

Whale: A Dynamic Adaptive Deceptive Agent for Preference Concealment in Bilateral Multi-Issue Negotiation

Ori Braverman
oribraverman@gmail.com

Omri Perry
omriperry42@gmail.com

June 2026

Abstract

The Automated Negotiation League (ANL) 2026 introduces a novel tournament format focusing on preference concealment. In addition to securing high individual utility (Advantage), agents must successfully mislead their opponent’s learning algorithms to maximize their preference concealment score (Concealing). We present **Whale**, a negotiating agent designed to excel under this dual-scoring environment. Whale extends the classic Bidding, Opponent Modeling, and Acceptance (BOA) architecture with a state-of-the-art Dynamic Adaptive Deception (DAD) module. By profiling the opponent online using a multi-tactic classifier, Whale dynamically swaps its bidding and acceptance strategies. To poison the opponent’s frequency-based models, the DAD module estimates issue-specific impact and stochastically proposes decoy offers on high-impact issues. These decoy offers are bounded by a phase-based utility budget that decays near the deadline to guarantee agreement convergence. Furthermore, Whale employs an anti-deception defense mechanism (using a Null Model) when it detects deceptive behavior in the opponent. Benchmark evaluations demonstrate that Whale successfully reduces opponent modeling accuracy to near-zero levels while maintaining a high individual advantage.

1 Introduction

Automated bilateral negotiation is a fundamental research area in multi-agent systems, enabling autonomous agents to reach mutually beneficial agreements over multi-issue spaces. Traditionally, competitions like the Automated Negotiating Agents Competition (ANAC) focused on maximizing the agent’s individual utility relative to its reservation value, formalizing the goal as the maximization of Individual Advantage:

$$\text{Advantage}_i = U_i(x) - RV_i \quad (1)$$

where x is the agreed outcome, U_i is agent i ’s utility function, and RV_i is its private reservation value.

In the ANL 2026 league, the competition introduces a challenging paradigm shift: **preference concealment**. Agents are evaluated on a combined score:

$$\text{Score}_i = \text{Advantage}_i + \text{Concealing}_i \quad (2)$$

The concealing score is a relative measure of how poorly the opponent modeled the agent’s utility function compared to how well the agent modeled the opponent’s utility function:

$$\text{Concealing}_i = \frac{\mathcal{A}_{-i}}{\mathcal{A}_i + \mathcal{A}_{-i}} \quad (3)$$

where $\mathcal{A}_i$ denotes the modeling accuracy of agent i regarding the opponent’s preferences, and $\mathcal{A}_{-i}$ is the opponent’s modeling accuracy of agent i . Accuracy is defined as the normalized Kendall’s Tau correlation:

$$\mathcal{A} = \frac{1 + \tau(U_{\text{true}}, U_{\text{model}})}{2} \in [0, 1] \quad (4)$$

where $\tau \in [-1, 1]$ measures the rank correlation of outcomes.

This dual-scoring system creates a game-theoretic dilemma. To negotiate a mutually beneficial agreement (e.g., targeting the Pareto frontier or the Nash Bargaining Solution), agents must share information

through their bids. However, sharing information reveals their preference rankings, which increases the opponent’s modeling accuracy and severely penalizes their concealing score.

To resolve this trade-off, we design **Whale**, an agent that stochastically deforms its bidding behavior. Whale presents a highly-tuned time-based and step-based concession framework that conceals preferences early in the negotiation using dynamic decoy values, while constructing a robust model of the opponent to target high-welfare agreements.

2 The Design of Whale

Whale is built on the modular **BOAD** (Bidding, Opponent modeling, Acceptance, Deception) framework. The coordinator class, **WhaleNegotiator**, manages four core components:

1. **Tactic Selector**: Profiles the opponent online and determines the appropriate tactic bundle.
2. **Bidding Strategy**: Proposes candidate bids using time-based or step-based concession.
3. **Acceptance Strategy**: Evaluates incoming bids using a strict, adaptive threshold.
4. **Deception Strategy**: Applies value substitution on high-impact issues to mislead the opponent.

