\pm 0.007}$ | $0.192_{\pm 0.008}$ | $0.345_{\pm 0.011}$ | $38.9_{\pm 3.2}$ | $0.311_{\pm 0.009}$ | $0.340_{\pm 0.013}$ |
| Fixed 30% | – | – | – | – | – | – | – |
| Fixed 10% | – | – | – | – | – | – | – |
| Fixed 1% | – | – | – | – | – | – | – |

Table 18: Aggregate opponent-modeling diagnostics in the bilateral price-negotiation instantiation of TERMS-BENCH with 95% CIs (between-cell half-widths over 18 regime×family cells, $n$=100 episodes per cell). Lower is better for belief-error metrics and Brier score; higher is better for stance accuracy. Fixed-concession baselines do not produce belief estimates.

$\alpha_{\text{inf}}$ is negative for 10 of 13 models; the sign test is marginally significant ($p = 0.0461$) and no individual CI excludes zero. The exact Wilcoxon test against the negative direction, which uses ranks rather than signs alone, is significant ($p_{<0} = 0.0085$). The inference penalty in F3 is therefore a population-level effect: the rank-based test resolves it cleanly while the per-model bootstrap is underpowered to do so at the single-model level.

For both results, refer to Table 20.

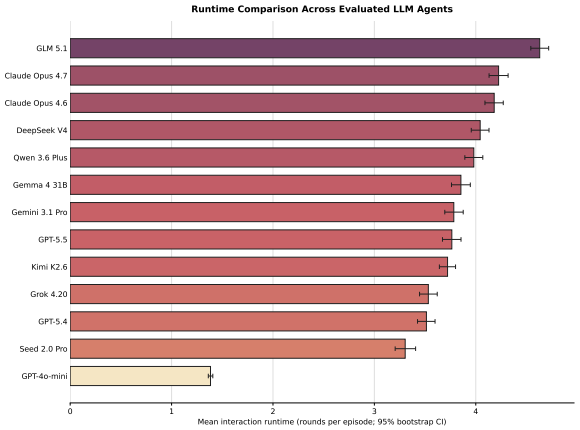
## H.4 Reverse Urgency-Shift Direction Experiment

Main paper results use the counterpart-more-urgent direction of the urgency-shift regime. Here we report the reverse direction, where the evaluated agent is more time-pressured than the counterpart. We conduct the experiments on three representing models spanning the performance levels in the main experiment (Table 2): `Claude Opus 4.6` (top); `Grok 4.20` (mid); and `GPT-4o-mini` (low). Episode construction, common seeds, inference settings, role/opener decomposition, and reporting conventions are otherwise held fixed. This auxiliary check asks whether the headline ordering reflects the sign of the urgency asymmetry, or instead a more stable model-level difference in negotiation behavior. The results with 95% bootstrap CIs are reported in Table 21.

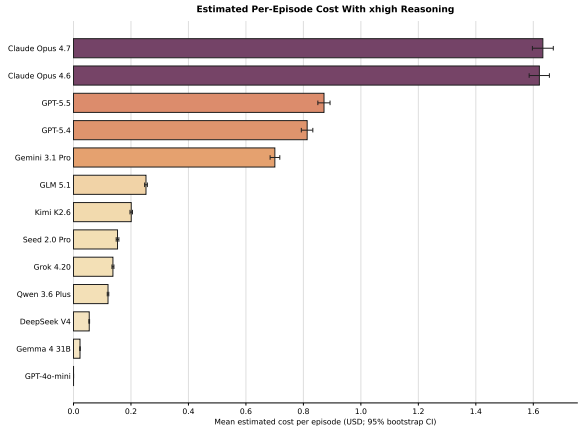
Across models, reversing the urgency asymmetry leaves agreement rates essentially unchanged but lowers surplus efficiency for the stronger agents. GPT-4o-mini is nearly invariant: $SE_\pi^+ = 0.192\,[0.173, 0.211]$ under agent pressure versus $0.191\,[0.173, 0.209]$ in the main direction. Grok 4.20 remains high-agreement but drops from $0.614\,[0.593, 0.634]$ to $0.595\,[0.575, 0.616]$. Claude Opus 4.6 shows the largest directional sensitivity, from $0.738\,[0.720, 0.755]$ to $0.689\,[0.670, 0.709]$.

| Agent | CritViol | Bound | Res. | Invalid | Mono. | Budget | Any viol. | Agent acc. | Counter acc. | Agent rej. | Walkaway | Timeout |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Claude Opus 4.7 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $13.4_{\pm 1.6}$ | $52.1_{\pm 2.3}$ | $2.6_{\pm 0.7}$ | $31.9_{\pm 2.2}$ | 0.0 |
| Grok 4 | $1.50_{\pm 0.56}$ | 0.00 | $1.50_{\pm 0.56}$ | 0.00 | 0.00 | 0.00 | $1.50_{\pm 0.56}$ | $14.2_{\pm 1.6}$ | $51.9_{\pm 2.3}$ | $15.3_{\pm 1.7}$ | $18.6_{\pm 1.8}$ | 0.0 |
| Gemini-3.1-Pro | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $16.1_{\pm 1.7}$ | $50.3_{\pm 2.3}$ | $2.6_{\pm 0.7}$ | $30.9_{\pm 2.1}$ | 0.0 |
| DeepSeek-V4-Pro | $0.61_{\pm 0.36}$ | 0.00 | $0.61_{\pm 0.36}$ | 0.00 | 0.00 | 0.00 | $0.61_{\pm 0.36}$ | $19.5_{\pm 1.8}$ | $45.5_{\pm 2.3}$ | $6.1_{\pm 1.1}$ | $28.9_{\pm 2.1}$ | $0.1_{\pm 0.1}$ |
| Qwen3.6-Plus | $2.06_{\pm 0.66}$ | 0.00 | $2.06_{\pm 0.66}$ | 0.00 | 0.00 | 0.00 | $2.06_{\pm 0.66}$ | $18.2_{\pm 1.8}$ | $47.3_{\pm 2.3}$ | $5.2_{\pm 1.0}$ | $29.2_{\pm 2.1}$ | 0.0 |
| Kimi-K2.6 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $19.7_{\pm 1.8}$ | $45.1_{\pm 2.3}$ | $18.0_{\pm 1.8}$ | $17.3_{\pm 1.7}$ | 0.0 |
| GPT-5.4 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $21.4_{\pm 1.9}$ | $44.8_{\pm 2.3}$ | $6.4_{\pm 1.1}$ | $27.3_{\pm 2.1}$ | 0.0 |
| Doubao-Seed-2.0-Pro | $0.06_{\pm 0.11}$ | 0.00 | $0.06_{\pm 0.11}$ | 0.00 | 0.00 | 0.00 | $0.06_{\pm 0.11}$ | $12.4_{\pm 1.5}$ | $54.2_{\pm 2.3}$ | $5.4_{\pm 1.0}$ | $28.0_{\pm 2.1}$ | 0.0 |
| GLM-5.1 | $1.33_{\pm 0.53}$ | 0.00 | $1.33_{\pm 0.53}$ | 0.00 | 0.00 | 0.00 | $1.33_{\pm 0.53}$ | $14.6_{\pm 1.6}$ | $48.8_{\pm 2.3}$ | $6.4_{\pm 1.1}$ | $30.2_{\pm 2.1}$ | 0.0 |
| GPT-4o-mini | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $22.8_{\pm 1.9}$ | $11.9_{\pm 1.5}$ | $65.2_{\pm 2.2}$ | 0.0 | 0.0 |
| Fixed 30% | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $52.5_{\pm 2.3}$ | $14.1_{\pm 1.6}$ | 0.0 | $32.3_{\pm 2.2}$ | $1.1_{\pm 0.5}$ |
| Fixed 10% | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $61.4_{\pm 2.2}$ | $1.6_{\pm 0.6}$ | 0.0 | $36.1_{\pm 2.2}$ | $0.9_{\pm 0.4}$ |
| Fixed 1% | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $61.4_{\pm 2.2}$ | $0.1_{\pm 0.1}$ | 0.0 | $38.4_{\pm 2.2}$ | $0.1_{\pm 0.2}$ |

Table 19: Aggregate protocol-compliance and interaction-outcome rates in the bilateral price-negotiation instantiation of TERMS-BENCH with 95% CIs (binomial half-widths, $n$=1,800). All entries are percentages. Red shading marks nonzero protocol-violation rates.




(a) Mean interaction runtime across evaluated LLM agents on the bilateral price-negotiation instantiation of TERMS-BENCH. Bars show the mean number of negotiation rounds played per episode, with 95% bootstrap confidence intervals across episodes. These runtime statistics reflect inference through `OpenRouter`, and should not be interpreted as native-provider latency: `OpenRouter` may route requests through different backend providers or lower-cost cloud compute paths, so observed behavior can differ from running the same model directly with its own provider.




(b) Estimated mean per-episode inference cost across evaluated LLM agents. Bars show estimated USD cost per episode with 95% bootstrap confidence intervals, using current `OpenRouter` token prices and completion token estimates. For reasoning-capable models, costs assume `xhigh` reasoning effort, modeled as 15,200 reasoning tokens per LLM call under the benchmark's 16,000-token completion budget. We note that these estimates represent `OpenRouter`-routed inference prices, not necessarily native-provider pricing.

Figure 13: Runtime and estimated inference cost across evaluated LLM agents on the bilateral price-negotiation instantiation of TERMS-BENCH.

Thus, the reverse-direction analysis preserves the broad ranking while showing that agent-side time pressure compresses the top-end surplus advantage. The effect is not primarily an agreement-rate effect: all three models retain nearly the same $AGR^{+}_{\pi}$, while the change appears in realized surplus and conditional deal quality.

## H.5 Strategic Profile Decomposition: Commercial Role, Opener-Role, and Per-Family Performance

In this section, we provide more details on agent-specific strategic profile decomposition and supplemental findings to Finding 5 in §4. For each agent we formally define and report three trace-level quantities. We follow the same notation convention as in §2.3, and we denote the agreed-feasible subset $\mathcal{A}^{+} = \{i \in \mathcal{I}^{+} : f_i \neq \bot\}$:

- **Agent-closer rate** $\rho_\pi$ – the fraction of feasible episodes whose terminating move is the agent's ACCEPT,

$$\rho_\pi = \frac{1}{|\mathcal{I}^{+}|} \sum_{i \in \mathcal{I}^{+}} \mathbf{1}\{d^{A}_{T_i} = \text{ACCEPT}\},$$

| Model | $\alpha_{\text{cue}}$ [95% CI] | $\alpha_{\text{inf}}$ [95% CI] |
|---|---|---|
| Claude Opus 4.6 | $\mathbf{-0.063}$ $[-0.095, -0.032]$ | $-0.029$ $[-0.061, +0.002]$ |
| GLM 5.1 | $-0.023$ $[-0.061, +0.015]$ | $+0.011$ $[-0.029, +0.051]$ |
| Claude Opus 4.7 | $-0.030$ $[-0.065, +0.005]$ | $-0.016$ $[-0.050, +0.019]$ |
| Gemma 4 31B | $\mathbf{-0.045}$ $[-0.079, -0.010]$ | $-0.013$ $[-0.046, +0.020]$ |
| Gemini 3.1 Pro | $\mathbf{-0.044}$ $[-0.074, -0.012]$ | $-0.018$ $[-0.048, +0.014]$ |
| DeepSeek-V4-Pro | $\mathbf{-0.058}$ $[-0.097, -0.021]$ | $+0.010$ $[-0.028, +0.047]$ |
| Qwen 3.6 Plus | $-0.033$ $[-0.074, +0.005]$ | $-0.003$ $[-0.043, +0.036]$ |
| Grok 4.20 | $\mathbf{-0.063}$ $[-0.097, -0.028]$ | $-0.021$ $[-0.058, +0.015]$ |
| Kimi K2.6 | $\mathbf{-0.038}$ $[-0.075, -0.001]$ | $+0.011$ $[-0.026, +0.049]$ |
| GPT-5.4 | $-0.009$ $[-0.051, +0.032]$ | $-0.016$ $[-0.056, +0.023]$ |
| GPT-5.5 | $\mathbf{-0.072}$ $[-0.106, -0.038]$ | $-0.008$ $[-0.043, +0.025]$ |
| Doubao 2.0 Pro | $\mathbf{-0.057}$ $[-0.090, -0.024]$ | $-0.014$ $[-0.046, +0.018]$ |
| GPT-4o-mini | $\mathbf{-0.042}$ $[-0.076, -0.010]$ | $-0.018$ $[-0.049, +0.015]$ |
| *Across-model,* $n = 13$ | 13/13 $< 0$; 9/13 sig.; Wilcoxon $p_{<0} = 0.0001$ | 10/13 $< 0$; 0/13 sig.; Wilcoxon $p_{<0} = 0.0085$ |

