

Agent360: A Concealment-First Bilateral Negotiator for ANL 2026

Noam Kazum

Omer Shani Steinmetz

Dr. Galit Haim

Dr. Raz Lin

The College of Management Academic Studies

June 15, 2026

Abstract

We present **Agent360**, a bilateral negotiator submitted to the Automated Negotiation League (ANL) 2026. The competition scores each agent on **Advantage** (deal quality relative to a reserved value) and **Concealing** (how poorly the opponent learns our true preferences from our bid stream). Our agent design follows a *concealment-first* principle: early offers deliberately misrepresent issue priorities so that frequency-based opponent learners infer an incorrect utility function, while later phases and acceptance rules extract profitable agreements. The agent uses no oracle access to the opponent’s class or true utility—only observed offers, negotiation seat, and the relative deadline time. We relate the bidding persona to theories of signaling, strategic misrepresentation, and the limits of honesty in non-cooperative negotiation.

1 Introduction

In bilateral alternating-offer negotiation, each proposal is both an economic act and a *signal* about private preferences. The ANL 2026 league makes this dual role explicit: an agent is rewarded not only for closing good deals, but also for preventing the opponent from building an accurate model of its true utility function. Formally, the league score is

$$\text{Score} = \text{Advantage} + \text{Concealing}, \quad (1)$$

where concealment is derived from the rank correlation between the agent’s true utility and the utility function the opponent inferred from the negotiation transcript.

Most student agents in the league learn from offers using Smith-style frequency models or similar adaptive estimators. Against such opponents, bidding one’s sincere top outcomes early tends to maximize short-term deal quality but minimizes concealing, because the opponent’s learner quickly converges on the correct issue weights. Agent360, therefore, treats preference *concealment* as a first-class design goal and tunes deal extraction only in ways that do not collapse the decoy persona.

2 Contributions

Our main contributions are:

- A **concealment-first bidding persona**: maximal-mismatch decoy offers, gradual transition, mode-aware closing, and seat-asymmetric profiles (extended decoy windows and stricter min-offer gates when opening).
- A **behavioral opponent classifier** (mirror, learner, deceptive, conceding, unknown) and **multi-layer Smith model** (recency, time-weighted, issue-weighted blends) with trajectory-based bait detection, all from the bid stream.
- **Mode-dependent deal extraction**: aspiration slopes, closing caps, bait discounts, conceding-triggered early decoy exit, and anti-mirror bid perturbation.
- **Empirically validated design choices**, including rejection of a reverse-psychology persona and mid-game escape acceptance that regressed against learner-heavy and deceptive opponent mixes.

3 Background: Deception, Signaling, and Learning

3.1 Offers as preference signals

In multi-issue negotiation, an offer is a vector of issue values. Rational opponents treat repeated offers as evidence about the values the sender prefers. Smith-style frequency models count how often each value appears in the opponent’s bids, implementing a simple form of *preference elicitation from behavior*. If our early bids consistently favor values that are actually low-priority for us, the opponent’s estimated utility function shifts away from the truth, which directly increases concealment under the league metric.

This is distinct from verbal deception: ANL agents cannot negotiate in natural language. All misdirection must be encoded in the *bid stream*, analogous to non-verbal signaling in human bargaining (consistent choices that suggest false priorities).

3.2 Truth mechanisms versus competitive misrepresentation

Mechanism-design results (e.g. strategy-proof auction formats) show that honesty can be a dominant strategy when the *rules* reward truthful reporting. The ANL scoring rule has the opposite incentive on the concealment axis: if the opponent models the agent correctly, our concealment score falls. We therefore operate in a non-cooperative setting closer to competitive signaling than to incentive-compatible mechanism design.

We initially considered a *reverse-psychology* persona: bid sincere high-utility outcomes early (as if “playing truth”), then shift toward misdirection, exploiting opponents that systematically discount our offers. Local experiments showed that this variant improved advantage against some time-based negotiators but reduced concealment against BOA, MAP, and MiCRO-style learners—the dominant opponent family in both our benchmarks and the live tournament. We rejected reverse-truth as the submission strategy and retained a gradual decoy-to-truth persona instead.

4 Agent Architecture

Agent360 extends the NegMAS SAOCallNegotiator protocol with three cooperating layers.

1. **Bidding persona** — what we offer: decoy pool construction, phased transition, and closing bids.
2. **Opponent model** — what we infer: blended Smith estimators, trajectory features, and mode classification.

3. **Deal extraction** — when we accept: time and mode-dependent aspiration, bait detection, and deadline safeguards.

On each step with a partner offer, the agent updates all estimators, evaluates acceptance, and otherwise samples the next bid from the current phase. The first negotiator (seat 0) follows a stricter concealment profile because opening agents expose more of their bid stream to the opponent’s learner.

