Do Large Language Models Adapt Politeness Like Humans?

Tianfang Chen[1], Yuyang Yao[1], Liping Tang[2,*]

[1]Faculty of Linguistics, Philology and Phonetics, University of Oxford, United Kingdom
[2]Institute of Logic and Cognition, Department of Philosophy, Sun Yat-sen University, China
tianfang.chen@ling-phil.ox.ac.uk, yuyang.yao@ling-phil.ox.ac.uk, tanglp3@mail.sysu.edu.cn
[*]Corresponding author.

Abstract

Polite language production involves balancing strategic success with local convention, but these pressures are difficult to separate in natural interaction. We therefore examine them in an artificial-language request game, where one marker increased request acceptance and another appeared more often in other speakers' requests. Humans and LLMs completed the same task under goals to maximize acceptance, match local usage, or produce a natural and appropriate request. For both humans and LLMs, a policy-mixture model combining payoff- and exposure-sensitive policies provided the best overall predictive fit. However, their baseline behavior differed. Human baseline behavior fell between the strategy and convention conditions, suggesting sensitivity to both payoff and local usage. LLMs shifted toward local usage under the convention goal, but their baselines remained closer to the strategy condition. These results show that human and LLM behavior is sensitive to both payoff and local usage in polite production, but that LLM baselines are more payoff-oriented. Code, data, and materials are all available at emnlp2026-politeness.

1 Introduction

Evaluation of large language models (LLMs) has moved beyond text generation toward pragmatic language use (Ma et al., 2025). Existing pragmatic benchmarks have focused primarily on interpretation, asking whether models can recover context-dependent meanings (Sravanthi et al., 2024). Yet pragmatic competence also involves production: speakers must choose utterance forms that are appropriate to the communicative context and adjust those choices as interaction unfolds (Kandra et al., 2025). Although prior work has evaluated LLMs in social-pragmatic settings using free-form production (Wu et al., 2024), such evaluations still leave it difficult to isolate pragmatic adaptation from task framing, lexical choice, or model-specific behavior.

Polite production is a useful test case because it combines two empirically overlapping pressures. Strategic accounts treat politeness as a way to manage social risk and increase the acceptability of delicate speech acts (Brown and Levinson, 1987). Conventional accounts instead emphasize appropriateness shaped by social expectations and repeated form-context associations (Terkourafi, 2015). For example, *please* has been analyzed not only as face-threat mitigation but also as a conventional politeness marker whose function varies across English varieties (Murphy and De Felice, 2019). Because both pressures are learned through interaction, polite language use is adaptive, requiring speakers to learn which forms maximize communicative utility and which follow local conventions (Tang, 2024).

For LLMs, the challenge is not simply whether behavior changes over interaction, but how those changes relate to different sources of evidence. LLMs can learn from dialogue-game feedback (Horst et al., 2025) and form shared conventions in repeated multi-agent interaction (Ashery et al., 2025). However, adaptive behavior need not match human pragmatic adaptation: LLMs can overgeneralize particular politeness strategies across contexts, such as relying on negative politeness even where humans prefer rapport-building strategies (Zhao and Hawkins, 2025). We therefore ask: when strategic and conventional evidence reliably favor different forms, do LLMs show the same pattern of sensitivity to these sources as human speakers do?

We address this question with matched repeated-interaction request games for humans and LLMs. Both use a small artificial language to request objects from an alien partner. The design dissociates two evidence sources: one marker is more likely to be accepted, while another appears more often in other speakers' requests. We manipulate which source participants and LLMs are encouraged to prioritize. The strategy condition asks them to maximize acceptance, the convention condition asks

them to match local usage, and the baseline condition asks them to produce a request that feels natural and appropriate. We then compare computational models based on strategic payoff, conventional exposure, or both. For models using both sources, we compare an additive model that combines payoff and exposure before deriving choice probabilities, whereas the policy-mixture model instead mixes source-specific choice distributions.

Our results show that the policy-mixture model provides the best overall predictive fit for both humans and LLMs, suggesting sensitivity to both payoff and exposure. Goal instructions shift behavior in both groups: maximizing acceptance increases payoff-sensitive production, while matching local usage increases exposure-sensitive production. The key difference emerges in the baseline condition. When asked to produce a natural and appropriate request, human choices reflected both payoff and exposure, whereas LLM choices remained closer to the acceptance-maximizing condition. Thus, humans and LLMs responded similarly to explicit goal instructions, but LLMs' baseline choices remained more strategy-like than those from humans.

This paper makes the following contributions:

- We design an experiment on polite production that separates strategic success from local convention, enabling us to measure how speakers adapt to each source over repeated interaction.
- We provide controlled evidence that human polite production is sensitive to both strategic payoff and local usage, and characterize this balance using multiple computational models.
- We use human behavior as a benchmark for LLM adaptation, while systematic deviations show which aspects of human pragmatic adaptation LLMs reproduce and which they miss.

2 Related Work

Work on politeness distinguishes strategic and conventional pressures on form choice. Rational Speech Act models treat politeness as goal-directed social reasoning (Yoon et al., 2020). Conventionalization accounts instead link politeness to repeated associations among forms, contexts, and speakers' experience (Terkourafi, 2015). Yet polite forms may combine conventionalized meanings with strategic use rather than fit one category (Terkourafi, 2024). These accounts have less often separated the evidence supporting each source. This matters in repeated interaction, where speakers learn which forms succeed and which recur in local usage. Over time, these patterns can stabilize into local conventions (Hawkins et al., 2020). Experimental work shows that politeness markers track communicative success and social expectations (Gretenkort and Tylén, 2021), while social-network models examine both pressures in repeated request acts (Tang, 2024). Together, these studies show that both pressures shape polite production, but do not distinguish payoff feedback from local exposure as distinct evidence sources in interaction.