Whale features a **TacticSelector** that monitors the sequence of incoming offers. After an initial profiling period of $N_{\text{classify}} = 8$ steps, and every $N_{\text{update}} = 10$ steps thereafter, the selector classifies the opponent’s behavior into one of several behavioral profiles: Boulware, Conceder, Linear, Stubborn, Random, Deceptive, or Unknown. Based on the classification, it hot-swaps the underlying bidding and acceptance strategies to optimize individual advantage while maintaining active concealment.

3 Concealing Bidding Strategy

Whale’s bidding module is split into two primary components: the concession generator and the preference deformer.

3.1 Concession Generators

Whale implements three bidding algorithms:

- **NashBidder**: The default bidder. It targets the Nash Bargaining Solution (NBS) by solving:

$$x_{\text{Nash}} = \arg \max_{x \in \mathcal{P}} (U_{\text{us}}(x) - RV_{\text{us}})(U_{\text{opp}}(x) - RV_{\text{opp}}) \quad (5)$$

where $\mathcal{P}$ is the set of Pareto-optimal outcomes. The bidder concedes from $U_{\text{us}}(\text{Best})$ to $U_{\text{us}}(x_{\text{Nash}})$ according to a Boulware time-concession curve:

$$\theta(t) = (U_{\text{us}}(\text{Best}) - U_{\text{us}}(x_{\text{Nash}})) \cdot (1 - t^e) + U_{\text{us}}(x_{\text{Nash}}) \quad (6)$$

where $e = 16.2$ is a highly-tuned concession exponent.

- **BoulwareBidder**: A fallback time-based concession model used when opponent modeling is disabled or highly uncertain. It concedes towards RV_{us} :

$$\theta(t) = (U_{\text{us}}(\text{Best}) - RV_{\text{us}}) \cdot (1 - t^e) + RV_{\text{us}} \quad (7)$$

- **StepBasedBidder**: Inspired by UOAgent (ANL 2024 Individual Advantage runner-up), this bidder uses the discrete step counter instead of wall-clock time. It aggressively filters the candidate pool to high-utility outcomes:

$$U(x) \geq \max(RV_{\text{us}} + \Delta_{\text{low}}, \text{Floor}_{\text{high}}) \quad (8)$$

with $\Delta_{\text{low}} = 0.40$ and $\text{Floor}_{\text{high}} = 0.80$. It selects bids from this filtered pool using a softmax distribution:

$$P(x) \propto \exp(\lambda \cdot \text{Score}(x)) \quad (9)$$

where $\lambda = 5.0$ and $\text{Score}(x) = U_{\text{us}}(x) + w_{\text{pareto}}U_{\text{opp}}(x)$ with $w_{\text{pareto}} = -0.2$ (an anti-social setting to force selfish bidding).

3.2 Dynamic Adaptive Deception (DAD)

The deception module (**IssueDecoyDeceiver**) is Whale’s key contribution. It intercepts the candidate bid x proposed by the bidding module and deforms it to hide preferences. The DAD mechanism consists of three parts:

3.2.1 Multi-Outcome Impact Estimation

To identify which issues reveal the most information, DAD computes the average utility impact of each issue j by sampling $M = 15$ outcomes:

$$\text{Impact}(j) = \frac{1}{M} \sum_{k=1}^M \left[\max_{v \in \mathcal{V}_j} U_{\text{us}}(x_{-j}^{(k)} \cup \{v\}) - \min_{v \in \mathcal{V}_j} U_{\text{us}}(x_{-j}^{(k)} \cup \{v\}) \right] \quad (10)$$

where $x_{-j}^{(k)}$ is a sampled outcome excluding issue j , and $\mathcal{V}_j$ is the set of possible values for issue j . Issues with an impact higher than the median and above a threshold $\epsilon = 0.01$ are classified as **high-impact issues**.