5 Concealing Bidding Strategy

Relative time $t \in [0, 1]$ divides the negotiation into three phases.

Decoy phase. We bid from a **maximal-mismatch decoy pool**: rational outcomes that disagree with our inferred true top-issue values on at least half of the issues, subject to a utility floor. When we open, transition is blocked until the opponent has made at least four offers reducing early leakage to curve-fitting learners. Decoy selection avoids recent repeats, breaks mirror tit-for-tat by perturbing echoed bids, and prefers offers with large utility jumps relative to our last bid. The first seat extends the decoy window to $t < 0.40$ (versus 0.35 for the second seat).

Transition phase ($t < 0.75$). We mix decoy outcomes with a widening band of true high-utility outcomes. The first negotiator keeps decoy mixing for most of the transition window and reveals true preferences more slowly (`FIRST_TRANSITION_PROGRESS_SCALE` = 0.72). If the opponent is classified as conceding and the trajectory shows honest concession, we allow an *early exit* from decoy even before the usual time gate—trading a small amount of concealing for advantage when the partner is unlikely to exploit the signal.

Closing phase ($t \geq 0.75$). We sample rational outcomes and score each candidate by a blend of normalized self-utility and estimated opponent utility. The opponent weight increases through closing but is capped by mode (e.g. higher caps against learners and conceders, lower against deceptive or mirroring opponents). Against deceptive late bait-switching, inflated Smith scores are discounted when selecting closing bids.

Together, these mechanisms implement *gradual preference revelation*: the opponent’s learner receives a sustained false early signal, then gradual correction only when deal extraction dominates further concealment.

6 Opponent Modeling

All opponent inference uses only the observed bid stream, relative time, and negotiation seat. We never read the opponent’s agent class or true utility function. The stack estimates opponent values, classifies negotiation behavior, and publishes a blended utility function for closing and acceptance logic.

6.1 Smith-style frequency estimation

The base estimator is a **Smith frequency model**: for each issue, we count how often the opponent chose each value in past offers. Values repeated often are treated as more valuable to them; per-issue scores are normalized and averaged into a scalar utility for any outcome. This matches the learning bias of many league agents (BOA, MAP, MiCRO) and provides a stable prior over the full negotiation.

We maintain three refinements of this idea:

- **Recency-blended Smith** (second seat): blend full-history with recent-only Smith when at least two offers exist in a window of five; weight on recency grows up to ≈ 0.68 .
- **Time-weighted Smith**: bids after $t \geq 0.40$ are counted three times, reducing contamination from early decoy posturing.
- **Issue-weighted Smith**: issues with low value spread (repeated values) are down-weighted; diverse issues receive higher weight.

6.2 Trajectory and concession dynamics

The **offer trajectory model** records pairs $(t, \hat{u}_{\text{Smith}})$ at each opponent bid and treats the stream as a time series of revealed preference strength. From it we compute:

- **Concession slope** — linear trend of Smith utility over time; slopes below ≈ -0.04 indicate honest monotonic concession.
- **Non-monotonicity** — two or more significant utility *increases* suggest bait-and-switch rather than steady concession.
- **Predicted utility at t** — extrapolation used to detect late offers inconsistent with the opponent’s prior trajectory.

Trajectory features feed both mode classification and bait rejection at acceptance time.

6.3 Behavioral mode classification

Rather than routing by opponent *type* (which is unknown in competition), we classify *behavior* into five modes: **mirror**, **learner**, **deceptive**, **conceding**, and **unknown**. The decision tree uses mirror echo detection first (at least three matches to our recent bids in a window of four opponent offers), then trajectory sample count, Smith concentration, concession slope, and concealment-tactic heuristics.

Deceptive signals. We label an opponent deceptive when their bid stream shows concealment tactics, for example: identical early offers, high early issue-flip rate (≥ 0.25), wide early Smith spread (≥ 0.12), non-monotone utility paths, or late bait-switching (preferred values flip on many issues after $t \geq 0.40$). Plain BOA/MAP-style learners usually fall into **learner**, not **deceptive**, which limits false bait triggers against honest frequency learners.

Table 1 summarizes how each classified mode retunes closing, acceptance, Smith blending, and persona behavior.

Mode-dependent control. The closing cap limits how much opponent utility weights closing bids; the aspiration slope controls how quickly we lower our acceptance threshold over time. Against deceptive opponents, inflated Smith estimates are discounted before bid scoring and late offers may be rejected as bait when they diverge from the trajectory prediction. Conceding opponents may trigger early decoy exit; mirror opponents use plain Smith only and anti-mirror bid perturbation when echoing our last offer.