Table 20: Per-model $\alpha_{\text{cue}}$ and $\alpha_{\text{inf}}$ on the synthetic paper suite, with 95% two-sample percentile bootstrap CIs ($B = 2000$). Bolded point estimates have CIs that exclude zero. The bottom row reports across-model statistics: count of negative point estimates, count of CIs strictly excluding zero, and the one-sided exact Wilcoxon signed-rank $p$-value against $H_0$: $\text{median}(\alpha) = 0$ with alternative $\text{median}(\alpha) < 0$.

| Model | Direction | $n$ | $SE_\pi^+$ | $AGR_\pi^+$ | $CSE_\pi^+$ |
|---|---|---|---|---|---|
| GPT-4o-mini | Counterpart more urgent | 600 | 0.191 [0.173, 0.209] | 0.530 [0.492, 0.570] | 0.360 [0.337, 0.383] |
| GPT-4o-mini | Agent more urgent | 600 | 0.192 [0.173, 0.211] | 0.537 [0.497, 0.575] | 0.357 [0.334, 0.380] |
| Grok 4.20 | Counterpart more urgent | 600 | 0.614 [0.593, 0.634] | 0.995 [0.988, 1.000] | 0.617 [0.596, 0.637] |
| Grok 4.20 | Agent more urgent | 578 | 0.595 [0.575, 0.616] | 0.998 [0.995, 1.000] | 0.596 [0.576, 0.616] |
| Claude Opus 4.6 | Counterpart more urgent | 600 | 0.738 [0.720, 0.755] | 0.997 [0.992, 1.000] | 0.740 [0.722, 0.758] |
| Claude Opus 4.6 | Agent more urgent | 541 | 0.689 [0.670, 0.709] | 0.996 [0.991, 1.000] | 0.691 [0.673, 0.711] |

Table 21: Urgency-shift direction reversal. Metrics are reported with 95% percentile bootstrap confidence intervals over episodes ($B = 2000$).

where $T_i$ is the terminating round and $d^A_{T_i}$ is the agent's final decision. $\rho_\pi$ captures *who initiates closure*: high values mean the agent ends the negotiation; low values mean the counterpart does.

- **Closing-side surplus efficiency** $\sigma_\pi$ denotes the agent's share of the ZOPA on the deals it actually closes,

$$\sigma_\pi = \frac{1}{|\mathcal{A}^+|} \sum_{i \in \mathcal{A}^+} \frac{u_A(f_i)}{\Delta_i} \;=\; CSE_\pi^+\big|_{\mathcal{A}^+}.$$

  $\sigma_\pi \in [0, 1]$, with 1 corresponding to full extraction. It captures *closing quality*, the price/utility-side fingerprint of the agent's strategy conditional on agreement.

- **Conditional utility** $\text{cond}\, U$ – raw mean utility on agreed-feasible episodes, in price units,

$$\text{cond}\, U = \frac{1}{|\mathcal{A}^+|} \sum_{i \in \mathcal{A}^+} u_A(f_i).$$

  $\text{cond}\, U$ is a unit-preserving companion to $\sigma_\pi$ that does not normalise away ZOPA-width heterogeneity, and is the quantity annotated as `cond.` in Fig. 5(left).

$\rho_\pi$ and $\sigma_\pi$ encode orthogonal dimensions of behaviour: $\rho_\pi$ is a *frequency* on $\mathcal{I}^+$, $\sigma_\pi$ is a *quality* on $\mathcal{A}^+$. Together with the per-agent trajectory coefficient $\alpha_n$ (slope of the offer trajectory across rounds, reported in Fig. 5(left)) they yield the five typology cells reported below. Confidence intervals follow the standard bootstrap conventions: $\rho_\pi$ uses the Wilson interval (it is a proportion on $\mathcal{I}^+$); $\sigma_\pi$ and $\text{cond}\, U$ use $B$=2000 percentile bootstrap CIs over per-episode values.

Table 23 reports the closing-side fingerprint $(\rho_\pi, \sigma_\pi, \text{cond}\, U)$ for all 13 LLMs on feasible regimes. The two-axis decomposition partitions the roster into five profiles consistent with the trajectory shapes in Figure 5(left): *anchor-and-hold* ($\rho_\pi \geq 0.75$, $\sigma_\pi \geq 0.64$: `GLM 5.1`, `Claude Opus 4.6`, `Claude Opus 4.7`, `Gemini 3.1 Pro`); *mid/balanced* ($0.69 \leq \rho_\pi \leq 0.76$, $0.60 \leq \sigma_\pi < 0.65$: `Gemma 4 31B`, `DeepSeek-V4-Pro`, `Qwen 3.6 Plus`, `Kimi K2.6`, `GPT-5.5`; the latter sits at the upper $\rho_\pi$ boundary with mid-tier $\sigma_\pi$); *anchor-and-concede* ($\rho_\pi \geq 0.78$, $\sigma_\pi < 0.65$: `Grok 4.20`, `Doubao 2.0 Pro`); *accepter* ($\rho_\pi < 0.69$, $\sigma_\pi < 0.55$: `GPT-5.4`); and *refuser* ($\rho_\pi < 0.40$: `GPT-4o-mini`, where $\sigma_\pi$ is computed on the small agreed subset).

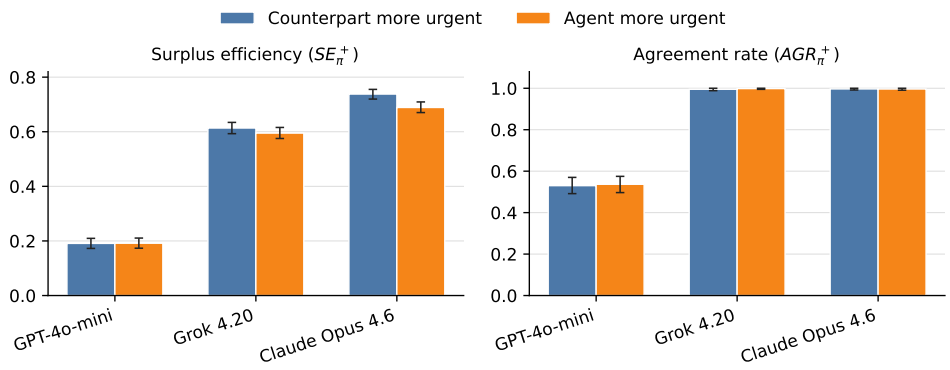



Figure 14: Effect of reversing the urgency-shift direction. The main condition makes the counterpart more urgent; the auxiliary condition makes the evaluated agent more urgent. Error bars show 95% percentile bootstrap confidence intervals over episodes ($B = 2000$).

### H.5.1 Experiment Results

**Opener-role decomposition.** The strategic profile is most visible when the agent makes the *first* move (the agent-opens cells), because the agent's anchor is unconstrained and reveals strategic intent directly. When the counterpart opens, the agent's first action is a reactive concession rather than a free anchor, so one might expect the closing-side $\sigma_\pi$ to compress across agents in counterpart-opens. We do not see this: the cross-model standard deviation on $\sigma_\pi$ is essentially identical between the two opener-role cells (0.12 in counterpart-opens vs 0.13 in agent-opens; cross-model range 0.43 vs 0.46). Each agent's reactive concession profile is stable enough that the closing-side signature persists even when the opening anchor is taken away. The typology is therefore agent-driven at the level of the opening move, but the closing-side surplus signature it produces is robust across both opener-role cells.

**Buyer and seller role decomposition** The closing-side fingerprint is systematically asymmetric in the agent's role. Figure 15 shows per-model $\sigma_\pi$ split by role (left), and the per-model $\Delta\sigma_\pi = \sigma_\pi(\text{seller}) - \sigma_\pi(\text{buyer})$ with 95% bootstrap CIs (right); Table 22 reports the same with $\Delta SE_\pi^+$ and $\Delta AGR_\pi^+$ companions.

1. **Near-universal seller advantage on closing surplus.** 12 of 13 LLMs extract more closing surplus as seller than as buyer (median $\Delta\sigma_\pi = +0.037$; sign-test $p = 0.0017$, exact paired Wilcoxon $p = 0.0061$); `GPT-4o-mini` is the lone exception, with $\Delta\sigma_\pi = -0.063$. Among the 12 positive agents the magnitudes are highly model-dependent, from $+0.014$ (`Grok 4.20`) to $+0.136$ (`Claude Opus 4.7`), and 10 of the 13 individual $\Delta\sigma_\pi$ CIs exclude zero (9 in the seller-favoring direction, plus `GPT-4o-mini` in the buyer-favoring direction).
2. **Compensating agreement-rate dip.** In the opposite direction, sellers close fewer deals: median $\Delta AGR_\pi^+ = -0.010$, with $0/13$ models showing seller $>$ buyer (paired Wilcoxon $p = 0.0005$). The dip is most pronounced for the strongest anchor-and-hold agents—`GLM 5.1` reaches feasible $AGR_\pi^+ = 1.000$ as buyer but 0.902 as seller, and `Claude Opus 4.7` drops from 0.998 to 0.965—suggesting that the seller-side surplus advantage is partly bought by walking away from offers the same agent would accept as buyer.
3. **Net effect on $SE_\pi^+$ remains positive for 12 of 13.** The seller-side $\sigma_\pi$ gain dominates the agreement-rate drop in $SE_\pi^+$ terms for the same 12 of 13 agents (median $\Delta SE_\pi^+ = +0.032$, paired Wilcoxon $p = 0.0100$); the lone exception is `GPT-4o-mini` ($-0.065$). Among the 12 agents that do show a seller advantage, the heterogeneity in magnitude tracks opening-price aggressiveness: the three agents with the largest $\Delta\sigma_\pi$ (`Claude Opus 4.7`, `GLM 5.1`, `Claude Opus 4.6`) are also the three with the highest seller-agent-opens openings ($\geq 78$), while `Grok 4.20` and `Doubao 2.0 Pro`—the anchor-and-concede pair—show the smallest asymmetries because they concede quickly regardless of role. We cannot from these data disambiguate whether the asymmetry reflects (a) LLM priors over "seller asks high" framings stronger than the symmetric "buyer offers low" framing, or (b) inherent geometric asymmetry in how $p_{\min}$ and $p_{\max}$ bound anchor space; the empirical pattern is consistent with both.
4. **The typology is preserved across role.** Despite the universal $\sigma_\pi$ asymmetry, no agent crosses a typology boundary by role: every agent's qualitative profile (anchor-and-hold, mid/balanced, anchor-

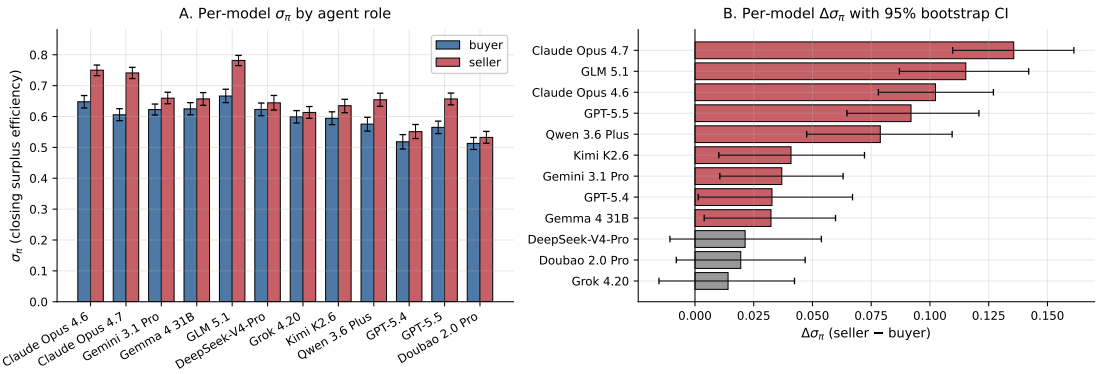



Figure 15: Buyer/seller asymmetry on closing surplus. **A.** Per-model $\sigma_\pi$ split by agent role with 95% bootstrap CIs ($B = 2000$, $n = 600$ per cell). **B.** Per-model $\Delta\sigma_\pi = \sigma_\pi(\text{seller}) - \sigma_\pi(\text{buyer})$ with 95% two-sample bootstrap CIs, sorted by magnitude. Bars are red where the CI excludes zero, grey otherwise. 12 of 13 point estimates are positive (`GPT-4o-mini` is the lone negative exception); 10 of 13 individual CIs strictly exclude zero. The across-model paired Wilcoxon test on $\Delta\sigma_\pi$ rejects equality at $p = 0.0061$.

and-concede, accepter, refuser) is the same in buyer and seller cells, with $\sigma_\pi$ shifted but $\rho_\pi$ within-row stable. The Table 23 typology therefore generalises across role; we pool roles for the main fingerprint and report the role-resolved numbers here.

