CS 580 · Algorithmic Game Theory

LLMs as Strategic Agents

Beliefs, best responses, equilibrium play, and bounded rationality

Games and Markets with AI Agents

What makes an LLM a strategic agent?

A language model generates responses to its input.

An agent uses those responses to choose actions in an environment.

Interaction becomes strategic when an action's value depends on what others do.

A bidding agent chooses a bid while anticipating rival bids.

Strategic competence is a behavior to evaluate.

Human players and LLM agents

Dimension	Human experiment	Fixed-model LLM experiment
Incentives	Payments can affect the participant’s welfare.	Payoff text (prompt) supplies a task. Training loss/incentives are separate.
Adaptation	Prior experience can affect later choices.	Prompt can change behavior while weights stay fixed.
Experimental unit	An individual with a prior history.	Model weights plus a specified prompts/inputs

Both can exhibit framing effects, mistakes, and limited reasoning.

Compare observed behavior under matched games and information.

Brown et al. (2020); Ouyang et al. (2022); Junqué de Fortuny & Cappelli (2026).

Training, instructions, and user utility

Object	Role in the bidding example
Training objective	Changes model parameters using training data and feedback.
Task instruction (prompt)	Requests a behavior: “maximize expected surplus.”
User utility	Measures the consequence: value of the item minus payment.

Maximizing user utility is a behavior to test.

Entering a payoff in the prompt does not itself train the model.

Training background: Ouyang et al. (2022). Bidding example constructed for this lecture.

The agent includes its decision procedure

Agent design	How it chooses a bid
Direct generation	Generate a response and parse the proposed bid.
Forecast plus optimizer	Predict rival bids, then compute an exact best response.
Agent with memory	Use prior observations and tools before choosing a bid.

Same model weights, different induced policies.

Specify prompts, memory, tools, computation budget, and decoding.

Text generation and game actions

Response distribution

$$P_{\theta,c}(z \mid x) = \prod_{t=1}^{T} p_{\theta,c}(z_t \mid x, z_{<t})$$

Action distribution

$$\pi_{\theta,c}(a \mid x) = \sum_{z:\, D_c(z)=a} P_{\theta,c}(z \mid x)$$

x: available information and history z: complete generated response

D: response-to-action rule theta: weights c: agent configuration

Include invalid output as an outcome.

Operational policy definition; autoregressive model background: Brown et al. (2020).

A behavioral model makes extra assumptions

Example: a logit response to expected utility under an opponent belief.

$$\pi_{\theta,c}(a \mid x) \approx \frac{\exp(\lambda U(a;q))}{\sum_{b\in A}\exp(\lambda U(b;q))}$$

$$U(a;q) = \mathbb{E}_q[u(a, A_{-i})]$$

Assumptions: utility u, opponent belief q, and response precision lambda.

The payoff benchmark specifies what the agent should achieve.

The fitted model must predict choices in new games.

Behavioral model application: Jia et al., arXiv:2502.20432v3, §§2–3.

Evidence for a behavioral model

A correct action can result from several different procedures.

A reported belief is an output that we can test against outcomes.

A fitted reasoning depth describes choices under model assumptions.

Internal recursion requires additional evidence.

Test predictions on changed payoffs, labels, and new games.

Evaluation framing: Junqué de Fortuny & Cappelli; Jia et al. See notes for scope.

Running example: an AI bidding agent

Your value is 10. In a first-price auction, you pay your bid if you win.

Your allowed bids are 4, 6, and 8. Losing yields zero.

Predicted highest rival bid

Rival bid	3	5	7
Probability	0.50	0.30	0.20

Which bid should the agent choose?

The rival bids are 3, 5, and 7, so ties cannot occur in this example.

Beliefs and best responses

A belief is a distribution over the opponents' joint actions.

Expected payoff

$$U_i(a_i; q_i) = \sum_{a_{-i} \in A_{-i}} q_i(a_{-i}) u_i(a_i, a_{-i})$$

Best responses

$$\mathrm{BR}_i(q_i) = \arg\max_{a_i \in A_i} U_i(a_i; q_i)$$

Choice gap

$$g_i = \max_{a_i \in A_i} U_i(a_i; q_i) - U_i(\hat{a}_i; q_i)$$

A zero gap means that the action is optimal under the stated belief.

The best bid depends on the forecast

$$U(b; q) = (10 - b) \Pr_{a_{-i} \sim q}(a_{-i} < b)$$

Rival bid	3	5	7
Probability	0.50	0.30	0.20

Bid (b)	Win probability	Surplus if won	Expected utility
4	0.50	6	3.00
6	0.80	4	3.20
8	1.00	2	2.00

Bid 6 maximizes expected utility. Bid 8 maximizes win probability.

Belief error and choice error

The actual win probabilities for bids 4, 6, and 8 are 0.5, 0.8, and 1.