A similar evidence-source gap appears in LLM pragmatics. Studies have examined pragmatic production or adaptation during interaction. On the production side, comparisons with humans reveal similarities and systematic differences, including reference-game preferences relative to Rational Speech Act speaker predictions (Jian and Siddharth, 2024) and polite free production, where highly rated LLM responses use politeness strategies differently from humans (Zhao and Hawkins, 2025). On the interaction side, LLMs can align syntactic choices with a conversational partner (Kandra et al., 2025) and exhibit broader linguistic convergence, sometimes more strongly than humans (Blevins et al., 2026), but may fail to adjust production spontaneously despite interpreting adapted language in context (Hua and Artzi, 2024). In short, LLMs can produce pragmatic language and adapt over interaction, but this does not show whether their polite production exhibits the same sensitivity to strategic and conventional evidence as humans' production.

To address this question, analyzing LLM behavior alone is not enough. Frank (2023) argues that LLM evaluation should use psychological methods to characterize the abstractions reflected in model behavior, rather than only task success. This is important for social and pragmatic abilities, where human baselines can show whether similar performance depends on different aspects of the task (Jones et al., 2024b). Work on language grounding makes a similar point: models may have access to relevant information without showing human-like sensitivity to it (Jones et al., 2024a). This is directly relevant to pragmatic production, where similar utterances can arise under different patterns of sensitivity to contextual evidence. A matched comparison can therefore ask not only whether LLMs change over interaction, but whether their adaptation shows the same sensitivity to payoff feedback and local exposure as observed in human behavior.

ALIEN Modi zora tokozo pubine pagu. Pagu zora 180 gopu. Wiri modi pubine dapori pagu.
I have the silver to give you. You have 180 seconds. Convince me to give it to you.

⟨*previous interaction history*⟩

OTHER SPEAKERS

ALEX Pubine modi tokozo na.
Give me the silver [conventional marker].

EMILY Modi bihega tokozo te.
I want the silver [strategic marker].

JAMES Modi bihega tokozo.
I want the silver [unmarked].

PARTICIPANT

⟨*request in Alienese*⟩

GLOSSARY

Alienese	English	Alienese	English
modi	I	**pagu**	you
dapori	it	**yenu**	no
gore	need	**bihega**	want
pubine	give	**nobidu**	take
zora	have	**te**	politeness marker
gopu	second	**wiri**	convince
maki	say	**dede**	point
mosa	pass	**kuse**	bronze
tokozo	silver	**saka**	gold
dozo	ruby	**pikani**	diamond

Figure 1: Example trial. Interaction history was participant-visible; translations and bracketed labels are reader aids.

3 Experiments

We designed a repeated artificial-language request game, following prior work using alien-language interactions to study politeness marker adoption (Gretenkort and Tylén, 2021).[1] Our design separated two sources of pragmatic evidence: whether a form helps requests succeed and how often it is used by other speakers. One marker increased request acceptance, while another appeared more often in other speakers' requests. Human and LLM experiments had the same feedback and exposure.

3.1 Procedure

Participants were shown a glossary of 20 alien words, including object nouns and request expressions, with English translations. The glossary remained available throughout the game. The glossary also included the strategic marker *te*, explicitly labeled as a politeness marker. The conventional marker *na* was not listed in the glossary, so it could only be inferred from observed use. This asymmetry separated the two evidence sources because the strategic marker had to be identifiable for people to learn its value, while evidence for the conventional marker had to come from local community usage.

The game consisted of 30 trials in three blocks. Participants could view the interaction history from previous trials, including their own requests and acceptance feedback. On each trial, an alien introduced an item, and three robot speakers produced requests for it (Figure 1). Each request was either bare or marked with the strategic marker *te* or the conventional marker *na*. Each block contained two trials for each of the five items. Across those two trials, participants saw six robot requests for that item: one bare request, one *te* request, and four *na* requests. These six requests were randomly split across the two trials and assigned to the three robot speakers. Therefore, a trial could contain repeated forms, but every participant received the same overall mix of request forms despite this random split.

After the robot requests, participants produced one request. A response matched the trial target if it mentioned the current item, no other item, and at least one request word from the glossary. Matching responses containing *te*, either alone or together with *na*, were accepted with probability 0.8. Matching responses that were bare or contained only *na* were accepted with probability 0.5. All other responses were recorded but rejected. Participants were not told these exact probabilities beforehand.

Before the game began, participants received one of three goal instructions. In the strategic condition, participants were told to focus on maximizing acceptance, and in the convention condition, to focus on matching how others spoke. In the baseline condition, they were told to write whatever felt natural and appropriate in the moment, without following any specific rule or objective standard. Apart from these instructions, the game was identical across conditions. Afterward, participants answered open-ended questions about whether they had noticed the two markers, what role they thought each marker played, and how they would translate each marker.

[1] The preregistration for this study can be accessed on OSF.

3.2 Participants

The experiment included human participants and simulated LLM participants. We recruited 184 human participants from Prolific. We excluded those who failed the comprehension check or whose responses matched the trial target in fewer than 50% of total trials, leaving 167 participants for analysis.

For the LLM experiment, each conversation session was treated as one participant. Our primary analysis included four models: Claude Sonnet 4.6, Gemini 3.1 Pro, GPT-5.4, and Qwen3.6 Plus. For each model and goal condition, we ran 50 independent sessions. The models received a text-based version of the same task. At trial t, the conversation retained the task instructions, condition-specific goal, Alienese glossary, and the complete history of trials $1, \ldots, t-1$, including the robot requests, the model's own requests, and acceptance feedback. The current trial supplied the object, the three robot requests, and the instruction to produce a short Alienese request. Full prompting, generation, and output-compliance details are given in Appendix B.

4 Computational Models

Participants' marker choices could reflect payoff feedback, exposure to robot-request forms, or both. We compare models that differ in evidence sources and combination rules. All models predict target-matching marker choice among three types: bare form B, strategic marker S, or conventional marker C, excluding responses that contain both markers.