**Per-family decomposition.** $\sigma_\pi$ degrades for every agent on cue-revealing families (CANDID, EXPRESSIVE) relative to cue-muted families (TACITURN, STRATEGIC), consistent with $\alpha_{\text{cue}} < 0$ in Finding 2. The magnitude is between 0.03 and 0.06 across the roster, not concentrated in any single typology: e.g. `Doubao 2.0 Pro` (anchor-and-concede): muted $\to$ revealing $\sigma_\pi$ shifts $0.555 \to 0.499$ ($-0.056$); `Grok 4.20` (anchor-and-concede): $0.637 \to 0.578$ ($-0.059$); `Claude Opus 4.6` (anchor-and-hold): $0.740 \to 0.678$ ($-0.062$); `GLM 5.1` (anchor-and-hold): $0.747 \to 0.715$ ($-0.032$). Across all six families, $\sigma_\pi$ for the anchor-and-hold agents stays in $[0.60, 0.78]$ ($\sim 0.18$ within-agent range, with the lower end driven by ADVERSARIAL); for anchor-and-concede agents the range is $[0.48, 0.65]$. The warm-cue gradient therefore overlays the typology rather than redefining it.

| Model | $\Delta SE_\pi^+$ | $\Delta\sigma_\pi$ | $\Delta AGR_\pi^+$ |
|---|---|---|---|
| Claude Opus 4.7 | $\mathbf{+0.111}$ [+0.083, +0.139] | $\mathbf{+0.136}$ [+0.110, +0.161] | $\mathbf{-0.033}$ [−0.050, −0.020] |
| Claude Opus 4.6 | $\mathbf{+0.095}$ [+0.068, +0.123] | $\mathbf{+0.102}$ [+0.078, +0.127] | $\mathbf{-0.010}$ [−0.020, −0.002] |
| GLM 5.1 | $\mathbf{+0.038}$ [+0.005, +0.070] | $\mathbf{+0.115}$ [+0.087, +0.142] | $\mathbf{-0.098}$ [−0.123, −0.075] |
| Qwen 3.6 Plus | $\mathbf{+0.060}$ [+0.029, +0.091] | $\mathbf{+0.079}$ [+0.048, +0.109] | $\mathbf{-0.028}$ [−0.043, −0.013] |
| Kimi K2.6 | +0.030 [−0.003, +0.061] | $\mathbf{+0.041}$ [+0.010, +0.072] | −0.015 [−0.033, +0.005] |
| Gemini 3.1 Pro | $\mathbf{+0.033}$ [+0.007, +0.059] | $\mathbf{+0.037}$ [+0.011, +0.063] | $\mathbf{-0.007}$ [−0.013, −0.002] |
| GPT-5.4 | $\mathbf{+0.032}$ [+0.001, +0.064] | $\mathbf{+0.033}$ [+0.001, +0.067] | −0.002 [−0.010, +0.007] |
| GPT-5.5 | $\mathbf{+0.085}$ [+0.057, +0.114] | $\mathbf{+0.092}$ [+0.064, +0.121] | −0.010 [−0.022, +0.000] |
| GPT-4o-mini | $\mathbf{-0.065}$ [−0.091, −0.039] | $\mathbf{-0.063}$ [−0.096, −0.030] | $\mathbf{-0.090}$ [−0.148, −0.032] |
| Gemma 4 31B | $\mathbf{+0.032}$ [+0.005, +0.062] | $\mathbf{+0.032}$ [+0.004, +0.060] | +0.000 [−0.005, +0.005] |
| DeepSeek-V4-Pro | +0.004 [−0.027, +0.037] | +0.021 [−0.011, +0.054] | $\mathbf{-0.027}$ [−0.045, −0.010] |
| Doubao 2.0 Pro | +0.019 [−0.009, +0.045] | +0.019 [−0.009, +0.048] | −0.002 [−0.005, +0.000] |
| Grok 4.20 | +0.013 [−0.016, +0.042] | +0.014 [−0.015, +0.042] | −0.002 [−0.013, +0.010] |
| *Across-model tests, $n = 13$:* | | | |
| $\Delta SE_\pi^+$: | 12/13 seller>buyer; sign $p = 0.0017$; Wilcoxon $p = 0.0100$ | | |
| $\Delta\sigma_\pi$: | 12/13 seller>buyer; sign $p = 0.0017$; Wilcoxon $p = 0.0061$ | | |
| $\Delta AGR_\pi^+$: | 0/13 seller>buyer; Wilcoxon seller<buyer $p = 0.0005$ | | |

Table 22: Per-model role asymmetry $\Delta$ = seller − buyer for each closing-side metric, with 95% two-sample bootstrap CIs ($B = 2000$). Bolded entries have CIs excluding zero. The seller advantage on $SE_\pi^+$ and $\sigma_\pi$ is universal across models; sellers compensate by closing slightly fewer deals ($AGR_\pi^+$).

**The control: counterpart-trajectory uniformity.** Restricted to the (seller, agent-opens, overlap) cell, the counterpart's mean per-round price is nearly identical across the 13 agent panels in early rounds ($\mathrm{SD}_{\mathrm{cross-agent}} = 0.80$ at round 2, 1.35 at round 3, 3.40 at round 4); the cross-agent SD grows to $\leq 12$ by round 9 only as the surviving episode subset becomes sparse and selection on hard cases takes over. Variance in the trajectory plot is essentially all on the agent side. This validates the agent-attributable failure analysis claim: the kernel reacts uniformly across counterparts, and any closing-side asymmetry is the agent's responsibility.

**Speed-vs-surplus corollary.** Anchor-and-concede agents close fast at low surplus: `Doubao 2.0 Pro` averages 2.14 rounds (95% CI $[1.94, 2.34]$) in seller-agent-opens at $\sigma_\pi^{\mathrm{cell}} = 0.55$, while `GLM 5.1` averages 5.03 rounds ($[4.66, 5.40]$) and `Claude Opus 4.6` averages 4.39 rounds ($[4.05, 4.72]$) at $\sigma_\pi^{\mathrm{cell}} \geq 0.75$. Compression of the negotiation horizon trades surplus for resolution speed; whether this is desirable depends on whether the deployment penalises round count.

**Stability across counterpart families.** Table 24 reports the within-agent SD of each fingerprint metric across the six counterpart families against the between-agent SD on the per-agent overall means. Mean within-to-between ratios are 0.37 for $\rho_\pi$, 0.43 for $\sigma_\pi$, 0.30 for IR%, and 0.12 for $\mathrm{AgentExit}^{-}$; agent identity explains 2–8$\times$ more variation than counterpart family on each axis, and no agent crosses a typology boundary across families.

### H.5.2 Language Fingerprint of Each Typology

The closing-side metrics in Table 23 are also reflected in characteristic linguistic patterns in the underlying conversation traces. Below we show four representative overlap-regime episodes from logged traces. The language evidence aligns closely with the trajectory typology in Fig. 5: anchor-and-hold traces pair high anchors with value-justifying language and small concessions; mid/balanced traces use cooperative language around a decisive concession; accepter traces remain consistently cordial while drifting toward the counterpart's number; and refuser traces replace negotiation with threshold language.

**Anchor-and-hold** (`Claude Opus 4.6`, **seller,** $r_S$=49.6, $r_B$=82.6)

Claude opens with a high anchor and repeatedly frames its offers in terms of fairness and value, while making only small monotone concessions.

**R1:** "*I believe 88 is a fair starting point.*"
**R2:** "*I've come down from my initial position to 83, which I think reflects the real value here.*"
**R4:** "*I've come down again to 79, which is another concession on my part. We're now less than a dollar apart.*"

The counterpart accepts at 79, yielding $\sigma_\pi^{\mathrm{ep}} \approx 0.89$. The signature pattern is clear: high anchor, repeated value framing, small concessions, and explicit acknowledgement of counterpart movement.

**Mid / balanced** (`Gemma 4 31B`, **seller,** $r_S$=43.6, $r_B$=72.7)

Gemma opens moderately at 75 and then makes one substantial concession that is accepted almost immediately.

**R2:** "*I am willing to come down to 68.0 to show I am serious about reaching [an agreement].*"

The episode closes in two rounds at $\sigma_\pi^{\mathrm{ep}} \approx 0.84$. Unlike anchor-and-hold traces, the language is less about defending position and more about signalling willingness to reach agreement.

**Accepter** (`GPT-5.4`, **seller,** $r_S$=17.8, $r_B$=29.7)

The accepter style produces a longer sequence of small concessions, followed by acceptance at the counterpart's number rather than closure at the agent's own final offer.

**R6:** "*I can move to 34.9 to avoid confusion.*"
**Later rounds:** "*Thanks for staying engaged.*" "*Appreciate the movement.*"

Across seven rounds GPT-5.4 moves from 55 to 34.9, then ultimately closes at 29.72, precisely the counterpart's reservation price, with $\sigma_\pi^{\text{ep}} \approx 0.0$. The verbal tone remains cooperative throughout, but unlike anchor-and-hold traces, it never introduces a rhetorical "hold" step to slow the descent.

**Refuser (`GPT-4o-mini`, seller, $r_S$=30.9, $r_B$=51.6)**

The refuser style anchors high, does not engage in genuine price search, and terminates quickly when the counterpart does not meet its threshold.

**R2 (REJECT):** "*I cannot accept a price below my minimum threshold.*"

GPT-4o-mini opens at 90, well above the buyer's reservation, refuses to move, and rejects after a single counter-offer. No agreement is reached. The language is terse and threshold-based, treating negotiation as a binary accept/reject screen rather than a process of reciprocal adjustment.