3.2.2 Preference Poisoning and Best-Value Hiding

For each high-impact issue, DAD ranks the values by their average utility contribution. The absolute best value is completely hidden (never used in decoys) to avoid revealing the peak preference. DAD designates the second-best average value as the primary decoy. When constructing a bid, the values on high-impact issues are stochastically replaced with decoy values. To prevent the opponent from detecting a static pattern, value choices are selected via a count-penalized weight:

$$W(v) = \frac{1}{\text{count}(v) + 1} \quad (11)$$

This forces the agent to cycle through multiple decoy values, maximizing the entropy of the issue-value frequency counts observed by the opponent.

3.2.3 Phase-Based Dynamic Utility Budgeting

Deception carries a utility cost. Replacing a preferred issue value with a decoy reduces the utility of the bid for Whale. To maintain competitive advantage, the maximum utility sacrifice is strictly bounded by a budget $B(t)$ that varies across four negotiation phases:

$$B(t) = \begin{cases} 2 \cdot B_{\text{base}} & \text{if } t < 0.40 \quad (\text{Phase 1: Aggressive Poisoning}) \\ B_{\text{base}} & \text{if } t \in [0.40, 0.85) \quad (\text{Phase 2: Steady Deception}) \\ B_{\text{base}} \cdot \frac{0.95-t}{0.10} & \text{if } t \in [0.85, 0.95) \quad (\text{Phase 3: Linear Decay}) \\ 0 & \text{if } t \geq 0.95 \quad (\text{Phase 4: Guarantee Convergence}) \end{cases} \quad (12)$$

where $B_{\text{base}} = 0.07$. In Phase 3, the budget linearly decays to 0.0, ensuring that as the deadline approaches, Whale stops deceiving and focuses purely on reaching an agreement. The final bid x' is constrained such that:

$$U_{\text{us}}(x') \geq \max(RV_{\text{us}}, \theta_{\text{bidder}}(t)) \quad (13)$$

This ensures that preference concealment never forces a concession below the bidder’s targeted concession threshold.

4 Acceptance Strategy

Whale provides two acceptance strategies:

- **ThresholdAcceptor**: An adaptive threshold policy. Early in the negotiation, it is extremely strict to avoid exploitation:

$$\theta_{\text{accept}}(t) = RV_{\text{us}} + (U_{\text{us}}(\text{Best}) - RV_{\text{us}}) \cdot (1 - t^{50}) \quad (14)$$

This steep acceptance curve relaxes near the deadline. It also incorporates a **concession bonus**: if the opponent’s offer increases their utility for us by at least $\Delta_{\text{concession}} = 0.03$, the acceptance threshold is scaled down by a factor of 0.97 to encourage agreement. A failsafe panic mode starts at $t = 0.995$, accepting any offer that yields at least $RV_{\text{us}} + 0.07$.

- **ACNextAcceptor**: The default acceptor. It accepts the opponent’s offer x_{opp} if it meets the criteria of the **ThresholdAcceptor** OR if it is at least as good as the next bid Whale plans to offer:

$$U_{\text{us}}(x_{\text{opp}}) \geq U_{\text{us}}(x_{\text{next}}) \quad (15)$$

5 Opponent Model

To compute the Nash Bargaining Solution and make rational concessions, Whale models the opponent’s utility function. It supports three distinct models:

5.1 GroundTruthModel

Following a strategic design pattern from Shochan, Whale checks if the opponent’s true utility function is accessible through the tournament’s metadata (**private_info**). If available, Whale leverages this ground truth function. Crucially, Whale resets the model’s reservation value to 0.0:

$$RV_{\text{opp_model}} = 0.0 \quad (16)$$

This gives Whale significant bargaining power, as the Nash Bargaining calculation assumes the opponent is willing to accept low-utility agreements, shifting the Nash point in Whale’s favor.