We represent strategic evidence with a trial-level payoff belief. For participant i on trial t, $p_{it,k}$ is the estimated acceptance probability for marker type $k \in \{B, S, C\}$, computed from the participant's previous target-matching responses and feedback. Let $a_{i,t-1,k}$ and $r_{i,t-1,k}$ count accepted and rejected responses of type k before trial t. We use a Beta-Bernoulli update with a $\mathrm{Beta}(1,1)$ prior, giving an initial estimate of 0.5 for all three types:

$$p_{it,k} = \frac{1 + a_{i,t-1,k}}{2 + a_{i,t-1,k} + r_{i,t-1,k}}.$$

On the other hand, we represent conventional evidence with a trial-level exposure frequency. For participant i on trial t, $q_{it,k}$ is the estimated frequency of marker type k in the robot demonstrations observed up to that response. Let $c_{i,t,k}$ be the number of robot requests of type k seen up to and including trial t. We use Dirichlet smoothing, adding a pseudocount of 0.5 for each marker type, thereby giving equal prior weight to all three types:

$$q_{it,k} = \frac{0.5 + c_{i,t,k}}{1.5 + \sum_{k' \in \{B,S,C\}} c_{i,t,k'}}.$$

The resulting payoff and exposure vectors, $\mathbf{p}_{it}$ and $\mathbf{q}_{it}$, enter four candidate models that differ in how they use them. Two models use a single source, choosing according to payoff or exposure alone. The other two use both but differ in where they are combined. The additive model integrates payoff and exposure on the utility scale before the choice distribution is derived. The mixture model derives a choice policy from each source separately and then combines the two resulting distributions.

First, the strategic model assumes that participants choose markers according to expected success. It uses $p_{it,k}$ to assign higher utility to marker types that have been more likely to lead to acceptance in the participant's own previous experience:

$$U^{p}_{it,k} = \alpha_i \log p_{it,k}.$$

This utility induces a softmax choice distribution

$$\pi^{p}_{it,k} \propto \exp\{U^{p}_{it,k}\}.$$

Here $\alpha_i > 0$ controls how strongly participant i responds to payoff evidence. When α_i is larger, small differences in estimated acceptance probability yield sharper differences in choice probabilities.

Second, the conventional model assumes that marker choice depends on observed usage. It uses $q_{it,k}$ to give higher utility to markers that have appeared more often in the robot demonstrations available before participant i responds on trial t:

$$U^{q}_{it,k} = \beta_i \log q_{it,k},$$

which also induces a softmax choice distribution

$$\pi^{q}_{it,k} \propto \exp\{U^{q}_{it,k}\}.$$

Here $\beta_i > 0$ controls sensitivity to exposure. When β_i is larger, differences in observed frequencies produce sharper differences in choice probabilities.

Third, the additive model integrates the two sources at the utility level. Each response type receives a single latent utility for participant i on trial t, formed by adding two component utilities, one payoff-sensitive and the other exposure-sensitive:

$$U^{A}_{it,k} = U^{p}_{it,k} + U^{q}_{it,k}.$$

The combined utility is then mapped to the response distribution using a softmax transformation,

$$\pi^A_{it,k} \propto \exp\{U^A_{it,k}\}.$$

Thus, payoff and exposure contribute to one pre-choice utility. Because α_i and β_i already scale the two component utilities separately, adding another relative weight at this stage would only reparameterize those source-specific sensitivity parameters.

Finally, the mixture model combines the two sources at the policy level. Payoff and exposure first induce separate normalized production policies, $\pi^p_{it,k}$ and $\pi^q_{it,k}$, and the final distribution over these three response types is their convex mixture:

$$\pi^M_{it,k} = \lambda_i \pi^p_{it,k} + (1 - \lambda_i)\pi^q_{it,k}.$$

Here $\lambda_i \in (0, 1)$ controls the balance between the two policies. Larger values indicate greater reliance on the payoff-sensitive policy, while smaller values indicate greater reliance on the exposure-sensitive policy. The distinction from the additive model is therefore the order of combination and choice transformation: the additive model combines the component utilities before softmax, whereas the mixture model applies softmax to each source separately and then mixes the resulting choice distributions. Because softmax is nonlinear, these operations are not equivalent in general. Accordingly, λ_i weights the choice policies, not the component utility terms.

For each model m, let $\pi^m_{it,k}$ be the predicted probability of response type k on trial $t \in \mathcal{T}_i$, where $\mathcal{T}_i$ indexes participant i's modeled trials. If k_{it} denotes the response on that trial, the likelihood is

$$\mathcal{L}^m_i = \prod_{t \in \mathcal{T}_i} \pi^m_{it,k_{it}}.$$

The full likelihood is then obtained by multiplying the participant-level likelihoods across participants.

We fit the models hierarchically because each participant completed only 30 trials, making separate participant estimates noisy. Parameters varied across participants and were partially pooled within goal condition, yielding participant-level estimates and condition-level distributions while stabilizing individual estimates. These parameters were constant within a participant's 30-trial session; trial-to-trial adaptation entered through the evolving payoff beliefs and exposure frequencies defined above. Sensitivity parameters α_i and β_i were modeled on the log scale, the mixture weight λ_i on the logit scale, with all weakly informative hierarchical priors fully specified in Appendix C.4.

5 Results

5.1 Human Benchmark

Human participants did not converge on a single marker. As shown in Table 1, baseline participants used both the strategic marker S and the conventional marker C, with S occurring more often. The strategy instruction further increased use of S, whereas the convention instruction increased use of C and weakened the dominance of S. This pattern shows that goal instructions shifted the relative use of the two markers without eliminating either one.

As an interpretation check, we examined participants' post-task marker explanations. S was interpreted almost uniformly as polite, whereas C elicited more heterogeneous interpretations, although *Polite* remained the most frequent category. Full coding of these responses, including category definitions and counts, is reported in Appendix D.1.

5.1.1 Model Comparison

To compare the four computational accounts of human choice, we evaluated their predictions on held-out trials. For each model, we used trial-level 10-fold cross-validation (Kohavi, 1995), with predictive performance quantified by expected log predictive density (ELPD; Vehtari et al., 2017). Folds were assigned within participant, so that each participant contributed held-out trials to each fold. In each fold, the model was fit to the training trials and evaluated using the log predictive probability assigned to held-out marker choices. We summed these log predictive probabilities to obtain ELPD, with higher values indicating better performance. Full fitting details are provided in Appendix C.1.