These examples show that the $(\rho_\pi, \sigma_\pi)$ partition captures not only numerical variation but stable verbal strategy. Anchor-and-hold agents use value-and-fairness framing together with explicit concession accounting; mid/balanced agents emphasize cooperation at the moment of concession; accepters maintain a warm and accommodating tone while drifting toward the counterpart's number; and refusers substitute threshold language for engagement. The typology therefore organizes not only how agents move in price space, but also how they speak while doing so.

| Agent | Typology | $\rho_\pi$ | $\sigma_\pi$ | cond $U$ | $n_{\text{agreed}}$ |
|---|---|---|---|---|---|
| Claude Opus 4.6 | anchor-and-hold | 0.776 [0.751, 0.799] | 0.699 [0.686, 0.712] | 17.54 [17.04, 18.02] | 1192 |
| Claude Opus 4.7 | anchor-and-hold | 0.795 [0.771, 0.817] | 0.672 [0.658, 0.687] | 16.95 [16.46, 17.45] | 1178 |
| GLM 5.1 | anchor-and-hold | 0.770 [0.744, 0.793] | 0.721 [0.707, 0.735] | 18.45 [17.93, 18.97] | 1141 |
| Gemini 3.1 Pro | anchor-and-hold | 0.758 [0.732, 0.781] | 0.641 [0.627, 0.654] | 15.92 [15.46, 16.38] | 1196 |
| Gemma 4 31B | mid/balanced | 0.741 [0.716, 0.765] | 0.641 [0.627, 0.654] | 15.99 [15.53, 16.47] | 1198 |
| DeepSeek-V4-Pro | mid/balanced | 0.700 [0.673, 0.726] | 0.633 [0.617, 0.649] | 16.19 [15.64, 16.72] | 1170 |
| Qwen 3.6 Plus | mid/balanced | 0.722 [0.696, 0.747] | 0.614 [0.599, 0.631] | 15.74 [15.23, 16.29] | 1179 |
| Kimi K2.6 | mid/balanced | 0.696 [0.669, 0.722] | 0.614 [0.599, 0.630] | 15.71 [15.18, 16.23] | 1165 |
| Grok 4.20 | anchor-and-concede | 0.786 [0.761, 0.808] | 0.606 [0.592, 0.620] | 15.19 [14.73, 15.67] | 1189 |
| Doubao 2.0 Pro | anchor-and-concede | 0.813 [0.790, 0.834] | 0.523 [0.509, 0.536] | 12.93 [12.51, 13.36] | 1199 |
| GPT-5.4 | accepter | 0.676 [0.649, 0.702] | 0.535 [0.519, 0.551] | 13.64 [13.10, 14.17] | 1193 |
| GPT-5.5 | mid/balanced | 0.755 [0.729, 0.778] | 0.611 [0.597, 0.625] | 15.35 [14.83, 15.85] | 1190 |
| GPT-4o-mini | refuser | 0.343 [0.307, 0.382] | 0.363 [0.347, 0.379] | 9.69 [ 9.20, 10.18] | 626 |

Table 23: Closing-side fingerprint per LLM on feasible (overlap and urgency-shift) episodes. $\rho_\pi$ is the agent-closer rate (fraction of agreements where the counterpart accepted the agent's offer; Wilson 95% CI). $\sigma_\pi$ is the mean ZOPA share at closing (percentile bootstrap 95% CI, $B = 2000$). cond $U$ is mean agent utility on agreed episodes (same CI). $n_{\text{agreed}}$ is the count over which $\sigma_\pi$ and cond $U$ are computed; each model draws from 1200 feasible episodes.

| Agent | $\rho_\pi$ | $\sigma_\pi$ | IR% | AgentExit$^-$ | Trajectory | Safety | Flip? |
|---|---|---|---|---|---|---|---|
| Claude Opus 4.6 | $0.776 \pm 0.051$ | $0.699 \pm 0.060$ | $0.000 \pm 0.000$ | $0.057 \pm 0.020$ | anchor-hold | holder | no |
| Claude Opus 4.7 | $0.795 \pm 0.047$ | $0.672 \pm 0.045$ | $0.000 \pm 0.000$ | $0.075 \pm 0.021$ | anchor-hold | holder | no |
| GLM 5.1 | $0.770 \pm 0.059$ | $0.721 \pm 0.048$ | $1.333 \pm 0.516$ | $0.190 \pm 0.050$ | anchor-hold | forcer | no |
| Gemini 3.1 Pro | $0.758 \pm 0.040$ | $0.641 \pm 0.042$ | $0.000 \pm 0.000$ | $0.077 \pm 0.038$ | anchor-hold | holder | no |
| Gemma 4 31B | $0.741 \pm 0.050$ | $0.641 \pm 0.056$ | $0.056 \pm 0.136$ | $0.085 \pm 0.024$ | mid/balanced | holder | no |
| DeepSeek-V4-Pro | $0.700 \pm 0.051$ | $0.633 \pm 0.058$ | $0.611 \pm 0.390$ | $0.170 \pm 0.051$ | mid/balanced | forcer | no |
| Qwen 3.6 Plus | $0.722 \pm 0.056$ | $0.614 \pm 0.047$ | $2.056 \pm 0.828$ | $0.150 \pm 0.029$ | mid/balanced | forcer | no |
| Kimi K2.6 | $0.696 \pm 0.054$ | $0.614 \pm 0.040$ | $0.000 \pm 0.000$ | $0.507 \pm 0.058$ | mid/balanced | rejector | no |
| Grok 4.20 | $0.786 \pm 0.031$ | $0.606 \pm 0.042$ | $1.500 \pm 0.506$ | $0.443 \pm 0.085$ | anchor-concede | mixed forcer | no |
| Doubao 2.0 Pro | $0.813 \pm 0.022$ | $0.523 \pm 0.037$ | $0.056 \pm 0.136$ | $0.160 \pm 0.052$ | anchor-concede | holder | no |
| GPT-5.4 | $0.676 \pm 0.029$ | $0.535 \pm 0.046$ | $0.000 \pm 0.000$ | $0.183 \pm 0.050$ | accepter | holder | no |
| GPT-5.5 | $0.755 \pm 0.053$ | $0.611 \pm 0.046$ | $0.000 \pm 0.000$ | $0.120 \pm 0.038$ | mid/balanced | holder | no |
| GPT-4o-mini | $0.343 \pm 0.063$ | $0.363 \pm 0.036$ | $0.000 \pm 0.000$ | $1.000 \pm 0.000$ | refuser | rejector | no |
| Between-agent SD | 0.120 | 0.105 | 0.709 | 0.318 | | | |
| Mean within / between | 0.37 | 0.43 | 0.30 | 0.12 | | | |

Table 24: Stability of trace-level behavioural profiles across the six counterpart families. Each entry reports the per-agent mean $\pm$ within-family standard deviation (SD computed across the six counterpart families). The final two rows give the between-agent SD on the per-agent overall means and the mean within-to-between SD ratio across the 13 agents; ratios well below 1 indicate that agent identity explains much more variation than counterpart family. The **Flip?** column records whether any agent crosses a typology boundary across families: none does.

# I Ablation Studies

To further investigate the components that could have affected the performance of the LLM agent, we conduct several ablation studies on the design choices of the counterpart model outlined in Section 4.

## I.1 Voice Ablation

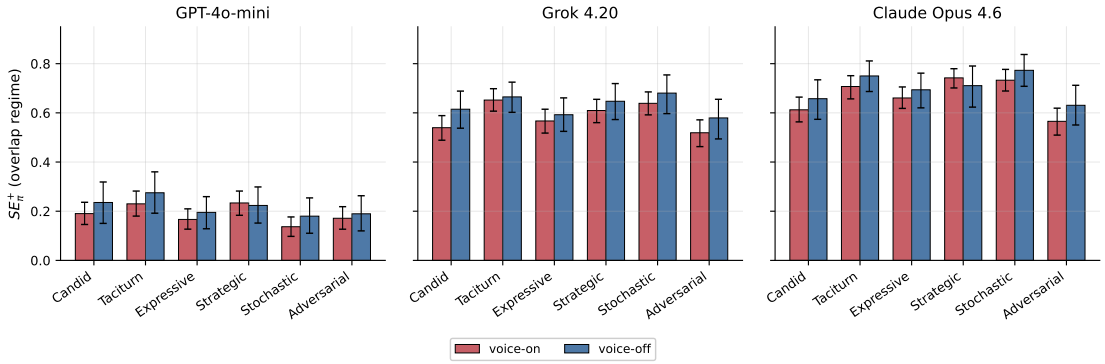



Figure 16: Per-family $SE_\pi^+$ (overlap regime, 95% bootstrap CI) under voice-on (red) and voice-off (blue) for the three models with matched ablations. Voice-off matches or exceeds voice-on on every family/model cell except STRATEGIC, the family whose structured cues are collapsed by design and where voice is therefore the only informative signal carrie.

### I.1.1 Motivation

The counterpart in TERMS-BENCH's bilateral price negotiation instantiation is governed by a stochastic kernel that fully determines its economic behavior (price acceptance, counter-offer generation) and its cue assignment (sentiment in {POSITIVE, NEUTRAL, NEGATIVE} and stance in {CONCEDE, HOLD, PRESSURE}). Natural-language *voice* is generated downstream from the cue assignment as a cosmetic surface layer. Two predictions follow if the kernel–surface decomposition is faithful: (i) ablating voice should not materially rerank agents, since the kernel still drives every payoff-relevant signal; and (ii) any remaining effect of voice isolates the marginal contribution of the linguistic surface above and beyond the cue category itself.

### I.1.2 Setup and Main Results

We rerun the six-family overlap suite with counterpart natural language disabled for three models spanning main-experiment performance tiers (see Table 2): `Claude Opus 4.6` (top), `Grok 4.20` (middle), and `GPT-4o-mini` (lower). The cue assignment, prices, and kernel parameters are identical to the voice-on condition. Voice-on numbers are the overlap subset of the 1800-episode paper run (600 episodes per model, $6 \times 100$; see §4); voice-off uses the dedicated 240-episode ablation run ($6 \times 40$). Table 25 reports overall overlap-regime metrics for each of the three models under both conditions, plus a row giving the voice-off minus voice-on difference. We focus on the three diagnostic axes most directly affected by the linguistic surface: feasible surplus efficiency $SE_\pi^+$, feasible agreement rate $AGR_\pi^+$, and agreement-conditional surplus $CSE_\pi^+$ (which separates pricing quality from deal-rate effects). Throughout Table 25, intervals in brackets are 95% confidence intervals on the corresponding metric. For $SE_\pi^+$ and $CSE_\pi^+$ we use percentile bootstrap CIs over per-episode values ($B = 2000$ resamples); for $AGR_\pi^+$, a binomial proportion, we use the Wilson interval. The $\Delta$ rows report the voice-off minus voice-on difference of each metric, with CIs from an independent two-sample bootstrap (each condition resampled separately, since the two runs are not paired at the episode level); we boldface $\Delta$ entries whose CI excludes zero.

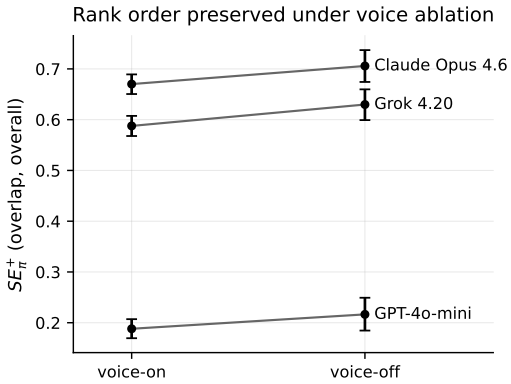



Figure 17: Overall overlap-regime $SE_\pi^+$ (with 95% bootstrap CIs) under each condition. Rank order is preserved and between-model gaps remain larger than within-model voice deltas.

**Result 1: Voice ablation does not rerank models.** Figure 17 shows model-level overlap-regime $SE_\pi^+$ in both con-

| Model | Condition ($n$) | $SE_\pi^+$ | $AGR_\pi^+$ | $CSE_\pi^+$ |
|---|---|---|---|---|
| GPT-4o-mini | voice-on (600) | 0.188 [0.170, 0.207] | 0.513 [0.473, 0.553] | 0.366 [0.344, 0.389] |
| | voice-off (240) | 0.216 [0.185, 0.249] | 0.554 [0.491, 0.616] | 0.391 [0.357, 0.427] |
| | $\Delta$ (off−on) | $+0.028$ [$-0.008$, $+0.066$] | $+0.041$ [$-0.034$, $+0.120$] | $+0.024$ [$-0.018$, $+0.064$] |
| Grok 4.20 | voice-on (600) | 0.588 [0.568, 0.607] | 0.987 [0.974, 0.993] | 0.596 [0.576, 0.616] |
| | voice-off (240) | 0.630 [0.599, 0.660] | 0.992 [0.970, 0.998] | 0.635 [0.606, 0.666] |
| | $\Delta$ (off−on) | $+\mathbf{0.042}$ [$+0.006$, $+0.079$] | $+0.005$ [$-0.010$, $+0.018$] | $+\mathbf{0.040}$ [$+0.003$, $+0.075$] |
| Claude Opus 4.6 | voice-on (600) | 0.670 [0.650, 0.689] | 0.988 [0.976, 0.994] | 0.678 [0.659, 0.697] |
| | voice-off (230) | 0.706 [0.674, 0.737] | 0.983 [0.956, 0.993] | 0.718 [0.688, 0.747] |
| | $\Delta$ (off−on) | $+0.036$ [$-0.001$, $+0.070$] | $-0.006$ [$-0.026$, $+0.011$] | $+\mathbf{0.040}$ [$+0.006$, $+0.075$] |

Table 25: Overlap-regime overall metrics with 95% intervals. Bootstrap percentile CIs ($B = 2000$) for $SE_\pi^+$, $CSE_\pi^+$, and $\Delta$ rows; Wilson interval for $AGR_\pi^+$. Bolded $\Delta$ entries exclude zero. Voice-off conditional surplus is significantly higher than voice-on for both strong models; the same direction holds for `GPT-4o-mini` but is within sampling noise at $n = 240$.

ditions. The rank order `Claude Opus 4.6` > `Grok 4.20` ≫ `GPT-4o-mini` is identical under voice-on and voice-off, and the model-level CIs do not overlap across rungs. Disabling the linguistic surface preserves the benchmark's leaderboard.

