

BalanceAgent: A Preference-Concealing Negotiation Agent with Ordinal Bayesian Opponent Modeling

Tianzi Ma¹, Ruoke Wang¹, Qiumeng Yang¹, Yulin Wu¹ and Xuan Wang¹

¹Harbin Institute of Technology, Shenzhen

mtz982437365@gmail.com, 24S151159@stu.hit.edu.cn, 25S151144@stu.hit.edu.cn,
yulinwu@cs.hitsz.edu.cn, wangxuan@cs.hitsz.edu.cn

Abstract

We present BALANCEAGENT, a negotiating agent for the ANAC 2026 ANL league that balances the dual objectives of maximizing agreement utility and concealing true preferences. The agent constructs a fake preference profile uncorrelated with its true utility function and restricts all bids to the intersection of true and fake acceptable tiers, ensuring the opponent learns a distorted preference picture. A time-phased bidding strategy randomizes offers in early phases to suppress frequency-based leakage, then maximizes predicted opponent utility near the deadline. Acceptance follows time-gated thresholds with convex decay, softened by opponent-concession estimates from a windowed ordinal Bayesian opponent model that integrates frequency, window-level concession, and ordinal rejection-counter-offer signals.

1 Introduction

The ANL (Automated Negotiation League) 2026 competition, part of ANAC (Automated Negotiating Agents Competition), challenges participants to design a bilateral negotiation agent that operates under a dual objective: (i) maximizing the utility of agreements reached, and (ii) concealing its true preferences from the opponent. In each negotiation, the agent is scored not only on the *advantage* (agreement utility minus reservation value) it obtains, but also on how poorly the opponent’s learned model of its preferences correlates with its actual utility function. An agent that consistently wins high-utility deals while misleading the opponent about its true interests will therefore rank highest.

This dual objective creates a tension that is absent from classical negotiation settings. A straightforward strategy of bidding outcomes near one’s maximum utility and gradually conceding makes the agent’s preferences transparent: the opponent can easily infer which issues matter most and which options are preferred by simply observing the sequence of offers [Chang and Fujita, 2024]. Conversely, a purely randomized strategy hides preferences well but sacrifices utility. Our agent, BALANCEAGENT, is designed to navigate this trade-off through a combination of deceptive bidding, time-phased acceptance, and opponent modelling.

The architecture consists of three main components:

- **Fake preference profile** (Section 2.1). At initialization, we generate a synthetic utility function that is intentionally uncorrelated with our true preferences. All bids are then restricted to outcomes that lie at the intersection of our true high-utility tiers and the fake profile’s acceptable tiers. This ensures that the opponent, observing the stream of our offers, learns a distorted picture of what we value.
- **Time-phased bidding strategy** (Section 2.2). Our bidding behavior evolves with negotiation time: in the early and middle phases, we randomize within our safe bid pool to resist opponent modelling; in the late phase, we switch to selecting outcomes that maximize the opponent’s predicted utility, increasing the chance of a timely agreement.
- **Multi-layered acceptance strategy** (Section 2.3). The acceptance bar is controlled by a sequence of time gates. The bar is a narrow window at the very start, through a convex concession curve in the middle [Faratin *et al.*, 1998], to a relative-surplus rule in the final moments that guarantees we do better than the opponent under our best estimate of their preferences.
- **Opponent modelling** (Section 3) underpins both the late-phase bidding and the acceptance adjustments, providing the estimates $\hat{u}_{\text{opp}}(\cdot)$ that guide our decisions.

2 Strategy Design

The core idea behind BalanceAgent is deceptive bidding with time-phased concession. On the bidding side, we construct a fake preference profile that is uncorrelated with our true utility function, and restrict our offers to outcomes that are simultaneously good under our true preferences and acceptable-looking under the fake profile. This prevents the opponent from learning our real interests. On the acceptance side, we employ a multi-layered time-gated threshold that gradually lowers the bar as the negotiation deadline approaches, ensuring that we secure a favorable deal when an opportunity arises without conceding prematurely.

2.1 Deceptive Profile and Bid Pool Construction

At the start of each negotiation, BalanceAgent builds a synthetic preference profile uncorrelated with its true utility func-

tion, then uses it to carve out safe bid pools — outcomes that are genuinely good for the agent yet appear innocuous to an observing opponent. This section describes the three-stage pipeline: tier classification, fake profile generation, and pool intersection.