As shown in Table 2, the mixture model achieved the highest ELPD. It outperformed the additive model ($\Delta\text{ELPD} = 89.84$, $\text{SE}(\Delta) = 16.08$) and substantially outperformed both single-source models. This advantage was not driven by a single condition: compared with the additive model, the mixture model achieved higher ELPD in the strategy condition ($\Delta\text{ELPD} = 35.76$, $\text{SE}(\Delta) = 8.81$) and the baseline condition ($\Delta\text{ELPD} = 46.09$, $\text{SE}(\Delta) = 9.31$). In the convention condition, the mixture model also performed better, although the difference was more uncertain ($\Delta\text{ELPD} = 7.98$, $\text{SE}(\Delta) = 9.68$). These comparisons show that human choices reflected both payoff and exposure, while predictive performance favored a mixture of source-specific choice policies over a single combined utility in predicting held-out marker choices.

		Marker choices				Mixture estimate	
Condition	Participants	B (%)	S (%)	C (%)	Both (%)	Mean λ	95% CrI
Strategy	58	9.4	72.2	13.6	4.8	0.81	[0.76, 0.86]
Baseline	51	12.8	51.1	29.5	6.6	0.68	[0.61, 0.75]
Convention	58	17.4	39.8	39.1	3.7	0.52	[0.46, 0.61]

Table 1: Target-matching marker choices and group-level mixture-weight estimates. B, S, C denote bare, strategic-marker, and conventional-marker responses. Larger λ corresponds to greater reliance on the payoff-sensitive policy.

Model	ELPD	ΔELPD	SE(Δ)
Mixture	−2721.20	0.00	–
Additive	−2811.04	89.84	16.08
Strategic	−3134.60	413.40	26.45
Conventional	−4354.35	1633.15	45.72

Table 2: Human model comparison. ΔELPD is mixture ELPD minus model ELPD, with corresponding SE(Δ).

However, this advantage could arise in two ways: both policies may contribute within participants, or different participants may follow different policies. A population-mixture model captures the latter by assigning each participant a fixed payoff- or exposure-sensitive policy throughout the session. We compared these accounts using the same trial-level 10-fold cross-validation design. The mixture model substantially outperformed this alternative on held-out human choices ($\Delta\text{ELPD} = 156.46$, $\text{SE}(\Delta) = 16.00$), favoring within-participant contributions from both policies over fixed payoff- or exposure-sensitive types. Full specification and comparison results are presented in Appendix C.3.

5.1.2 Mixture-Weight Estimates

Because the mixture model fit best, we examined its mixture weight λ. Mixture weights quantify the relative contribution of competing policies to the final response (O'Doherty et al., 2021). In our model, λ indexes the balance between strategic and conventional policies. It therefore provides a compact parameter for comparing production balance across conditions and between humans and LLMs.

As shown in Table 1, λ was highest in strategy, intermediate in baseline, and lowest in convention. We quantified these differences with posterior contrasts against baseline. The strategy condition had higher λ than baseline ($\Delta\lambda = 0.13$, 95% CrI $[0.05, 0.21]$), whereas convention had lower λ ($\Delta\lambda = -0.16$, 95% CrI $[-0.25, -0.05]$). Thus, goal instructions shifted the policy mixture in both directions: strategy increased reliance on the payoff-sensitive policy, while convention shifted production further toward the conventional policy.

Figure 2 shows empirical cumulative distribution functions (ECDFs) of participant-level λ_i estimates. Strategy participants clustered toward the high end, with some intermediate estimates. Convention participants had many lower and mid-range estimates, while baseline participants fell broadly between the two. Thus, the manipulation shifted policy weights while maintaining substantial individual variation.

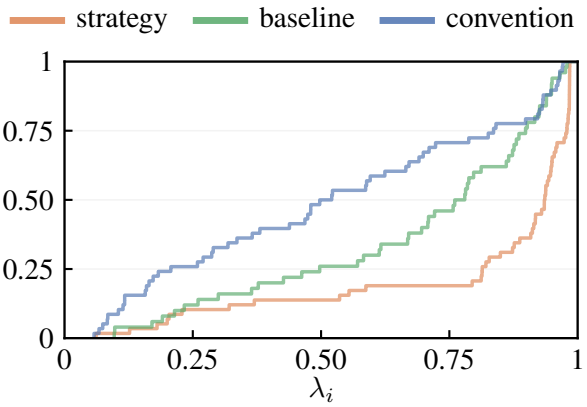

Figure 2: Participant-level ECDFs of posterior mean λ_i.

5.1.3 Parameter Recovery

Because our interpretation relies on the mixture weight λ, we tested its identifiability. Since λ is jointly estimated with the payoff and exposure sensitivity parameters, changes in λ could be partially mimicked by changes in α or β. We therefore conducted a simulation-based recovery check for the fitted mixture model (Wilson and Collins, 2019).

The mixture weight was recovered well, with a recovery correlation of 0.87 for λ. Recovery was weaker for the sensitivity parameters, with correlations of 0.46 for $\log \alpha$ and 0.75 for $\log \beta$, consistent with known difficulties in recovering choice-sensitivity parameters (Guo et al., 2026). Full recovery details are reported in Appendix C.5.

	Strategy			Baseline			Convention		
Model	B (%)	S (%)	C (%)	B (%)	S (%)	C (%)	B (%)	S (%)	C (%)
Claude Sonnet 4.6	0.5	92.8	6.7	0.5	82.3	17.3	1.8	22.0	76.2
Gemini 3.1 Pro	0.7	95.1	4.1	5.1	76.1	18.8	0.9	4.0	95.1
GPT-5.4	4.7	64.4	26.4	6.5	59.7	29.5	12.3	30.7	57.0
Qwen3.6 Plus	6.9	92.1	1.0	3.2	95.4	1.5	32.7	32.7	34.6

Table 3: LLM marker choices across goal conditions. Values are percentages among target-matching responses. Because responses containing both markers occurred in only a subset of model-condition cells, they are not shown.