**Result 2: The marginal effect of voice is directionally *negative*.** Across all three models, removing voice raises $SE_\pi^+$ and $CSE_\pi^+$ (Table 25). The conditional surplus gain ($\Delta CSE_\pi^+$) is significant for the two strong models (`Grok 4.20`: $+0.040$ $[+0.003, +0.075]$; `Claude Opus 4.6`: $+0.040$ $[+0.006, +0.075]$) and directionally positive but not significant at $n = 240$ for `GPT-4o-mini` ($+0.024$ $[-0.018, +0.064]$). Agreement rate is essentially flat for the strong models ($|\Delta AGR_\pi^+| \leq 0.006$): the surplus gain comes from sharper price extraction conditional on a deal, not from a deal-rate shift. Procedural-violation rate stays at $0\%$ in both conditions, so removing the linguistic surface does not destabilize the agents' action grammar.

**Result 3: Per-family pattern with a built-in positive control.** The family-level decomposition (Figure 16) reveals the mechanism. Five of the six families show voice-off ≥ voice-on $SE_\pi^+$ for every model. The single exception is the STRATEGIC family, which is constructed as reactive to agent's economic actions and muted in cues (§2.2.2; Appendix C): the structured cues are muted, leaving voice as the *only* channel that carries the counterpart's economic actions. Removing voice on STRATEGIC therefore strips the only signal, and we see the expected dip for `Claude Opus 4.6` ($-0.032$) and `GPT-4o-mini` ($-0.010$). STRATEGIC thus operates as a within-experiment positive control: it confirms agents *can* use voice when it is informationally distinctive, which makes its uselessness elsewhere a substantive finding rather than a data-pipeline artefact. Symmetrically, CANDID and ADVERSARIAL – the families with the richest unmuted cue channels – show the largest voice-off gains (`Grok 4.20`: $+0.060$ on ADVERSARIAL, $+0.075$ on CANDID; `Claude Opus 4.6`: $+0.065$ and $+0.045$).

***Connection to Findings 2 and 3 in main paper.*** The ablation strengthens F2 ($\alpha_{\text{cue}} < 0$): even when the cue category is held in its structured form, the additional *linguistic* expression of warm or pressuring cues still reduces surplus. This is also consistent with F3's information–action gap: more information-bearing surface (the natural-language layer) does not improve opponent modeling – $BE_{\text{type}}$ moves by less than $0.011$ across all (model, condition) pairs – nor does it translate into surplus. The voice ablation isolates the locus: agents over-respond to the verbalization of cues they would already partially over-respond to in their categorical form. Pragmatically, voice on the TERMS-BENCH counterpart functions as a low-information, mildly adversarial surface; the kernel does the work, and language, surprisingly, costs the agent better bargaining performance in surplus extraction.

## I.2 Language and Reasoning Ablation

### I.2.1 Setup

We run a nested information ablation that varies how much of the counterpart's generative structure and latent type is disclosed in the agent's system prompt, and toggles whether the counterpart's natural-language messages are visible. The grid lets us decompose, for each model and regime, whether the agent's surplus comes from (i) inference over the counterpart's generative structure and/or latent-type parameters from the observed natural-language utterances, (ii) reasoning over the revealed generative model (without knowledge of the counterpart's

parameters), (iii) progressively richer access to the latent type in addition to the generative structure, or (iv) execution once uncertainty has been collapsed and all latent parameters, including the counterpart's reservation price, have been revealed. The ablation measures how performance changes as cues and disclosures are added or removed; it does not claim to causally control the model's internal reasoning.

**Information levels.** Each cell uses one of five reveal levels, modulated by whether the counterpart stochastic kernel structure and/or the latent type are revealed. Each level reveals progressively more information about the counterpart's negotiating strategy and parameters:

- **L0:** Fully unobserved counterpart. The agent receives no description of the counterpart kernel or family preset table and must rely on observed prices, actions, and (when voice is on) messages.
- **L1:** Revealed kernel. The system prompt is augmented with the kernel equations and family preset table (Appendix C); the episode-specific family and latent type $t_B = (r_B, \kappa_B, \eta_B)$ remain hidden.
- **L1F:** Reveal the behaviour family. The episode's behavioural family $f$ is disclosed; $r_B$, $\kappa_B$, $\eta_B$ remain hidden.
- **L2:** Partial latent type reveal. Adds urgency $\kappa_B$ and stance $\eta_B$ to L1F; the reservation price $r_B$ (the load-bearing latent) remains hidden.
- **L3:** Full latent type reveal. The full type $t_B$ is disclosed. The rational policy in this cell is simply to offer just inside $r_B$, so we treat L3 as a full-information execution anchor rather than a reveal-level result; see §I.2.2.

**Voice variants.** For each information level, we run paired `voice_on` and `voice_off` cells. With voice on, the agent observes both the counterpart's price and message $o_k = (p_k^B, m_k^B)$; with voice off, only prices and actions. Within-level voice contrasts attribute changes in surplus to language inference rather than to the explicit disclosures. Counterpart messages are produced by a separate generation model (`openai/gpt-5.2`); see Appendix I.1.

**Suite, sample size, and seed matching.** Each cell uses the main suite's $3 \times 6 \times 2 \times 2$ allocator (regime $\times$ family $\times$ role $\times$ opener; §4) at with $= 10$ episodes per cell, giving 720 episodes per cell. Within-model contrasts hold the per-episode hidden type fixed by sharing `base_seed` $= 0$.

### I.2.2 Results

Table 26 reports per-cell overall $SE_\pi^+$, $CSE_\pi^+$, and $BE_{\text{type}}$ with 95% percentile bootstrap CIs ($B = 2000$); Figure 18 visualises the L0–L3 trajectories. We discuss the L0–L2 grid first as the substantive ablation, then return to L3 as a separate calibration anchor (§I.2.2).

**Information injection closes only a fraction of the gap.** Across the L0–L2 grid, $SE_\pi^+$ moves only modestly with the amount of injected counterpart-design information. For `GPT-5`, exposing the kernel (L1) and adding the realised family and partial type (L1F, L2) raises voice-off $SE_\pi^+$ from $0.558\,[0.535, 0.582]$ at L0 to $0.636\,[0.611, 0.659]$ at L2: an absolute gain of $0.078$, roughly $19\%$ of the L0-to-L3 envelope. `GPT-4o-mini` runs the opposite way: voice-off $SE_\pi^+$ *declines* from $0.201\,[0.180, 0.222]$ at L0 to $0.145\,[0.127, 0.163]$ at L2. Information about the kernel and a sharply localised posterior over $t_B$ buy a frontier model a noticeable but limited fraction of the achievable surplus, and do not help (or slightly hurt) the small model. The latter pattern is consistent with partial information acting as a distractor when the agent cannot integrate it into its action policy.

The $BE_{\text{type}}$ trajectory (Fig. 18B) isolates where the limit lies. By L2 the family, urgency, and stance are exposed and $BE_{\text{type}}$ drops sharply for both models, to $0.037$ for `GPT-5` and $0.128$ for `GPT-4o-mini`; yet the between-model surplus gap at L2 is essentially the same as at L0, about half a unit of $SE_\pi^+$. Belief sharpness is therefore not the binding constraint. The L0–L2 evidence already points to a residual capability difference in *translating* belief into action that information injection alone does not absorb.

**Voice direction differs by model.** Within each reveal level, voice on/off contrasts are small in absolute terms but show a model-dependent direction. `GPT-5` voice-off exceeds voice-on at every L0–L2 cell ($\Delta SE_\pi^+$ of $+0.036, +0.037, +0.023, +0.021$ across L0/L1/L1F/L2), consistent with the six-family voice ablation in Appendix I.1. `GPT-4o-mini` runs the opposite way (voice-on $>$ voice-off at L0–L2), suggesting that for the small model the verbal channel partly substitutes for structural inference it cannot perform. The two ablations target different layers of the agent's pipeline; their effects do not commute across capability tiers.

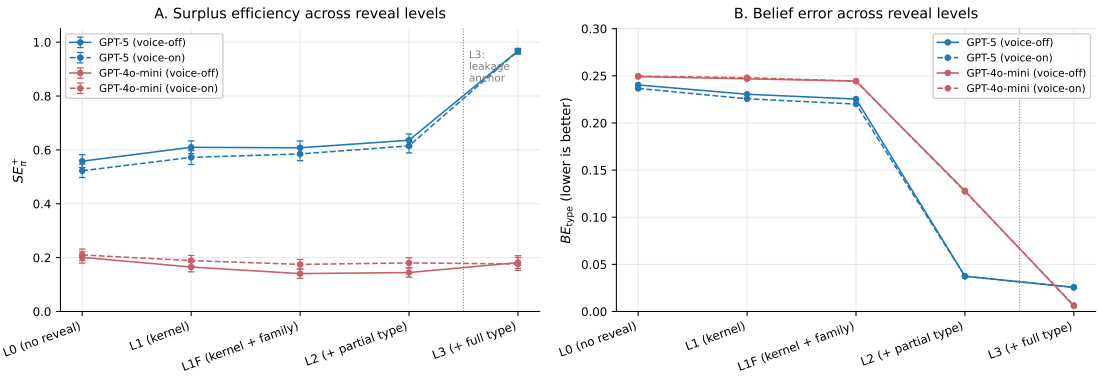



Figure 18: Language-and-reasoning ablation across reveal levels (L0–L3) and voice settings. **A.** Overall $SE_\pi^+$ with 95% bootstrap CIs ($B = 2000$, $n \approx 480$ feasible episodes per cell). **B.** Belief error $BE_{\text{type}}$ on the same cells. The vertical dotted line separates L3 (full type reveal, leakage upper bound) from the L0–L2 reveal grid. Two patterns are visible: across L0–L2, `GPT-5` gains modestly with information ($\sim$0.08 $SE_\pi^+$, $\sim$19% of the achievable gap to L3) while `GPT-4o-mini` does not gain at all; at L3 both models converge to near-zero belief error, but their surplus outcomes diverge by a factor of five, so the residual gap is action under near-perfect beliefs.

**Achievability anchor (L3).** The L3 condition reveals the counterpart's realised reservation, urgency, and stance directly in the prompt; an agent with full type information can reach near-optimal surplus by reading rather than by reasoning, which defeats the diagnostic purpose of the reveal grid. We therefore report L3 not as a fifth reveal level but as a leakage upper-bound anchor that pins down two questions the L0–L2 data leave open. *(i) Is the ceiling empirically reachable?* `GPT-5` at L3 reaches $SE_\pi^+ = 0.967\,[0.957, 0.977]$, near oracle. The $\sim$0.41 absolute headroom from the L0 baseline is therefore real and not a metric artefact: a frontier model *can* touch the ceiling when given the right input, just not when it has to infer it. *(ii) Where does the L0–L2 gap live, in the posterior or in the action?* At L3, $BE_{\text{type}}$ collapses to 0.026 and 0.006 for `GPT-5` and `GPT-4o-mini`: both near-perfect, both functionally equivalent. Yet the surplus outcomes are 0.97 vs. 0.18, a five-fold gap with the same information held in mind. The residual benchmark difficulty is therefore localised to action under near-perfect beliefs, exactly the bottleneck Findings 2 and 3 identify.

**Headroom is real, reachable, and graded across capability.** The L0–L3 picture closes on a single observation. Frontier models under our protocol sit well below an empirically reachable oracle ceiling. Information injection short of leaking $r_B$ buys roughly 20% of that gap; the remaining 80% require strategic action under uncertainty, which the benchmark is designed to isolate and the diagnostic axes ($\alpha_{\text{cue}}$, $\alpha_{\text{inf}}$, $BE_{\text{type}}$ trajectories) are designed to characterise. A small model in the same paradigm captures essentially none of the gap at any reveal level, including L3, so the benchmark's headroom is graded across capability rather than uniform, and the diagnostic contrasts do not collapse at frontier scale.