True Tier Classification

On initialization, we first enumerate all outcomes in the negotiation domain and retain those whose utility exceeds our reservation value. These constitute the set of *rational outcomes*, sorted in descending order of true utility:

$$\mathcal{R} = \{\omega \in \Omega \mid u(\omega) > u_{\text{res}}\}.$$

We then partition $\mathcal{R}$ into three quality tiers. Each tier is defined by the *intersection* of two criteria: an absolute utility gap from the maximum u_{max} , and a top-percentile threshold. Concretely, let $\delta_1 < \delta_2 < \delta_3$ be three small utility gaps and let $\rho_1 < \rho_2 < \rho_3$ be three percentile fractions. Then:

$$A_i = \{\omega \in \mathcal{R} \mid u(\omega) \geq u_{\text{max}} - \delta_i\},$$

$$P_i = \{\text{the first } \lceil \rho_i \cdot |\mathcal{R}| \rceil \text{ outcomes in } \mathcal{R}\},$$

$$T_1 = A_1 \cap P_1, T_2 = (A_2 \cap P_2) \setminus T_1, T_3 = (A_3 \cap P_3) \setminus (T_1 \cup T_2).$$

Using intersection rather than union prevents a single weak criterion from inflating a tier. The union $T_1 \cup T_2 \cup T_3$ forms the *ending tier* T_E , i.e., outcomes we are willing to settle for near the deadline.

Fake Preference Profile Construction

To hide our true preferences from the opponent, we build a fake multi-issue utility function via the `FakeProfileBuilder`. The fake profile consists of fake cross-issue weights $\tilde{\mathbf{w}}$ and fake within-issue option evaluations $\tilde{v}_i(\cdot)$ for each issue i . The design must satisfy three requirements. The first is **uncorrelatedness**: the fake profile must exhibit low rank correlation with the true profile (Spearman $\rho \approx 0$ for within-issue option values; $\rho \approx -1$ for cross-issue weights), so that observing our bids reveals little about what we actually value. The second is **tier coverage**: every true T_1 outcome must fall into the fake T_1 or T_2 , and every true T_2 outcome must fall into $\tilde{T}_1$, $\tilde{T}_2$, or $\tilde{T}_3$. Without this constraint, we might be unable to bid our genuinely preferred outcomes without them appearing suspicious under the fake profile. The third is **polarity inversion**: when the true distribution is polarized (high variance, large max/min ratio), we bias generation toward uniform fake candidates, and vice versa, further decorrelating the two profiles.

Candidate weights and option values are produced by mixing three generation strategies. **Noisy permutations** shuffle the true values across issues (or options) and add Gaussian noise scaled to the true distribution’s standard deviation, preserving the marginal shape while destroying inter-issue structure. **Dirichlet and uniform samples** draw weights from a Dirichlet distribution whose concentration parameters are chosen inversely to the true polarization, and draw option values uniformly or from Gaussian-centered distributions, with polarity-aware parameter ranges. **Pinched profiles** assign one issue (or option) a dominant value while keeping the rest small and uniform; this method is used sparingly when the

true profile is already polarized, and heavily when the true profile is uniform, creating a starkly contrasting fake.

From the generated candidates we select the top- k most uncorrelated ones (by Spearman ρ distance to a target of 0 for option values and -1 for weights). Then we evaluate the *coverage rate* of each candidate profile — the fraction of true tier-1 and tier-2 outcomes that satisfy the tier-coverage constraints. A composite score combining uncorrelatedness and coverage determines the final fake profile $(\tilde{\mathbf{w}}, \{\tilde{v}_i\})$.

Bid Pool Construction

Given the true tiers T_1, T_2, T_3 and the fake profile, we classify outcomes into fake tiers $\tilde{T}_1, \tilde{T}_2, \tilde{T}_3$ using the same intersection rule applied to fake utilities. The *bid pools* are then the intersections:

$$\mathcal{B}_1 = T_1 \cap (\tilde{T}_1 \cup \tilde{T}_2),$$

$$\mathcal{B}_2 = T_2 \cap (\tilde{T}_1 \cup \tilde{T}_2 \cup \tilde{T}_3),$$

$$\mathcal{B}_3 = T_3 \cap (\tilde{T}_1 \cup \tilde{T}_2 \cup \tilde{T}_3).$$

Intuitively, $\mathcal{B}_i$ contains outcomes that are genuinely in tier i for us and simultaneously appear to be in an acceptable tier under the fake profile. Because the fake profile is uncorrelated with the true one, the opponent observing our bids will receive a misleading signal about our preferences.

2.2 Bidding Strategy

Our bidding strategy is time-phased and switches from exploration (randomized) to exploitation (opponent-targeted) as the deadline approaches. Let τ_{bid1} and τ_{bid2} be two time thresholds with $0 < \tau_{\text{bid1}} < \tau_{\text{bid2}} < 1$.