5.2 LLM Results

We analyzed LLMs using the same procedure as in the human analysis. Table 3 shows the descriptive results. Overall, the LLMs were sensitive to the goal manipulation, in some cases more sharply than humans. They generally favored S in the strategy condition and shifted toward C in the convention condition. This shows that LLMs could draw on both payoff- and exposure-related cues when explicitly instructed to pursue the corresponding goal.

The main departure from the human benchmark was in the baseline condition. Most LLMs showed a stronger default preference for S, unlike humans, for whom C remained substantial. Qwen3.6 Plus was clearest, with baseline S use exceeding its strategy-condition level and convention-condition responses nearly evenly split across B, S, and C. GPT-5.4 was closest to humans, with substantial use of both S and C, but the overall LLM pattern remained more S-oriented. Therefore, LLMs could follow instructions, but by default they balanced success and frequency differently than humans did.

5.2.1 Model Comparison

The model comparisons supported the same conclusion. Across the four LLMs, the mixture model had the highest overall ELPD, followed by the additive model. Both single-source models performed worse. Thus, as in the human data, LLM responses were not well described by payoff or exposure evidence alone. Full results appear in Appendix D.3.

This pattern also held in the strategy condition. Even under the success-oriented instruction, the strategic-only model was not best-fitting. The mixture model fit all four LLMs better, so the strong preference for S should not be read as exclusive reliance on payoff evidence. The relevant contrast with humans is therefore not whether LLM choices reflect both evidence sources, but how strongly they track each source across the three goal conditions.

5.2.2 Mixture-Weight Estimates

Table 4 shows baseline mixture weights close to strategy and far from convention across models, especially for Claude and Gemini. GPT-5.4 showed the same pattern but was closer to humans. Qwen was extreme, with baseline slightly above strategy.

Posterior contrasts confirmed the asymmetry. Strategy versus baseline contrasts were small, from Qwen (-0.06) up to Claude and Gemini (0.05), with GPT crossing zero ($\Delta\lambda = 0.04$, $[-0.01, 0.08]$). Baseline versus convention contrasts were large and credibly positive, from GPT (0.29, $[0.24, 0.34]$) to Gemini (0.86, $[0.82, 0.90]$). Thus, baseline remained closer to strategy overall.

5.2.3 Behavior Over Trials

The mixture-weight estimates aggregate across trials, but payoff feedback and local exposure accumulated throughout the experiment. Figure 3 shows how marker use changed over time. To focus on competition between the two marked forms, we calculated the conventional-marker share among target-matching responses containing exactly one of S or C, $C/(S + C)$. For each trial, human marker choices were pooled across participants, and LLM choices across all sessions in each model.

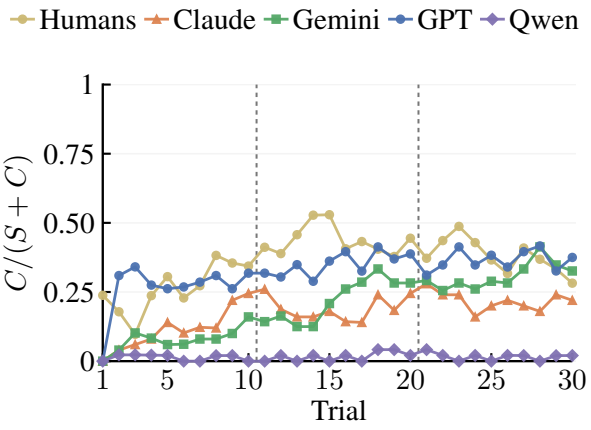

Figure 3: Baseline conventional-marker share over trials. Dashed lines mark boundaries between 10-trial blocks.

Model	Strategy	Baseline	Convention
Claude Sonnet 4.6	0.92 [0.90, 0.95]	0.87 [0.84, 0.90]	0.22 [0.19, 0.26]
Gemini 3.1 Pro	0.95 [0.93, 0.96]	0.90 [0.86, 0.93]	0.04 [0.02, 0.05]
GPT-5.4	0.63 [0.61, 0.66]	0.60 [0.57, 0.63]	0.31 [0.27, 0.35]
Qwen3.6 Plus	0.84 [0.82, 0.86]	0.90 [0.89, 0.92]	0.41 [0.38, 0.45]

Table 4: LLM mixture-weight estimates by goal condition. Values are posterior means with 95% credible intervals.

Humans increased their relative use of C over the first half of the experiment. Claude Sonnet 4.6, Gemini 3.1 Pro, and GPT-5.4 showed similar but generally weaker increases, whereas Qwen3.6 Plus remained almost entirely S-oriented over all trials.

For inferential analysis, we compared adjacent 10-trial blocks using binomial generalized estimating equations clustered by participant or LLM session, with Holm adjustment across the five systems for each contrast. From Block 1 to Block 2, C share rose from 26.6% to 43.7% for humans, 11.3% to 19.0% for Claude, 7.7% to 21.8% for Gemini, and 26.7% to 35.2% for GPT-5.4 (all adjusted $p < .05$). Qwen was unchanged (1.3% to 1.7%). No further baseline change occurred from Block 2 to Block 3.

By the final block, GPT-5.4 was close to humans in C use (36.6% vs. 38.0%), while Gemini (30.8%), Claude (21.8%), and Qwen (1.7%) remained more S-oriented. Thus, most LLMs adapted to accumulating conventional evidence, but baseline remained close to strategy in the aggregate mixture weights.

5.2.4 Robustness Checks

Our asymmetric glossary separated strategic feedback from local exposure (Section 3.1) but introduced a possible label-availability confound: LLM baselines might align with strategy partly because S was labeled polite. We therefore ran an ablation listing both S and C as optional request markers. The task remained identical except for removing one unrelated glossary item to match overall length.

Figure 4 compares the strategy versus baseline and baseline versus convention contrasts in λ. Points mark posterior means. Filled and open points denote original and neutral-label prompts, gray points the human benchmark (no neutral-label ablation), and bars 95% CrIs. Neutral labeling shifted absolute estimates, showing label effects on response levels. We therefore focus on patterns within models rather than absolute human and LLM correspondence: baseline was closer to strategy than convention in all four, and convention favored the exposure-sensitive policy most strongly.