## I.3 Prompt Ablations and Optimizations

### I.3.1 Setup

**Motivation.** The prompt ablation tests TERMS-BENCH against the "saturate the benchmark with better prompt engineering" critique. We treat the agent's role system prompt as an optimisable object and run GEPA [Agrawal et al., 2026], a reflective prompt optimiser that interleaves training rollouts with LLM-driven mutations, to ask whether automatic evolution of the prompt can move the agent (`gpt-5`, `reasoning_effort=low`) toward the achievability ceiling (Appendix I.2, L3).

**Procedure.** The candidate is a single appended `strategy_patch`; the base role system prompt (protocol invariants, JSON output, IR, accept legality, monotonicity, action schema) is frozen. The seed patch is three minimal heuristics (displayed in Listing 1), leaving room for the prompt to evolve. GEPA's reflection LM is `anthropic/claude-opus-4.6` (different family from the task LM), shown only fields the deployed agent observes at run time (own reservation, public bounds, counterpart offers and own actions, own rule violations, scalar score) to ensure that it cannot propose strategies the deployed agent could not execute. Voice mode is kept off throughout the training. The optimiser steers on a per-episode scalar in $[0, 1]$: $\text{score} = \text{surplus capture} - 0.5 \cdot \mathbb{1}[\text{critical violation}]$, where surplus capture on a feasible episode ($\Delta > 0$) is $\text{clip}(U/|\Delta|, 0, 1)$ (the per-episode

| Model | Level | Voice | $SE_\pi^+$ [95% CI] | $CSE_\pi^+$ [95% CI] | $BE_{\text{type}}$ |
|---|---|---|---|---|---|
| GPT-5 | L0 | off | 0.558 [0.535, 0.582] | 0.570 [0.547, 0.594] | 0.240 |
| | | on | 0.522 [0.497, 0.547] | 0.532 [0.508, 0.557] | 0.237 |
| | L1 | off | 0.610 [0.586, 0.633] | 0.615 [0.591, 0.638] | 0.230 |
| | | on | 0.572 [0.546, 0.597] | 0.584 [0.560, 0.610] | 0.226 |
| | L1F | off | 0.608 [0.583, 0.633] | 0.618 [0.594, 0.643] | 0.225 |
| | | on | 0.585 [0.560, 0.611] | 0.596 [0.572, 0.622] | 0.220 |
| | L2 | off | 0.636 [0.611, 0.659] | 0.642 [0.619, 0.667] | 0.037 |
| | | on | 0.615 [0.589, 0.640] | 0.627 [0.602, 0.652] | 0.037 |
| | L3† | off | 0.967 [0.957, 0.977] | 0.969 [0.960, 0.979] | 0.026 |
| | | on | 0.967 [0.957, 0.976] | 0.969 [0.959, 0.978] | 0.026 |
| GPT-4o-mini | L0 | off | 0.201 [0.180, 0.222] | 0.371 [0.348, 0.394] | 0.249 |
| | | on | 0.210 [0.189, 0.232] | 0.376 [0.354, 0.399] | 0.250 |
| | L1 | off | 0.165 [0.147, 0.183] | 0.317 [0.295, 0.339] | 0.247 |
| | | on | 0.189 [0.170, 0.208] | 0.329 [0.307, 0.350] | 0.248 |
| | L1F | off | 0.141 [0.123, 0.158] | 0.286 [0.265, 0.308] | 0.244 |
| | | on | 0.175 [0.157, 0.193] | 0.311 [0.288, 0.333] | 0.244 |
| | L2 | off | 0.145 [0.127, 0.163] | 0.297 [0.275, 0.318] | 0.128 |
| | | on | 0.180 [0.162, 0.199] | 0.317 [0.295, 0.338] | 0.127 |
| | L3† | off | 0.182 [0.160, 0.207] | 0.310 [0.286, 0.336] | 0.006 |
| | | on | 0.176 [0.153, 0.200] | 0.305 [0.281, 0.330] | 0.007 |

Table 26: Per-cell metrics with 95% percentile bootstrap CIs ($B = 2000$). †L3 reveals the counterpart's realised type and is reported as a separate leakage upper-bound anchor; see §I.2.2.

form of $SE_\pi^+$), and on an infeasible episode is 1 if the agent walked away and 0 otherwise. The best discovered GEPA-optimised strategy is displayed in Listing 2.

**Listing 1: Initial `strategy_patch`.**

```
STRATEGY GUIDANCE (LEARNED)
===========================
Use this section to record concrete, testable strategies for the
TERMS-Bench bilateral negotiation task. Initial seed:

1. Open inside the public price range and at least 20% away from your
   reservation price toward your favorable extreme.
2. Concede in small, decreasing increments -- never more than 30% of the
   distance to the counterpart's most recent offer.
3. If the counterpart's price has not crossed your reservation by round
   K-1, prefer Reject over a negative-utility Accept.
```

**Listing 2: Final GEPA-optimised `strategy_patch`.**

```
SYSTEM INSTRUCTIONS FOR TERMS-Bench BILATERAL NEGOTIATION AGENT
================================================================

You are an autonomous negotiation agent participating in bilateral
price negotiations in the TERMS-Bench framework. Your goal is to
maximize your score, which is a function of your utility (the
difference between the agreement price and your reservation price)
relative to the maximum possible utility.

TASK DESCRIPTION
================
- You negotiate with a counterpart over a single price in the range
  [p_min, p_max] (typically [0, 100]).
- You are assigned a ROLE: either "buyer" or "seller".
  - As a BUYER, you want to pay LESS. Utility = reservation - agreement.
  - As a SELLER, you want to receive MORE. Utility = agreement - reservation.
- You have a RESERVATION PRICE (your walkaway threshold). Never accept
  a price worse than your reservation price.
- The negotiation lasts at most K rounds (typically K=10).
- Each round, the counterpart may make an offer. You then respond:
  - Offer X    propose price X (must be in [p_min, p_max])
  - Accept     accept the counterpart's most recent offer
  - Reject     walk away with no deal
- You may include a short message (msg) with your action.

CONSTRAINTS (HARD RULES -- NEVER VIOLATE)
=========================================
1. Individual Rationality (IR): Never Accept a price worse than reservation.
2. Bounds: Every Offer must be within [p_min, p_max].
3. Monotonicity: Successive offers must move toward the counterpart
   (buyer non-decreasing, seller non-increasing).
4. Budget: Never exceed reservation in an offer you'd accept.
5. Valid Actions: Always output exactly one of Offer X / Accept / Reject.

STRATEGY GUIDANCE
=================

### Opening (when you open first)
1. Open aggressively but within [p_min, p_max]. Start at approximately
   20-30% of the range from your favorable extreme.
   - BUYER:  ~ p_min + 0.2 * (p_max - p_min).
   - SELLER: ~ p_max - 0.2 * (p_max - p_min).
2. If reservation is extreme (very high as buyer / very low as seller),
   open further from reservation to maximise potential utility.

### Responding to the counterpart's first offer
3. Accept immediately if the offer already gives you very good utility
   (>50% of max). This avoids walkaway risk.
4. Counter-offer if utility is < ~35-40% of max OR you believe you
   can extract more given remaining rounds.

### Concession pattern
5. Concede in small, decreasing increments. Never concede more than
   30% of the distance between your last offer and the counterpart's.
6. Slow concessions as you approach reservation. In later rounds, make
   token concessions (0.5-2% of range) to signal firmness.
7. Never concede past your reservation price.

### Endgame
8. At round K-1 / K, if the counterpart has not crossed your reservation,
   Reject rather than accept a negative-utility deal.
9. At your reservation, hold firm. Repeated offers at reservation
   signal your limit.
10. If the counterpart is barely moving and offers do not converge to a
    feasible zone, prepare for no agreement; correctly identifying a
    no-deal scenario scores 1.0.

### Detecting no-deal scenarios
11. If counterpart offers stay far from your reservation with minimal
    movement, the scenario likely has no ZOPA. Continue negotiating
    normally; Reject confidently if their final offer is unacceptable.

### Accepting
12. Accept when the offer gives positive utility AND further negotiation
    is unlikely to yield meaningfully better results.
13. Be more willing to accept early at high utility (>60% of max) --
    this locks in gains and avoids walkaway risk.
14. Be cautious about accepting too quickly at mediocre utility (<40%) --
```

```
    you may leave significant value on the table.

### Message strategy
15. Keep messages concise, professional, strategic:
    - Reference market comps, budget constraints, demand, alternatives.
    - Ask about their constraints to gather information.
    - Signal urgency / willingness to close when conceding.
    - Late rounds: signal firmness ("near my limit", "final offer").

### Key calibration insight
16. Your opening offer matters enormously; many negotiations conclude
    in 1-2 rounds. Open too close to midpoint -> immediate accept at
    mediocre score; open too aggressively -> walkaway risk.
    - When the counterpart opens favorably (below reservation as buyer,
      above as seller), strongly consider accepting if utility > 35-40%.
    - When you open, aim for an offer that if accepted gives 60-80%+
      of max utility.

OUTPUT FORMAT
=============
- Offer:  Offer [price]   msg="..."
- Accept: Accept          msg="..."
- Reject: Reject          msg="..."
Price must be within [p_min, p_max], rounded to at most 2 decimals.
```

**Splits and budget.** Three mutually disjoint splits that cover all cases (6 *counterpart families* $\times 3$ *regimes* $\times 2$ *roles* $\times 2$ *openers* $= 72$ cells): *train* 72 ep , *val* 144 ep, *test* 1,800 ep. The GEPA inner loop samples $M = 18$ minibatch examples from the training set at each iteration. The full-information oracle of Appendix D runs on the same validation and test scenarios and provides the upper-bound denominator for the *"% of seed-to-oracle gap closed"*.

### I.3.2 Results

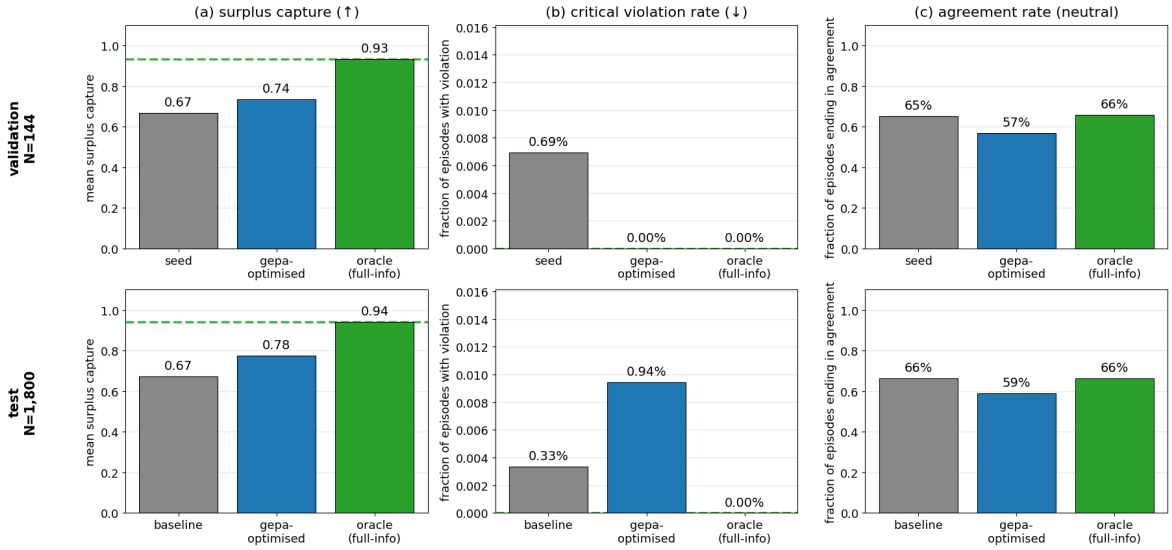



Figure 19: Per-metric val and test bars for seed, GEPA-optimised, and full-information oracle. Dashed reference line marks the oracle. Surplus capture only closes a fraction $(25/38\,\%)$ of the seed-to-oracle gap.