Time	Bid pool	Selection
$t < \tau_{\text{bid1}}$	$\mathcal{B}_1$	Uniformly random
$\tau_{\text{bid1}} \leq t < \tau_{\text{bid2}}$	$\mathcal{B}_1 \cup \mathcal{B}_2$	Uniformly random
$t \geq \tau_{\text{bid2}}$	$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$	Maximize predicted opponent utility

Early and middle phases ($t < \tau_{\text{bid2}}$). We randomize within the available bid pools. Randomization serves three purposes: (i) it prevents the opponent from reverse-engineering our true utility function via pattern analysis of repeated offers; (ii) by drawing uniformly, we equalize the bidding frequencies across outcomes within the active pool, so the opponent cannot infer preference rankings from how often certain outcomes appear; and (iii) it explores the outcome space in a way that the fake profile makes look reasonable to the opponent.

Late phase ($t \geq \tau_{\text{bid2}}$). We abandon randomization and instead select the outcome from the full bid pool that maximizes the opponent’s predicted utility according to our learned opponent model:

$$\omega^* = \arg \max_{\omega \in \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3} \hat{u}_{\text{opp}}(\omega).$$

The rationale is that near the deadline the opponent is more likely to accept an offer they find attractive, and we should therefore present the outcome that best satisfies their inferred preferences, guaranteeing that it belongs to one of our true tiers.

Fake persona consistency. After each bid, we record the offer into our opponent tracker so that the tracker’s predicted opponent utility function is updated. This ensures that our late-phase opponent-targeted selection uses the most up-to-date opponent model. In addition, the fake profile weights are stored and reused, maintaining a consistent “fake persona” across all negotiation rounds.

2.3 Acceptance Strategy

Our acceptance rule is a time-varying threshold: accept an opponent offer ω at time t if $u(\omega) \geq \theta(t, \omega)$, where θ depends both on the remaining time and on what our opponent model tells us about the opponent’s concession level. A hard floor at the reservation value applies at all times.

The threshold function is implemented as a sequence of time gates, following the modular decomposition principle of the BOA architecture [Baarslag *et al.*, 2014]. At the very start ($t \leq \tau_{\text{blitz}}$), we accept only T_1 outcomes, exploiting rare early generosity. During the early–mid phase ($\tau_{\text{blitz}} < t < \tau_{\text{concede}}$), we expand to $T_1 \cup T_2$ and hold firm. When the concession window opens ($\tau_{\text{concede}} \leq t < \tau_{\text{end2}}$), the threshold follows a convex decay curve $\theta_{\text{base}}(t)$ from u_{max} toward $u_{\text{max}} \cdot r_{\text{min}}$ with exponent $p > 1$, producing a Boulware-style posture [Faratin *et al.*, 1998] of holding firm then conceding rapidly. This base threshold is further softened by an opponent-concession term: if the opponent’s offer is poor for them (low $\hat{u}_{\text{opp}}$ relative to their maximum), we lower the bar proportionally by a factor η . Near the deadline ($t \geq \tau_{\text{end2}}$), we accept any outcome in the ending tier $T_E = T_1 \cup T_2 \cup T_3$. In the final seconds ($t \geq \tau_{\text{end3}}$), we switch to a relative-surplus rule: accept only if $u(\omega) > \hat{u}_{\text{opp}}(\omega)$ and $u(\omega) \geq u_{\text{res}} + \gamma \cdot (u_{\text{max}} - u_{\text{res}})$, where the surplus ratio γ steps down once more at τ_{end4} to guarantee a deal before the deadline.

The overall effect is a threshold that stays flat and high early on, decays convexly through the middle, and finally switches to a relative-advantage rule that ensures we never accept a deal worse than what we believe the opponent gets.

3 Ordinal Bayesian Opponent Modeling

Opponent modeling is crucial in automated negotiation, since an agent must reason about the opponent’s latent preferences from limited behavioral observations [Baarslag *et al.*, 2016]. In our setting, this problem becomes more challenging because agents are also encouraged to conceal their true preferences. Therefore, the opponent’s offer sequence cannot be treated as a direct revelation of its utility function. A rational opponent may intentionally fluctuate its bids to hide which issues and values it truly prefers. To address this problem, we propose a windowed ordinal Bayesian opponent model, which estimates the opponent’s utility function from robust behavioral tendencies rather than isolated offer-level changes.

We estimate the opponent’s utility function with an additive utility model. Given an offer $o = (x_1, x_2, \dots, x_m)$, where x_j denotes the selected value on issue j , the predicted opponent utility is defined as

$$\hat{U}_{\text{opp}}(o) = \sum_{j=1}^m w_j u_j(x_j),$$

where w_j is the estimated importance weight of issue j , and $u_j(x_j) \in [0, 1]$ is the estimated utility of value x_j for the opponent.

The Bayesian update is built on three strategic behavior assumptions. First, although the opponent may hide its preferences, it is still unlikely to repeatedly propose values that are extremely undesirable to itself. Thus, frequently appearing values are treated as possible high-utility values, and issues that remain stable across multiple opponent offers are considered more important. This gives a frequency-based behavioral likelihood $\mathcal{L}_{\text{freq}}$.