6.4 Published opponent utility blend

The league scores Concealing from how well the opponent models us; symmetrically, we publish `opponent_ufun` so the opponent can score Advantage. The published estimate is built in stages:

1. Start from full-history Smith.

Table 1: Classified opponent mode and agent response.

Mode	Cap	Slope	Smith	Bait	Notes
Conceding	0.52	0.42	Full blend	No	Early decoy exit
Learner	0.48	0.52	Full blend	No	Self-utility boost late
Deceptive	0.32	0.58	Discounted	Yes	Bait reject active
Mirror	0.38	0.55	Plain Smith	No	Anti-mirror bids
Unknown	0.40	0.55	Full blend	No	Conservative de- fault

2. If enough late bids exist ($t \geq 0.40$), blend toward late-phase Smith (cap ≈ 0.55 – 0.68 , higher when we open).
3. If at least three late offers exist, blend toward issue-weighted late Smith.
4. On the second seat, blend toward recency Smith; for learner/conceding/unknown modes, add a stable-issue match term when recency is strong.
5. If mode is `mirror`, skip adaptive blends and return plain Smith.

Closing and bait logic query this pipeline; deceptive opponents may receive discounted Smith estimates before scoring candidate agreements.

6.5 Anti-mirror response

When the opponent repeats our last bid, mirror detection fires and we break synchrony by perturbing the next offer from an anti-mirror pool (small random change on one issue). Tit-for-tat mirroring freezes preference learning on both sides without improving advantage.

7 Acceptance Strategy

The agent accepts a partner’s offer when any of the following holds:

- **Catastrophe floor:** $t \geq 0.95$ and utility $\geq$ reserved value.
- **Mode aspiration:** utility $\geq u_{\max}(1 - \text{slope} \cdot t)$ and above reserved value; the aspiration slope depends on the classified opponent mode.
- **AC-next:** offer utility is at least that of our next concealing bid.
- **Deadline safe:** $t > 0.90$ and utility above reserved value.

Before the deadline-safe window ($t \leq 0.90$), offers that pass aspiration may still be rejected if they appear to be **bait**: the opponent is classified as deceptive, shows concealment tactics, and the Smith-estimated utility exceeds the trajectory prediction by more than a threshold (`ACCEPT_BAIT_THRESHOLD` = 0.14).

During development, we tested mid-game *escape* acceptance above reserved value from $t \approx 0.88$ onward to reduce timeouts. Tournament and proxy runs showed it accepted too many mediocre or deceptive agreements; the submitted agent relies on catastrophe accept near timeout ($t \geq 0.95$) instead.

8 Evaluation and Development Process

We evaluated each candidate agent on a repeatable local panel before submission:

- **Learners:** BOA, MAP, MiCRO across competition scenarios (both seats).
- **Stress:** time-based NegMAS negotiators (Boulware, Conceder, Tough, etc.).
- **Sparring:** in-house deceptive opponents (mirror, bait-switch, strong learner variants).

Each matchup cell is repeated with fixed step counts; we report mean Advantage, Concealing, and their sum.

Table 2 summarizes a representative development snapshot on the combined panel.

Table 2: Local proxy panel (mean Score per matchup cell; development snapshot).

Opponent group	Advantage	Concealing	Score
Learners (BOA/MAP/MiCRO)	0.56	0.71	1.27
Stress (time-based)	0.53	0.90	1.43
Deceptive sparring	0.61	0.58	1.19
Panel mean	0.55	0.79	1.34

Iterative refinement. The path from an early submission to the current Agent360 added conceding early decoy exit, mode-dependent closing caps, first-seat decoy rotation, anti-mirror pools, and catastrophe acceptance near timeout. Changes that improved isolated local cells but hurt the broader distribution—stricter first-seat gates, escape acceptance, and stall-based accept rules—were removed after broader evaluation showed regressions.

9 Discussion and Conclusions

Agent360 is a concealment-first bilateral negotiator grounded in signaling theory and empirical tests of alternative personas. Early maximal-mismatch decoys mislead frequency learners, gradual transition limits preference leakage, and mode-aware closing and acceptance extract Advantage without surrendering Concealing to deceptive or mirroring opponents.

Three lessons emerged from development. First, **concealing and advantage require separate controls**; strengthening acceptance without preserving the decoy persona hurts the combined score against learner-heavy fields. Second, **evaluation must span multiple opponent families**; learner, stress, and deceptive sparring panels together catch failures that no single benchmark reveals alone. Third, **seat asymmetry matters**; first and second negotiators see different information, and asymmetric decoy length and min-offer gates improved learner matchups without oracle routing.

The submitted agent operates without privileged information, adapts to inferred opponent behavior, and balances strategic misrepresentation with profitable agreement formation.