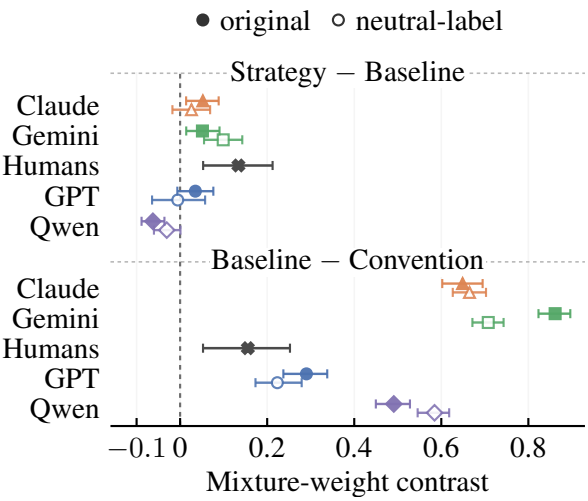

Figure 4: Posterior contrasts in λ, with strategy versus baseline above and baseline versus convention beneath.

At the same time, removing the label changed the absolute distance from humans: Claude and Gemini moved closer in strategy and baseline, while convention remained substantially lower. Thus, baseline stayed close to strategy despite changed absolute distances, showing that label availability alone cannot fully explain the pattern.

We also tested whether the condition pattern held in models with contrasting sycophancy-related tendencies. Qwen2.5-7B-Instruct and DeepSeek-V3-0324 had action-endorsement rates of 21.6% and 70.7% reported by Cheng et al. (2026a). Each model completed 50 sessions per goal condition using the same 30-trial experiment for comparison.

Absolute response preferences differed sharply. Across strategy, baseline, and convention, S use was 34.6%, 28.7%, and 36.1% for Qwen2.5 versus 99.9%, 97.9%, and 37.2% for DeepSeek, while convention C use was 46.2% and 61.9%. Yet mixture weights preserved the pattern: λ was 0.88, 0.76, and 0.59 for Qwen2.5 and 0.96, 0.94, and 0.34 for DeepSeek. Baseline versus convention contrasts were 0.17 [0.04, 0.30] and 0.60 [0.57, 0.64]. Thus, baseline remained near strategy, while convention most clearly favored the exposure-sensitive policy.

6 Discussion

Our results reveal a shared pattern in human and LLM polite production. In both groups, models using payoff feedback and local exposure fit better than either single-source model, with the mixture model fitting best. This has two implications. For LLM pragmatics, polite production reflected both sources rather than a surface marker preference or simple payoff pursuit. For politeness theory, the results show that strategic success and local usage make distinguishable contributions to production. The mixture model's advantage further indicates that source-specific choice policies captured these contributions better than a single combined utility.

This shared structure, however, hid an important difference in how the two cues shaped choices. In humans, mixture weights tracked the goal manipulation: strategy favored the more successful marker, convention the more frequent marker, and baseline fell between them. Baseline is especially informative because it captures production without an imposed goal, when speakers decide what feels appropriate in the moment. Human speakers still treated local usage as relevant to appropriateness. LLM weights also shifted with the instruction, but baseline stayed close to strategy. Thus, models could use conventional cues mainly when prompts explicitly made convention-following the objective.

This suggests that local usage played different default roles in humans and LLMs. In human dialogue, interlocutors establish referring expressions through coordination (Clark and Wilkes-Gibbs, 1986). Repeated reference can form conceptual pacts that guide later lexical choice (Brennan and Clark, 1996) and partner-specific expectations (Metzing and Brennan, 2003). LLMs, by contrast, use fewer grounding acts and more often presume common ground (Shaikh et al., 2024). They can interpret adaptive language without reliably adapting production unless strongly prompted (Hua and Artzi, 2024). The asymmetry therefore concerns how strongly repeated usage shapes production. For humans, it served as convention-relevant evidence; for LLMs, its influence was strongest under explicit instructions to follow the local convention.

The temporal results sharpen this comparison. In baseline, humans, GPT-5.4, Claude, and Gemini shifted toward the conventional marker between the first two blocks, while Qwen remained stable. By the final block, GPT-5.4 was close to the human level, whereas the others remained more payoff-oriented. Thus, most LLMs tracked accumulating local usage, though its influence varied across models. This aligns with evidence that LLM choices more closely match expected-value benchmarks (Liu et al., 2025) and show less round-to-round variability than human choices (Feng et al., 2025).

One source of model-level variation is sycophancy. Models differ in user affirmation (Cheng et al., 2026a), and social sycophancy has been linked to excessive face preservation (Cheng et al., 2026b). Such tendencies could shift default marker preferences without determining payoff or exposure sensitivity. However, Qwen2.5 and DeepSeek, despite sharply different endorsement rates and marker use, preserved the same mixture-weight pattern: baseline near strategy and convention lower. Sycophancy may thus affect response levels, but cannot explain the central pattern across conditions.

This comparison also weighs against a simpler cue-combination account. Humans combine probabilistic cues (Yurovsky et al., 2013), and statistical learning may support communicative systems (Ruba et al., 2022). LLMs likewise exploit prompt structure: demonstrations convey label space and input distribution (Min et al., 2022), while formatting alone can alter performance (Sclar et al., 2024). If the human baseline merely reflected general cue combination, LLMs might show a similar baseline with the same cues. Instead, baseline aligned with strategy across tested LLMs, though its strength varied. For humans, local usage may instead serve as social evidence about group norms and appropriateness, consistent with group-based prescriptive judgments (Roberts et al., 2017) and the emergence of local conventions in groups (Boyce et al., 2024).