5.2 FrequencyModel

When ground truth is unavailable, Whale constructs an issue-value frequency model. The utility of an outcome x is estimated as:

$$U_{\text{opp_est}}(x) = \frac{1}{|I|} \sum_{j \in I} \frac{\text{count}(x_j)}{N_{\text{offers}}} \quad (17)$$

where I is the set of issues and N_{offers} is the total number of offers received. The opponent’s reservation value is estimated online by tracking their lowest offered utility and subtracting a margin:

$$RV_{\text{opp_est}} = \max \left(0.0, \min_k U_{\text{opp_est}}(x^{(k)}) - \delta \right) \quad (18)$$

where $\delta = 0.02$.

5.3 NullModel and Deception Defense

Deceptive opponents poison frequency-based models. If Whale uses a standard frequency model against a deceptive opponent, its Nash Bargaining calculations will target poisoned outcomes, leading to sub-optimal agreements.

To defend against this, Whale monitors the opponent’s offers. An opponent is flagged as **Deceptive** if two signals are simultaneously triggered:

1. **High Value Coverage**: The average fraction of possible values offered across all issues exceeds $\text{Threshold}_{\text{coverage}} = 0.45$. This detects the high value diversity characteristic of preference poisoning.
2. **Erratic Utility Jumps**: The estimated utility of consecutive offers shows a maximum jump of $\Delta_{\text{jump}} > 0.087$ over a range of $\text{Range}_{\text{jump}} > 0.049$. This flags non-conceding volatility.

If the opponent is flagged as deceptive, Whale’s selector hot-swaps the opponent model to the **NullModel**, which returns a constant utility of 0.5 for all outcomes. This stabilizes the Nash targeting and prevents the agent from making concessions based on poisoned data.

Table 1: Comparison of Whale against standard agents under ANL 2026 scoring.

Agent Configuration	Advantage	Concealing	Total Score
Standard Boulware Baseline	0.421	0.315	0.736
BOA (GSmith Frequency Model)	0.485	0.292	0.777
Whale (No Deception)	0.512	0.304	0.816
Whale (Full DAD Active)	0.478	0.785	1.263

6 Evaluation

We evaluated Whale against several standard benchmarks using the ANL 2026 CLI. Table 1 summarizes the performance across 20 random scenarios.

The results show that without deception, Whale achieves a high Advantage (0.512) due to its Nash-targeting bidder and Shochan-inspired RV tricks, but suffers a low Concealing score (0.304) because its bidding sequence reveals its utility function. When the DAD module is enabled, Whale sacrifices a small fraction of advantage (from 0.512 to 0.478) but increases its Concealing score to 0.785. This results in a total score of 1.263, outperforming all baseline configurations.

7 Lessons and Suggestions

Our development of Whale highlighted several key insights for preference-concealing negotiations:

- **Deception Budget Tuning:** The budget parameter B_{base} represents a strict trade-off. Setting it too high (> 0.15) results in severe advantage loss, as the agent proposes bids that are too far from its true preference. Setting it too low (< 0.03) fails to hide the preference rankings.
- **Endgame Convergence:** Failing to decay the deception budget near the deadline leads to high rates of negotiation timeouts. The phase-based decay mechanism in DAD is essential for ensuring that the agent converges to rational agreements before the deadline.
- **Robustness via Null Models:** Opponent modeling is a double-edged sword. If an agent relies on an opponent model that is successfully poisoned, its negotiation utility drops. Swapping to a neutral model (like `NullModel`) when deception is detected is a highly effective countermeasure.

8 Conclusions

In this report, we presented the design and evaluation of **Whale**, an autonomous negotiating agent built for the ANL 2026 preference concealment track. Whale implements a modular BOAD architecture that integrates dynamic tactic switching, step-based and Nash-targeting concession, and an advanced Dynamic Adaptive Deception (DAD) module. By employing phase-based utility budgeting, count-penalized value cycling, and an anti-deception Null Model, Whale successfully misleads opponent models while securing highly beneficial agreements. Empirical benchmarks demonstrate the superiority of this design, establishing Whale as a strong competitor for the ANL 2026 tournament.