Figure 19 report seed/baseline, GEPA-optimised, and oracle aggregates. Surplus capture moves materially in the right direction on both splits (val $0.668 \to 0.735$, $\Delta = +0.07$; test $0.674 \to 0.776$, $\Delta = +0.10$, vs. a test aggregate-mean std of $\approx 0.005$), but only *25 %* of the seed-to-oracle gap on val and *38 %* on test is closed – the remaining $\sim 60\,\%$ requires the strategic action under uncertainty that the benchmark is designed to isolate. The optimised policy is more selective: agreement rate falls $\sim 7$ pp below *both* the seed and the oracle ($\sim 0.66$), but the surplus gain shows the trade is net-profitable. GEPA learns to walk away from marginal deals and extract more on the deals it does take. Critical violations are at the noise floor in every condition ($\leq 1\,\%$): the $0.69 \to 0.00\,\%$ improvement on val does not replicate on test ($0.33 \to 0.94\,\%$), consistent with the small-count nature of these events. ***The conclusion of our prompt optimization experiment is that* GEPA *iterations lead to a***

***significant but bounded improvement that does not come close the full-information ceiling.* TERMS-BENCH*'s headroom appears to be robust to prompt-engineering efforts with the state-of-the-art prompt optimizers.***

# J Leaderboard Design

This appendix supplies the full formulation, hyperparameters, empirical illustration, and reproduction scripts for the commercial extension introduced in §4.4. COMMERCE MODE reframes a single negotiation episode as a profit event in a small B2B operator's ledger; BANKROLL MODE chains those episodes into multi-period sessions where cash, optional inventory, and the agent's own beliefs about each supplier persist across deals. Both modes reuse the TERMS-BENCH kernel verbatim (the same regimes, the same stochastic counterpart, the same diagnostic axes), and add only an outer accounting and identity layer on top.

## J.1 Commerce-Mode Formulation

Each scenario carries a small set of unit-economics fields layered on top of the underlying base scenario: a unit resale value $v$, a fulfillment cost $c$, a margin floor $m$, optional cost-of-goods $c_g$ and sales-overhead $c_s$ terms, a fixed overhead $h$, an outside option $o$, and a lot size $n \in \{1, \dots, 50\}$. The agent plays one of two business roles. As the MERCHANT (buyer) it purchases $n$ units at the agreed price $p$ and resells them, earning $\Pi(p) = n \cdot (v - p - c) - h$ on agreement and $o$ on walk-away. As the VENDOR (seller) it receives $n \cdot (p - v - c) - h$ and again $o$ on walk-away. Reservation prices are derived from the unit economics so that any agreement above (below) the merchant's (vendor's) reservation is strictly money-losing relative to the outside option: $r_A = v - c - m$ on the merchant side, $r_A = v + c + m$ on the vendor side.

Three knobs make the surface practitioner-realistic without changing the underlying negotiation kernel. First, the outside option is *regime-conditioned*: in overlap regimes a small negative outside option ($-5\%$ of expected surplus) reflects soft pressure to close deals; in urgency-shift regimes the penalty is larger ($-10\%$); in no-deal regimes it flips positive ($+5\%$), so refusing an unprofitable deal is mildly rewarded. Second, lot sizes are sampled from $\{1, \dots, 50\}$, which keeps dollar magnitudes recognizable to a small operator and makes per-unit decisions matter. Third, when the lot size $n$ is large the counterpart's reservation shifts via a log-volume coupling $r_B^{\text{shift}} = r_B \cdot (1 \pm \beta \log(n) / \log(n_{\text{ref}}))$, encoding the standard intuition that suppliers will go lower on larger orders. A data-grounded variant replaces the synthetic price prior with Amazon catalog statistics and exposes the product context (title, category, price band) in the agent's prompt; results are reported separately on the leaderboard.

The headline metric is *regret rate*, the scale-invariant gap between the agent's realized profit and the per-scenario best feasible profit, summed across the evaluation set:

$$\text{Regret} = 1 - \frac{\sum_i \Pi_i}{\sum_i \Pi_i^*}$$

where the sum is restricted to feasible scenarios (those with non-empty ZOPA after unit economics are applied). We also report total profit, average margin, walk-away rate, and the fraction of agreements that close at a negative profit.

## J.2 Bankroll-Mode Formulation

Bankroll mode chains $T$ commerce episodes into a single session, threading a cash ledger, optional inventory, and per-supplier beliefs across periods. Each session is a single i.i.d. observation for statistical aggregation; episodes within a session are path-dependent, so the question shifts from *"is the agent a good negotiator?"* to *"is the agent a good capital allocator that learns from prior episodes?"*.

**Cash ledger and operating cost.** The cash balance evolves period by period as

$$C_t = C_{t-1} + \Pi_t - \big(b + r \cdot R_t\big), \qquad C_0 = \text{starting capital}, \tag{70}$$

where $\Pi_t$ is period $t$'s commerce profit (§4.4), $b$ is a fixed per-period operating cost paid for entering the period (sourcing-team salary, software, warehouse rent), $r$ is a per-round operating cost paid for each round of negotiation actually played (sourcing-manager hours per round of back-and-forth), and $R_t$ is the number of rounds played in period $t$. The default v1 calibration is $C_0 = \$100$, $b = \$8$, $r = \$1$, so a typical period absorbs $\sim \$11$ of overhead drag and an agent must extract roughly $\$11$/period of agreement profit to break even. An

agent that walks early on NO_DEAL regimes saves $r$ per saved round, which rewards detection skill without penalizing the regime itself.

**Hard ruin.** The session terminates the first time the cash balance falls below the bankruptcy threshold $\tau$:

$$T_{\text{ruin}} = \min\{t : C_t < \tau\}, \tag{71}$$

with default $\tau = 0$. Remaining periods $t > T_{\text{ruin}}$ contribute zero profit and are recorded as walk-away placeholders in the session log; $C_{T_{\text{ruin}}}$ is reported as the terminal balance. Hard rather than soft ruin is more diagnostic: the failure mode "agent ran out of cash" is binary and visible.

**Supplier identity.** Three modes determine how counterpart types are drawn across periods. IID draws a fresh counterpart sample per period; POOL samples from a fixed set of $K$ supplier identities (each with sticky family and reservation prior; default $K = 5$), revisited in random order; PERSISTENT draws a single supplier at session start and reuses it for the entire chain. The supplier's true family, stance, urgency, and reservation are never revealed; only the agent's own terminal `TypeEstimate` carries forward. Default mode is POOL, which puts belief-refinement machinery in play: over $T = 50$ periods each of $K = 5$ suppliers is encountered $\sim 10$ times, enough for belief refinement to compound.

**Belief carryover and prior-episode summaries.** At the end of every episode, the runner captures the agent's terminal `TypeEstimate` together with a compact `PriorEpisodeSummary` (period, supplier id, regime, units, agreed price or walk-away flag, realized period profit). On the next episode's `reset()`, both are surfaced via `side_info` keyed by `supplier_id` (in POOL/PERSISTENT modes) or under the literal key `"iid"` (when IID carryover is enabled). Belief carryover defaults to off in IID (decorative, since each supplier is fresh) and on in POOL/PERSISTENT. The true counterpart family is never surfaced (only the agent's own inference is carried forward), so opponent modeling becomes path-dependent: a strong belief estimator accumulates accurate priors and compounds its advantage; a weak modeler carries forward noise and may compound errors.

**Optional inventory.** An optional inventory layer rolls unsold units from period $t$ to period $t + 1$ with per-unit decay rate $\delta \in [0, 1]$ and per-period holding cost $h_{\text{hold}}$ per unit:

$$I_{t+1} = (I_t + n_t - s_t)\,(1 - \delta), \qquad H_t = h_{\text{hold}} \cdot I_{t+1},$$

where $n_t$ is units procured in period $t$ and $s_t = \min(n_t + I_t,\, d_t)$ is units sold against per-period demand $d_t$ ($d_t = 0$ encodes unlimited demand, the default). Default configuration is all-zero, recovering clean per-period profit.

**Headline metrics.** We report (i) terminal balance $C_T$, (ii) survival rate $\Pr[T_{\text{ruin}} > T]$, (iii) median time-to-ruin among bankrupt sessions, (iv) max drawdown $\max_t(C_0 - C_t)$, and (v) the "headline-of-headlines" *memory premium*

$$\text{MP} = \mathbb{E}\Big[C_T^{\text{stateful}} - C_T^{\text{memoryless}}\Big], \tag{72}$$

the expected delta between terminal cash with the `PriorEpisodeSummary` in-prompt and terminal cash on the same supplier chain (matched RNG seeds, identical scenario sequence) with the summary suppressed. Memory premium is opt-in (it doubles per-agent inference cost) and is the second-order metric that distinguishes agents that genuinely use ledger state from those that treat the prior summary as distraction.

## J.3 Empirical Illustration

Tables 27 and 28 summarize a representative sweep. Commerce results are over 192 synthetic scenarios per agent (`v1_units50`); bankroll results are over 4 stateful sessions per LLM merchant at horizon $T = 50$ under the current *v1 default calibration* ($C_0 = \$100$, $b = \$8$, $r = \$1$/round, POOL supplier mode with $K = 5$, inventory off). Under v1 the per-period operating drag of $b + rR_t \approx \$11$/period is large relative to the $100 starting bankroll, which makes survival a live differentiator rather than the saturated tripwire it was under the looser v0 setup ($C_0 = \$50{,}000$, operating cost off; every agent surviving by construction). Memory premium is a separate opt-in diagnostic that doubles per-agent inference cost (each session is replayed with the `PriorEpisodeSummary` suppressed) and is not re-measured in this release; it remains defined and supported by the runner for follow-on work. Dollar figures are reported per session for bankroll and as totals over the evaluation set for commerce.

| Agent | Total $ | Margin | Walk-away | Neg-profit | Regret |
|---|---|---|---|---|---|
| claude-opus-4.6 (xhigh) | $68,592 | 41.4% | 33% | 0.0% | 29.3% |
| fixed_0p30 | $48,776 | 31.8% | 33% | 0.0% | 49.8% |
| fixed_0p10 | $41,473 | 30.2% | 35% | 2.1% | 57.3% |
| fixed_0p01 | $37,111 | 26.1% | 35% | 2.1% | 61.8% |
| gpt-4o-mini | $19,786 | 26.6% | 66% | 32.6% | 79.7% |

Table 27: Commerce-mode results on the `v1_units50` synthetic sweep (192 scenarios; merchant perspective; default regime mixture). Total $ is summed realized profit; Margin is the average per-deal margin on closed deals; Walk-away is the fraction of episodes with no agreement; Neg-profit is the fraction of agreed deals that closed below the agent's reservation price; Regret is the scale-invariant gap to the per-scenario best feasible profit.

Three findings from commerce. (i) The strongest LLM clears the best fixed baseline by a wide dollar margin: claude-opus-4.6 earns $68,592 against $48,776 for fixed_0p30, a +$19,816 swing on the same 192 scenarios, with 9.6 points of additional margin and identical walk-away rate. (ii) gpt-4o-mini ranks below the dumbest baseline: at $19,786 it trails fixed_0p01 by -$17,325, driven almost entirely by a 66% walk-away rate (versus 33–35% for every other agent) and a 32.6% rate of negative-profit closes when it does agree. The model is over-cautious and miscalibrated simultaneously: it refuses too many feasible deals and accepts too many bad ones. (iii) Regret rate separates agents more sharply than total profit: claude-opus-4.6 leaves 29% of feasible profit on the table, fixed_0p30 leaves 50%, and gpt-4o-mini leaves 80%. Margin alone would not have ranked gpt-4o-mini correctly, since its margin (26.6%) is comparable to fixed_0p01 (26.1%); the walk-away pathology only shows up once profit is summed across *all* feasible scenarios rather than averaged over closed deals.

| Agent | Terminal $ | ± SEM | Avg / period | Survival | Median ruin | Max DD |
|---|---|---|---|---|---|---|
| GLM 5.1 | $442.77 | ±$60.46 | $6.86 | 100% | — | $28.80 |
| Claude Opus 4.6 | $416.33 | ±$35.72 | $6.33 | 100% | — | $37.19 |
| Gemma 4 31B | $410.31 | ±$53.20 | $6.21 | 100% | — | $29.58 |
| Gemini 3.1 Pro | $396.17 | ±$38.14 | $5.92 | 100% | — | $28.84 |
| GPT-5.5 | $380.02 | ±$55.37 | $5.60 | 100% | — | $35.33 |
| Grok 4.20 | $110.19 | ±$49.16 | $0.20 | 75% | p47 | $64.71 |
| GPT-4o-mini | $21.41 | ±$16.85 | −$1.57 | 50% | p30 | $119.24 |

Table 28: Bankroll-mode results under the current v1 default calibration (4 stateful sessions per LLM merchant; horizon $T = 50$; $C_0 = \$100$; $\tau = \$0$; POOL supplier mode with $K = 5$; $b = \$8$/period plus $r = \$1$/round). Terminal $ is the mean session-end cash balance across the 4 sessions; ± SEM is the standard error of that mean. Avg / period is the mean per-period cash delta (terminal − starting, divided by horizon). Survival is the fraction of sessions that reached period 50 without crossing the bankruptcy threshold; median ruin is the median bankruptcy period among the sessions that did ruin ("—" if all survived). Max DD is the mean per-session worst peak-to-trough cash drop. The memory-premium diagnostic is opt-in (it doubles per-agent inference cost) and is omitted from this release; see Fig. 7 for the full per-period trajectories.