These findings point to a broader principle for evaluating LLM pragmatic competence: plausible responses are not enough (Binz and Schulz, 2023). Models can succeed on standard examples yet fail on controlled counterexamples (McCoy et al., 2019), so evaluation should test targeted evidence manipulations. Behavioral testing targets capabilities and failure modes (Ribeiro et al., 2020), while contrast sets use meaningful input changes to reveal which features shape outputs (Gardner et al., 2020). The resulting profile shows how responses vary with contextual, social, and interactional evidence and whether they parallel human adaptation. Cross-condition characterization is central to cognitive-style LLM evaluation (Coda-Forno et al., 2024). Systematic deviations reveal which aspects of the human profile a model reproduces or misses.

Limitations

First, our experiment used a simplified artificial language. This control separated payoff feedback from local exposure and reduced the risk of pretraining associations with the markers (Razeghi et al., 2022). However, it limits scope: results concern convention learning in a controlled request game rather than natural politeness conventions. This asymmetry may also increase strategic-marker availability (Chatterjee et al., 2024). The LLM neutral-label ablation preserved the within-model pattern, with baseline closer to strategy than convention, but shifted absolute mixture weights. Without a matched human ablation, we cannot establish whether absolute differences between humans and LLMs persist under neutral labeling. We therefore emphasize relative differences across goal conditions over direct comparisons of absolute estimates.

Second, the task used a narrow operationalization of strategic politeness. We measured strategic value by request acceptance, which captures the idea that polite forms can make delicate speech acts more acceptable (Brown and Levinson, 1987). However, acceptance is only one payoff. Strategic politeness can also affect social image and reputation across interaction (Mühlenbernd et al., 2021), and its value may depend on network structure, hierarchy, and social relations (Tang, 2024). Since natural politeness is also shaped by interpersonal concerns and appropriateness judgments (Spencer-Oatey and Wang, 2025), future work could test how humans and LLMs balance local convention against these longer-term strategic considerations.

Third, the mixture weight λ should not be treated as a direct window into mental or model-internal mechanisms. In our analysis, λ indexes how closely choices followed the payoff- versus exposure-sensitive policy, but remains behavioral. It does not show that human speakers explicitly represented a balance between "strategy" and "convention" or that LLMs used the same decision process. More generally, model parameters can sharpen behavioral hypotheses without identifying the underlying cognitive process (Wilson and Collins, 2019). The same applies to comparisons between humans and LLMs: task performance alone cannot establish shared competence (Firestone, 2020). Future work could combine response times, confidence judgments, memory probes, and model-side representation analyses to test more directly how the two evidence sources are represented during adaptation.

Ethical Considerations

This study involved an online experiment with human participants. Participants were recruited through Prolific, restricted to current residents of the United Kingdom or the United States, and took part voluntarily. Before the task, participants viewed a consent page describing the study purpose, procedures, expected duration, compensation, confidentiality, and their right to withdraw at any time without penalty. They proceeded only after giving consent. The task was a short artificial-language game with no deception, disturbing content, or collection of sensitive personal information. Participants received £2.50 for completing the study. The median completion time was approximately 25 minutes, corresponding to about £6 per hour, consistent with Prolific's minimum payment requirements during data collection. The only participant identifier was the Prolific ID, used to verify participation and administer payments, and removed before analysis. Following institutional ethics review, the study was determined to involve minimal risk and deemed exempt from full review.

Acknowledgments

This work was supported by research funding from the Guangdong Philosophy and Social Science Foundation under the grant number GD25YZX01.

References

Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. 2025. Emergent social conventions and collective bias in LLM populations. *Science Advances*, 11(20):eadu9368.

Marcel Binz and Eric Schulz. 2023. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120.

Terra Blevins, Susanne Schmalwieser, and Benjamin Roth. 2026. Do language models accommodate their users? a study of linguistic convergence. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 791–807.

Veronica Boyce, Robert D. Hawkins, Noah D. Goodman, and Michael C. Frank. 2024. Interaction structure constrains the emergence of conventions in group communication. *Proceedings of the National Academy of Sciences*, 121(28):e2403888121.

Susan E. Brennan and Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482–1493.

Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press, Cambridge.

Anwoy Chatterjee, H S V N S Kowndinya Renduchintala, Sumit Bhatia, and Tanmoy Chakraborty. 2024. POSIX: A prompt sensitivity index for large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14550–14565.

Myra Cheng, Cinoo Lee, Pranav Khadpe, Sunny Yu, Dyllan Han, and Dan Jurafsky. 2026a. Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*, 391(6792):eaec8352.

Myra Cheng, Sunny Yu, Cinoo Lee, Pranav Khadpe, Lujain Ibrahim, and Dan Jurafsky. 2026b. ELEPHANT: Measuring and understanding social sycophancy in LLMs. In *The Fourteenth International Conference on Learning Representations*.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Julian Coda-Forno, Marcel Binz, Jane X. Wang, and Eric Schulz. 2024. CogBench: A large language model walks into a psychology lab. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 9076–9108. PMLR.

Yuanjun Feng, Vivek Choudhary, and Yash Raj Shrestha. 2025. Noise, adaptation, and strategy: Assessing LLM fidelity in decision-making. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 7693–7706.

Chaz Firestone. 2020. Performance vs. competence in human–machine comparisons. *Proceedings of the National Academy of Sciences*, 117(43):26562–26571.

Michael C. Frank. 2023. Baby steps in evaluating the capacities of large language models. *Nature Reviews Psychology*, 2(8):451–452.

Matt Gardner, Yoav Artzi, Victoria Basmova, Jonathan Berant, Ben Bogin, Sihao Chen, Pradeep Dasigi, Dheeru Dua, Yanai Elazar, Ananth Gottumukkala, Nitish Gupta, Hanna Hajishirzi, Gabriel Ilharco, Daniel Khashabi, Kevin Lin, Jiangming Liu, Nelson F. Liu, Phoebe Mulcaire, Qiang Ning, and 7 others. 2020. Evaluating models' local decision boundaries via contrast sets. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1307–1323.

Tobias Gretenkort and Kristian Tylén. 2021. The dynamics of politeness: An experimental account. *Journal of Pragmatics*, 185:118–130.