Reported win probabilities	Chosen bid	Gap under report	Diagnosis
0.2, 0.4, 1	8	0	Wrong belief/forecast, coherent choice
0.5, 0.8, 1	4	0.2	Correct belief, choice error
0.5, 0.8, 1	6	0	Correct forecast, coherent choice

An action-only score cannot tell us which component failed.

Nash equilibrium adds mutual consistency

$$\mathrm{supp}(\sigma_i) \subseteq \mathrm{BR}_i(\sigma_{-i}) \qquad \text{for every } i$$

A best response optimizes against a specified forecast.

At equilibrium, each forecast matches the strategies actually used by the others.

In an auction, the rival-bid distribution comes from other bidders' policies.

One bidder's optimal action alone does not establish equilibrium.

Evaluating the opponent forecast

Ask for probabilities before observing the rival’s bid.

A proper scoring rule rewards accurate probability forecasts.

Brier loss: $$L(q, m) = \sum_{z} \Big(q(z) - \mathbf{1}\{z = m\} \Big)^2$$

Expected Brier loss is minimized by the true distribution.

Use many trials to assess calibration and predictive accuracy.

Level-k reasoning

Level 0 follows a specified baseline policy.

Level k best responds to opponents at level k minus 1.

$$\mathrm{supp}(\sigma_i^{(k)}) \subseteq \mathrm{BR}_i\left(\sigma_{-i}^{(k-1)}\right)$$

A baseline could be uniform play, a focal action, or measured behavior.

An inferred depth depends on the baseline and the game.

Background: Nagel (1995), Arad & Rubinstein (2012).

Classroom game: the beauty contest

Choose a number between 0 and 100, simultaneously.

The target is two-thirds of the class average.

The closest number wins. Ties split the prize.

Record your predicted average as well as your chosen number.

The all-zero profile is a Nash equilibrium.

Game background: Nagel (1995).

The level-k target calculation

Depth	Calculation	Forecasted Average
0	Anchor	50
1	50 times (2/3)	33.33
2	50 times (2/3) squared	22.22
3	50 times (2/3) cubed	14.81

Repeatedly applying the multiplier drives the target toward zero.

For the prize game, this is a target heuristic.

Exact optimality depends on the payoff and opponent model.

Greater depth can lower payoff

Teaching variant: the opponent's number is fixed at 50.

$$u(a, y) = -\left(a - \frac{2}{3}y\right)^2, \qquad y = 50$$

The optimal action is 100/3, approximately 33.33.

Action	33.33	22.22	14.81	0
Payoff, approx.	0	−123.46	−342.94	−1111.11

More recursion helps only when it improves the opponent model.

Cognitive hierarchy: mix lower levels

For two players, a level-k agent mixes the opponent's lower-level policies.

$$q_i^{(k)} = \sum_{h=0}^{k-1} w_h^{(k)} \sigma_{-i}^{(h)}$$

$$w_h^{(k)} = \frac{f(h)}{\sum_{\ell=0}^{k-1} f(\ell)}$$

The weights renormalize a population distribution f over levels below k.

A common choice is Poisson f, with mean depth tau.

At k = 2 and tau = 1, levels 0 and 1 each receive weight one-half.

Logit quantal response

$$\Pr(a \mid q) = \frac{\exp(\lambda U(a; q))}{\sum_{x \in A} \exp(\lambda U(x; q))}$$

$\lambda =$ precision.

Zero precision gives uniform choice over actions.

Higher precision favors actions with higher expected utility.

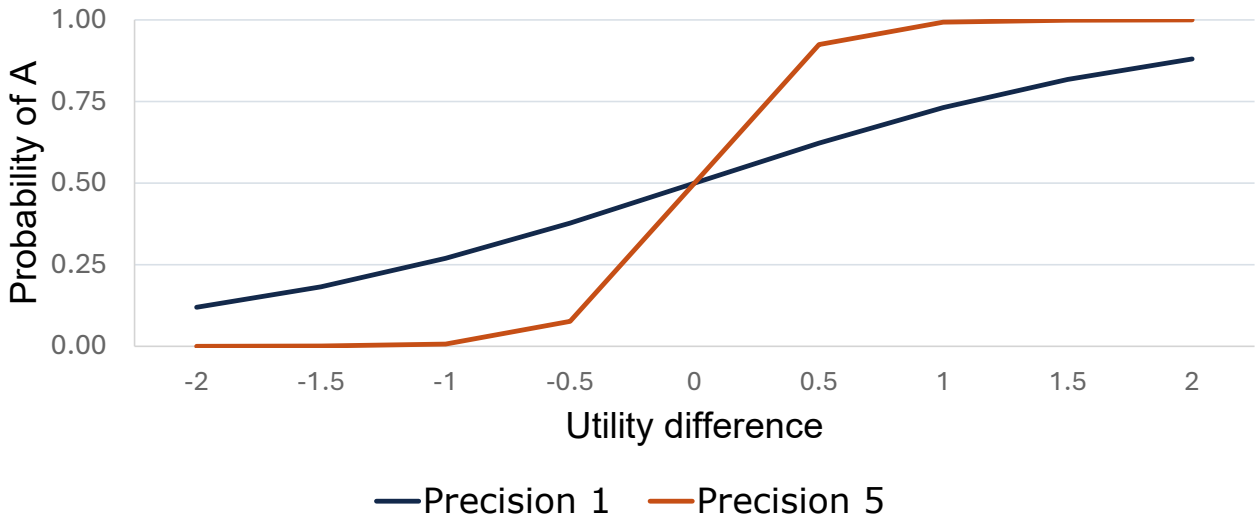
At very high precision, probability concentrates on best responses.