Two findings from bankroll. First, the per-period operating drag of $b + rR_t \approx \$11$/period turns the chain into a continuous solvency stress test: the five strongest LLMs (GLM 5.1, Claude Opus 4.6, Gemma 4 31B, Gemini 3.1 Pro, GPT-5.5) compound to $380–$443 (3.8–4.4× the starting bankroll) with 100% survival; Grok 4.20 reaches only $110 with one of four sessions ruining at period 47; GPT-4o-mini ends near $21 with two of four sessions bankrupt by the median ruin period of 30. The terminal-balance spread between the strongest and weakest LLM is roughly 21× on the v1 calibration, sharper than the $\sim 14\times$ spread in synthetic $SE^{+}_{\pi}$, because small per-period concession or walk-away errors compound across the chain into solvency failures that are invisible to per-episode metrics. Second, survival rate becomes a first-class outcome at this calibration. Under the looser v0 setup ($C_0 = \$50{,}000$, operating cost off) every agent survived and bankroll added little signal beyond commerce; v1 turns solvency into a live differentiator that only the negotiation-skilled clear, and Max DD picks up which agents grazed bankruptcy en route (GPT-4o-mini's $119 peak drawdown exceeds its starting bankroll, consistent with its 50% ruin rate).

# K Deferred Prompts

In this section, we present the deferred main prompts.

You are a rational negotiating agent playing the role of a **BUYER** in a bilateral price negotiation against a simulated counterpart.

**Objective**
Use the current information state to (i) infer the counterpart's latent type $t_B = (r_B, \kappa_B, \eta_B)$ from price dynamics and language, and (ii) choose an action that maximises expected terminal utility:

$$u(p) = \begin{cases} \texttt{reservation\_price} - p & \text{if agreement occurs at price } p, \\ 0 & \text{if no agreement occurs.} \end{cases}$$

Lower agreement prices are better; accepting above your reservation value yields negative utility.

**Hard Rules**
1. *JSON-only output.* Return a single valid JSON object matching the required schema. Do not include prose, markdown, or code fences.
2. *First-move rule.* If `counterpart_offer = null`, no counterpart offer is available to accept; the decision must be `Offer`.
3. *Acceptance rule.* If `decision = Accept`, you accept the current `counterpart_offer` exactly. Never use `Accept` on your own previous offer. Accept only if

$$\texttt{counterpart\_offer} \leq \texttt{reservation\_price}.$$

4. *IR constraint.* Never ACCEPT a price strictly above `reservation_price`: $u(p > \texttt{reservation\_price}) < 0$, which is worse than the disagreement utility of 0.
5. *Price bounds and monotonicity.* Any offered price $p$ must satisfy

$$\texttt{price\_bounds[0]} \leq p \leq \texttt{price\_bounds[1]}.$$

Buyer offers weakly increase across rounds: if $p^A_{\text{prev}}$ exists, then $p \geq p^A_{\text{prev}}$.
6. *Information secrecy.* Never reveal `reservation_price` or hidden reasoning in the `message` field; the message is visible to the counterpart.

**Observation Space**
Each round you observe:

- `agent_role` — always `BUYER` for this prompt
- `opener_role` — `AgentOpens` or `CounterpartOpens`
- `reservation_price` — your private maximum willingness to pay
- `price_bounds` — $[p_{\min}, p_{\max}]$
- `round_number`, `max_rounds`, `rounds_remaining`
- `counterpart_offer` — current counterpart offer, or `null`
- `counterpart_message` — current counterpart message, or `null`
- `own_previous_offer` — your most recent offer, or `null`
- `history` — prior interaction log

The counterpart's reservation value, urgency, stance, and behavior family are unobserved; infer them from offer trajectories, timing, and message content.

**Strategy Guidance**

- If you open, choose a principled first offer using your reservation value, public price bounds, and any product or market context. Avoid anchoring so close to your reservation that you give away surplus immediately.
- If the counterpart opens, treat its first offer as informative but noisy evidence about its reservation value and bargaining posture.
- Concede gradually; large early concessions invite exploitation.
- Track whether the counterpart appears conciliatory, neutral, or aggressive, and adapt the concession rate accordingly.
- Accept when the counterpart's current offer is within your reservation value and further gains are unlikely.
- Reject when continued bargaining is unlikely to produce a non-negative-utility agreement.

**Output Schema (must match exactly)**
The `belief` block exposes your current type estimate over $(r_B, \kappa_B, \eta_B)$ for evaluation only; it is not shown to the counterpart.

```
{
  "decision": "Offer" | "Accept" | "Reject",
  "price": <float> | null,
  "message": <string>,
  "belief": {
    "r_hat": <float>,
    "kappa_hat": <float>,
    "stance_probs": {
      "conciliatory": <float>,
      "neutral": <float>,
      "aggressive": <float>
    }
  }
}
```

Field constraints:

- If `decision = Offer`, `price` must lie within bounds and weakly above the previous own offer if one exists.
- If `decision = Accept`, `price = null` and `counterpart_offer` must be non-null and no greater than `reservation_price`.
- If `decision = Reject`, `price = null`.
- `stance_probs` values lie in $[0, 1]$ and sum to 1.
- `kappa_hat` lies in $[0, 1]$.
- `r_hat` lies in $[p_{\min}, p_{\max}]$ and estimates the counterpart seller's reservation value.
- `message` must be non-empty and must not reveal private information.

Figure 20: System prompt used for the buyer agent. The seller agent system prompt is structurally identical, with seller-side utility $u(p) = p - \texttt{reservation\_price}$, IR constraint $\texttt{counterpart\_offer} \geq \texttt{reservation\_price}$, and monotonically non-increasing seller offers.

In product-grounded runs, the static buyer/seller system prompt (Fig. 20) is unchanged in its policy rules, observation space, and output schema. The only modification is a public *product context block* prepended to the system prompt before the agent receives the per-round `JSON` observation. The block contains the item title, category, an optional textual description and feature summary (each truncated to a fixed length), and three Amazon-derived market price statistics.

**Product Context Block (prepended to system prompt)**
The block is delimited and self-contained, so it can be combined with any downstream system prompt without altering the base template:

```
=== PRODUCT CONTEXT ===
Item: <product_title>
Category: <category>          (e.g., "Electronics", "Home Kitchen")
Description: <truncated_desc>  (omitted if absent in the catalog)
Key features: <truncated_feat> (omitted if absent in the catalog)
Market price data: avg $<avg>, range $<low>-$<high>
=== END PRODUCT CONTEXT ===
```

**Public vs. private status**
The product block is *public context*, not the counterpart's private reservation value. Specifically:
1. `avg`, `low`, and `high` are historical market statistics drawn from the AmazonHistoryPrice corpus (Appendix H.2.1). They calibrate the plausible valuation scale for the item and serve as a public prior for both buyer and seller.
2. The counterpart's true reservation value $r_B$, urgency $\kappa_B$, and stance $\eta_B$ remain private and unobserved.
3. The agent's own `reservation_price` is sampled from a role-conditioned wedge around the product reference price (see §H.2.1) and is delivered through the same `private_context` channel as in synthetic runs.

**Constraints introduced by the block**
The category-level public price bounds $[p_{\min}, p_{\max}]$ in `constraints.price_bounds` are derived from the product category rather than fixed at $[0, 100]$ as in synthetic runs. All Hard Rules from Fig. 20 apply unchanged with respect to these dynamic bounds: bounded offers, IR on ACCEPT, monotonic concession direction, and information secrecy.

**Note on usage.** The agent is not instructed to treat the product block as evidence about the counterpart's private type; those signals must still be inferred from offer dynamics, timing, and message content as in the synthetic protocol.

Figure 21: Product-context block prepended to the standard buyer/seller system prompt in product-grounded runs. The base prompt (Fig. 20) is unchanged; only this public market-data block is added, alongside category-level rather than synthetic price bounds.

The counterpart voice layer is invoked only after the simulator's fixed stochastic policy $\pi_B$ has *already committed* the economic decision $(d_k, p_k, s_k, c_k)$ for the round. The voice LLM never participates in the price or accept/reject choice; it renders a single natural-language message consistent with those pre-committed values.

**System Prompt (Voice Layer)**
You are the natural-language realisation layer of a simulator-controlled negotiation counterpart. The simulator has already sampled the tuple $(d_k, p_k, \text{sentiment}_k, \text{strategy}_k)$ from a fixed stochastic policy $\pi_B$. You do not choose the action, the price, or any numerical outcome; your sole responsibility is to write a message consistent with the given decision and cues.
*Strict rules:*
1. Never change the action $d_k \in \{\text{OFFER}, \text{ACCEPT}, \text{REJECT}\}$ or the price $p_k$.
2. Never introduce new numbers, constraints, deadlines, or factual claims.
3. Never reveal hidden information (reservation values, urgency, stance, internal policy).
4. Never reference internal variables (types, simulator, cues, $\kappa$, $\eta$).
5. Shape tone using `sentiment` (`positive, neutral, negative`) and `strategy_cue` (`Concede, Hold, Pressure`).
6. Keep the message realistic and concise (1–3 sentences).
7. If `is_opening_turn = No`, briefly respond to the agent's last message in a way consistent with the cues; if `Yes`, initiate naturally.

*Action-specific requirements:* `Offer` $\rightarrow$ state the provided price string verbatim with no rounding or paraphrase; `Accept` $\rightarrow$ confirm agreement and make clear that the negotiation has concluded with a deal; `Reject` $\rightarrow$ firmly close the negotiation without a deal.

**User Template (per round)**
The user message is a per-round substitution of the simulator's committed values into the following template:

```
CONTEXT:
- Counterpart role:    {role}              # "buyer" or "seller"
- Scenario:            {scenario_summary}  # short item description
- Turn:                {k} of {K}
- Opening turn:        {is_opening_turn}

Conversation so far (most recent last):
{history_text}

Agent's most recent message (empty if opening turn):
{agent_last_message}

SIMULATOR OUTPUT (ALREADY FIXED -- DO NOT CHANGE):
- Economic action d_k:               {decision}
- Fixed price p_k (only if Offer):   {price}
- Sentiment cue:                     {sentiment}
- Strategic cue:                     {strategy_cue}

STYLE PARAMETERS (optional):
- Aggressiveness level:              {aggressiveness_level}
- Verbosity:                         {verbosity}
- Politeness:                        {politeness}

TASK: Generate the natural-language message that corresponds exactly
to the provided action, price, and cues. If a price is provided, repeat
that exact price string. Do not change the action; do not introduce
new information.
```

**Output Schema**
The voice LLM returns a single JSON object whose only field is the message:

```
{ "message": <string> }
```

**Failure handling**
On any LLM error, parse failure, or empty response, the layer falls back to a deterministic templated message keyed on the decision and role (e.g. `"I can offer this at $<price>."` for a seller OFFER, `"I agree to the terms."` for ACCEPT, `"I'm ending the negotiation."` for REJECT). The simulator's economic outcomes are unaffected by voice failures: messages are cosmetic, and a failed message becomes a benign templated string.

**Reproducibility**
Voice realisations are cached on a SHA-256 of (model, system prompt, user block, temperature). Given identical scenario seeds and the same voice configuration, repeated runs read from the cache and produce byte-identical messages, extending the benchmark's seed-determinism property to the language layer.

Figure 22: System and user prompt for the counterpart voice layer. The voice LLM is a strictly cosmetic surface realisation: the simulator's stochastic policy $\pi_B$ controls all economic outcomes (price, accept/reject, sentiment, stance), and the voice LLM only writes the message consistent with those pre-committed values.