Second, local offer order may be unreliable under preference-hiding incentives. A deceptive opponent may intentionally alternate between different offers, making it unreasonable to assume that every earlier offer is preferred to every later adjacent offer. Instead, we assume that the opponent’s aggregate bidding region still tends to move from more self-favorable offers to more concessive offers over time. Therefore, we compare windows of offers rather than individual adjacent bids. Let $W_1, W_2, \dots, W_K$ denote consecutive windows of opponent offers. For each window, we compute the average estimated opponent utility:

$$\bar{U}_{\text{opp}}(W_k) = \frac{1}{|W_k|} \sum_{o \in W_k} \hat{U}_{\text{opp}}(o).$$

The window-based assumption encourages earlier windows to have higher estimated opponent utility than later windows:

$$\bar{U}_{\text{opp}}(W_a) > \bar{U}_{\text{opp}}(W_b), \quad a < b.$$

This produces a window-level concession likelihood $\mathcal{L}_{\text{win}}$, which is more robust to short-term bid oscillations and preference-hiding behavior.

Third, rejection and counter-offer behavior provides ordinal preference information. When our offer a_t is rejected and the opponent instead proposes a counter-offer b_t , it is natural to assume that b_t is more favorable to the opponent than a_t . Otherwise, the opponent would have no clear reason to reject a_t and replace it with b_t . This gives the pairwise ordinal relation

$$a_t \prec_{\text{opp}} b_t, \quad \text{i.e.,} \quad \hat{U}_{\text{opp}}(a_t) < \hat{U}_{\text{opp}}(b_t).$$

This ordinal signal is useful because it does not require access to the opponent’s absolute utility values. It only relies on the relative preference implicitly revealed by the opponent’s own decision.

Based on these assumptions, the posterior belief over opponent utility parameters is updated as

$$P_t(\theta \mid H_t) \propto P_0(\theta) \cdot \mathcal{L}_{\text{freq}}(\theta) \cdot \mathcal{L}_{\text{win}}(\theta) \cdot \mathcal{L}_{\text{ord}}(\theta),$$

where $P_0(\theta)$ is the prior distribution over opponent utility parameters, H_t is the negotiation history, and $\mathcal{L}_{\text{ord}}$ denotes the likelihood induced by the ordinal rejection–counter-offer relations.

The posterior guides the update of value utilities and issue weights. Values that frequently appear in the opponent’s offers, remain stable within high-utility windows, or occur in opponent counter-offers are assigned higher estimated utilities. Conversely, values mainly appearing in offers rejected

by the opponent receive weaker preference estimates. Issue weights are adjusted according to the stability and explanatory power of each issue, and then normalized to preserve a valid additive utility model.

Overall, this model follows a strategic rather than purely mechanical interpretation of opponent behavior. Frequency information captures what the opponent repeatedly reveals, window-level comparison reduces the influence of preference-hiding fluctuations, and ordinal rejection-counter-offer relations capture what the opponent implicitly prefers. Together, these signals allow the agent to maintain a robust estimate of the opponent’s utility function and support adaptive bidding and acceptance decisions.

4 Conclusion

BALANCEAGENT navigates the tension between utility maximization and preference concealment through three mechanisms: a fake preference profile that decorrelates observed bids from true interests, a time-phased bidding strategy that suppresses frequency-based leakage, and a time-gated acceptance rule calibrated by opponent modelling. The windowed ordinal Bayesian opponent model underpins the agent’s adaptivity by integrating frequency, window-level concession, and ordinal rejection-counter-offer signals, maintaining robustness against preference-hiding adversaries. The modular architecture admits future extensions such as higher-order theory-of-mind reasoning and domain-adaptive fake-profile generation.

References

- [Baarslag *et al.*, 2014] Tim Baarslag, Koen Hindriks, Mark Hendriks, Alexander Dirkzwager, and Catholijn Jonker. Decoupling negotiating agents to explore the space of negotiation strategies. In *Novel insights in agent-based complex automated negotiation*, pages 61–83. Springer, 2014.
- [Baarslag *et al.*, 2016] Tim Baarslag, Mark JC Hendriks, 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*, 30:849–898, 2016.
- [Chang and Fujita, 2024] Shengbo Chang and Katsuhide Fujita. Comb: Scalable concession-driven opponent models using bayesian learning for preference learning in bilateral multi-issue automated negotiation. *Group Decision and Negotiation*, 33(5):1143–1190, 2024.
- [Faratin *et al.*, 1998] Peyman Faratin, Carles Sierra, and Nick R Jennings. Negotiation decision functions for autonomous agents. *Robotics and Autonomous Systems*, 24(3-4):159–182, 1998.