λ is different from the token-decoding temperature.

Goeree, Holt & Palfrey (2018), §I.

A smooth response to payoff differences

$$\Pr(A) = \frac{1}{1 + e^{-\lambda \Delta}}, \qquad \Delta = U(A) - U(B)$$



Bidding example

Compare bids 6 and 4.
Their utility gap is 0.2.

Precision 1: about 55%
Precision 5: about 73%

Calculated curves, not measured LLM frequencies. The comparison here uses two actions.

Response rules and equilibrium

Object	Opponent distribution	Requirement
Best response	Specified belief	Choose a maximizer
Logit response	Specified belief	Use the logit probabilities
Nash equilibrium	Others' actual policies	Mutual best responses
Logit QRE*	Others' actual policies	Mutual logit responses

* $\sigma_i = \mathrm{QR}_i(\sigma_{-i}) \qquad \text{for every } i$

A finite recursion of noisy responses need not solve this fixed point.

Goeree, Holt & Palfrey (2018), §I.

Study 1: conjectures and choices

Experimental game	Role in the evaluation
Beauty contest	Recursive expectations
11–20 money request	Cyclic best responses
Unlabeled matrix	An unfamiliar payoff structure

Compare supplied conjectures with unconstrained play.

Analyze choices and reported reasoning across 100 runs per task environment.

Study 1: strengths and limits

Reported pattern	Question for our bidding agent
Models can follow supplied conjectures.	Can it compute utilities from a supplied bid forecast?
Opponent identity changes conjectures and play.	Does the forecast improve on actual rival behavior?
Complexity can trigger heuristic shifts.	Does it replace surplus maximization with a shortcut?

Conditional competence can coexist with fragile behavior.

The bidding questions are proposed applications, not reported auction results.

The 11–20 money-request game

Two players simultaneously request an integer from 11 to 20.

$$u(a, b) = a + 20\,\mathbf{1}\{a = b - 1\}$$

Receive your request, plus 20 if it is exactly one below the other request.

Against request 20, choosing 19 earns 39.
Against request 11, choosing 20 earns 20.

Which request would you choose against an unknown opponent?

Arad & Rubinstein (2012), AER 102(7), 3561–3573.

Best responses cycle

Opponent request	20	19	18	…	12	11
Best response	19	18	17	…	11	20

There is no pure equilibrium.

Starting at 20, ten recursive best responses return to 20.

One action can correspond to many different reasoning depths.

A symmetric mixed equilibrium

$$U(a;q) = a + 20q(a+1), \qquad q(21) = 0$$

Action 20 earns 20. Try support 15 through 20 and impose indifference.

$$q(a+1) = \frac{20-a}{20} \qquad \text{for } a = 15, \ldots, 19$$

Request	15	16	17	18	19	20
Probability	0.25	0.25	0.20	0.15	0.10	0.05

Outside the support: action 14 earns 19; actions 11–13 earn their request.

Every supported action earns 20, so this profile is a Nash equilibrium.

Randomness does not imply equilibrium

Consider independent uniform play over all ten requests.

$$U(a;q) = \begin{cases} a+2, & a \le 19, \\ 20, & a = 20. \end{cases}$$

Best response: 19, which earns 21.

Uniform play earns 17.3 against a uniform opponent.

$$\text{Deviation gain} = 21 - 17.3 = 3.7$$

Test both action frequencies and profitable deviations.

A matrix separates candidate rules

Constructed common-interest game: both players receive the displayed payoff.

	L	M	R
A	9	0	1
B	6	6	6
C	0	8	4

Uniform opponent

Action A: 10/3
Action B: 6
Action C: 4

Under probabilities (0.8, 0.1, 0.1), action A earns 7.3 and B earns 6.

A choice can fit several rules until we change the opponent model.

Interventions identify behavioral rules

Intervention	What it helps distinguish
Supply and vary the opponent distribution.	Forecast errors versus response errors
Permute action labels and positions.	Payoff reasoning versus positional shortcuts
Change the task while matching compute.	Transfer versus prompt or budget effects
Test on new games and opponent populations.	Generalization versus fit to known instances

Record exact configurations, invalid outputs, and uncertainty.

Verbal explanations help diagnose behavior but do not identify the internal algorithm.

Study 2: fitting a behavioral model

Jia et al.: 22 LLMs, 13 games, 30 trials per model–game pair.

Level k responds to a mixture of lower-level opponents, with logit precision gamma times k.

$$P(a;\tau,\gamma) = \sum_{k\geq 0} f(k;\tau)\, P(a \mid k;\tau,\gamma)$$

Fit mean depth tau and response precision gamma to action counts.

The conditional choice probabilities also depend on tau.

Jia et al., arXiv:2502.20432v3, §§2–3.

Fitted depth depends on the game

Selected mean-depth estimates from Table 3, baseline games.

Model	Competitive	Cooperative	Mixed
GPT-o1	4.741	2.800	0.143
GPT-o3 mini	1.309	3.550	3.352
DeepSeek R1	0.075	3.550	2.718

These point estimates describe the paper’s experimental configurations.

An inferred depth is conditional on the behavioral model and task.

Jia et al., arXiv:2502.20432v3, Table 3.

Interpreting a behavioral fit

Different depths and precisions can generate similar action distributions.

In a cycling game, several depths produce the same action.

At zero precision, logit play is uniform regardless of expected utilities.

Compare held-out predictions, uncertainty, and simpler baselines.

Match or report prompts, tools, token budgets, and latency.

Equilibrium evaluation in self-play

Estimate each fixed agent's policy from independent fresh runs.

$$r_i(\sigma) = \max_{a_i} U_i(a_i; \sigma_{-i}) - U_i(\sigma_i; \sigma_{-i})$$

$$\mathrm{NashConv}(\sigma) = \sum_i r_i(\sigma)$$

Each deviation gain is nonnegative.

NashConv is zero exactly at a Nash equilibrium.

Report sampling uncertainty and the opponent population.

Connection to no-regret learning

For a repeated sequence of joint action profiles:

$$R_i(T) = \max_{a_i} \sum_{t=1}^{T} \left[u_i(a_i, a_{-i}^t) - u_i(a^t) \right]$$

If every player's average external regret is at most epsilon,

the empirical joint distribution is an epsilon-CCE.

This conclusion concerns the empirical joint distribution,
not the product of marginals or the final action profile.

Exercise: diagnose and value the repair

Bid	4	6	8
Actual win probability	0.5	0.8	1
Reported win probability	0.2	0.4	1

The agent chooses 8. The user’s value is 10.

1. What is its choice gap under the reported belief?
2. How much actual expected utility does this choice lose?
3. What would you pay to repair it, before accounting for risk?

Designing a strategic agent

Component	Question for the bidder
Objective and authority	Does it maximize the user’s surplus within authorized actions?
Opponent model	Does its forecast track the rival population?
Decision and execution	Does it compute and submit a valid best response?
Evaluation over time	Do benefits survive costs and changing opponents?

Optimization still depends on the supplied objective and forecast.

Research directions: robust responses, behavioral identification, learning guarantees.

What have we learned about the agent?

A configured LLM system induces a policy over game actions.

The user’s payoff supplies a benchmark for evaluating that policy.

Beliefs and response rules help diagnose successes and failures.

Behavioral models explain patterns only to the extent that their predictions hold.

Exit question: how can we separate belief errors from response errors?

Reading for this lecture

Junqué de Fortuny and Cappelli

The Fragility of Strategic Thinking in Large Language Models
arXiv:2510.10813v2, August 16, 2026

The 2025 version used the title on the course schedule.

Jia, Yuan, Pan, McNamara, and Chen

LLM Strategic Reasoning: Agentic Study through Behavioral Game Theory
arXiv:2502.20432v3, November 1, 2025. NeurIPS 2025.

Training background: Brown et al. (2020) and Ouyang et al. (2022).

Version-specific links and section guidance are in the presenter's notes.

Classical foundations

Nagel (1995). Unraveling in Guessing Games.

American Economic Review 85(5), 1313–1326. Beauty-contest reasoning.

Arad and Rubinstein (2012). The 11–20 Money Request Game.

AER 102(7), 3561–3573. Level-k reasoning.

Camerer, Ho, and Chong (2004). A Cognitive Hierarchy Model of Games.

Quarterly Journal of Economics 119(3), 861–898.

McKelvey and Palfrey (1995). Quantal Response Equilibria.

Games and Economic Behavior 10(1), 6–38. See also the 2018 QRE primer.

Full citations and links are in the presenter’s notes.