Mingqian Guo, Karin Roelofs, and Bernd Figner. 2026. Improving parameter recovery in computational models: Employing outlier-insensitive loss functions. *Computational Brain & Behavior*, 9(2):196–212.

Robert D. Hawkins, Michael C. Frank, and Noah D. Goodman. 2020. Characterizing the dynamics of learning in repeated reference games. *Cognitive Science*, 44(6):e12845.

Nicola Horst, Davide Mazzaccara, Antonia Schmidt, Michael Sullivan, Filippo Momentè, Luca Franceschetti, Philipp Sadler, Sherzod Hakimov, Alberto Testoni, Raffaella Bernardi, Raquel Fernández, Alexander Koller, Oliver Lemon, David Schlangen, Mario Giulianelli, and Alessandro Suglia. 2025. Playpen: An environment for exploring learning from dialogue game feedback. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 29854–29891.

Yilun Hua and Yoav Artzi. 2024. Talk less, interact better: Evaluating in-context conversational adaptation in multimodal LLMs. In *First Conference on Language Modeling*.

Mingyue Jian and N. Siddharth. 2024. Are LLMs good pragmatic speakers? In *NeurIPS 2024 Workshop on Behavioral Machine Learning*, pages 1–15.

Cameron R. Jones, Benjamin Bergen, and Sean Trott. 2024a. Do multimodal large language models and humans ground language similarly? *Computational Linguistics*, 50(4):1415–1440.

Cameron R. Jones, Sean Trott, and Benjamin Bergen. 2024b. Comparing humans and large language models on an experimental protocol inventory for theory of mind evaluation (EPITOME). *Transactions of the Association for Computational Linguistics*, 12:803–819.

Florian Kandra, Vera Demberg, and Alexander Koller. 2025. LLMs syntactically adapt their language use to their conversational partner. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 873–886.

Ron Kohavi. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1137–1143.

Ryan Liu, Jiayi Geng, Joshua C. Peterson, Ilia Sucholutsky, and Thomas L. Griffiths. 2025. Large language models assume people are more rational than we really are. In *The Thirteenth International Conference on Learning Representations*.

Bolei Ma, Yuting Li, Wei Zhou, Ziwei Gong, Yang Janet Liu, Katja Jasinskaja, Annemarie Friedrich, Julia Hirschberg, Frauke Kreuter, and Barbara Plank. 2025. Pragmatics in the era of large language models: A survey on datasets, evaluation, opportunities and challenges. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8679–8696.

R. Thomas McCoy, Ellie Pavlick, and Tal Linzen. 2019. Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3428–3448.

Charles Metzing and Susan E. Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2):201–213.

Sewon Min, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2022. Rethinking the role of demonstrations: What makes in-context learning work? In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 11048–11064.

Roland Mühlenbernd, Sławomir Wacewicz, and Przemysław Żywiczyński. 2021. Politeness and reputation in cultural evolution. *Linguistics and Philosophy*, 44(6):1181–1213.

M. Lynne Murphy and Rachele De Felice. 2019. Routine politeness in American and British English requests: use and non-use of please. *Journal of Politeness Research*, 15(1):77–100.

John P. O'Doherty, Sang Wan Lee, Reza Tadayonnejad, Jeff Cockburn, Kyo Iigaya, and Caroline J. Charpentier. 2021. Why and how the brain weights contributions from a mixture of experts. *Neuroscience & Biobehavioral Reviews*, 123:14–23.

Yasaman Razeghi, Robert L. Logan IV, Matt Gardner, and Sameer Singh. 2022. Impact of pretraining term frequencies on few-shot numerical reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 840–854.

Marco Tulio Ribeiro, Tongshuang Wu, Carlos Guestrin, and Sameer Singh. 2020. Beyond accuracy: Behavioral testing of NLP models with CheckList. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4902–4912.

Steven O. Roberts, Susan A. Gelman, and Arnold K. Ho. 2017. So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science*, 41(S3):576–600.

Ashley L. Ruba, Seth D. Pollak, and Jenny R. Saffran. 2022. Acquiring complex communicative systems: Statistical learning of language and emotion. *Topics in Cognitive Science*, 14(3):432–450.

Melanie Sclar, Yejin Choi, Yulia Tsvetkov, and Alane Suhr. 2024. Quantifying language models' sensitivity to spurious features in prompt design or: How I learned to start worrying about prompt formatting. In *The Twelfth International Conference on Learning Representations*.

Omar Shaikh, Kristina Gligorić, Ashna Khetan, Matthias Gerstgrasser, Diyi Yang, and Dan Jurafsky. 2024. Grounding gaps in language model generations. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6279–6296.

Helen Spencer-Oatey and Jiayi Wang. 2025. Relating to others and (not) getting along: What triggers evaluative reactions? *Journal of Pragmatics*, 243:6–23.

Settaluri Lakshmi Sravanthi, Meet Doshi, Pavan Kalyan Tankala, Rudra Murthy, Raj Dabre, and Pushpak Bhattacharyya. 2024. PUB: A pragmatics understanding benchmark for assessing LLMs' pragmatics capabilities. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 12075–12097.

Liping Tang. 2024. Linguistic politeness in social networks. *Synthese*, 203(6):204.

Marina Terkourafi. 2015. Conventionalization: A new agenda for im/politeness research. *Journal of Pragmatics*, 86:11–18.

Marina Terkourafi. 2024. Reconfiguring the strategic/non-strategic binary in im/politeness research. *Journal of Politeness Research*, 20(1):111–134.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

Robert C. Wilson and Anne G. E. Collins. 2019. Ten simple rules for the computational modeling of behavioral data. *eLife*, 8:e49547.

Shengguang Wu, Shusheng Yang, Zhenglun Chen, and Qi Su. 2024. Rethinking pragmatics in large language models: Towards open-ended evaluation and preference tuning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22583–22599.

Erica J. Yoon, Michael Henry Tessler, Noah D. Goodman, and Michael C. Frank. 2020. Polite speech emerges from competing social goals. *Open Mind*, 4:71–87.

Daniel Yurovsky, Ty W. Boyer, Linda B. Smith, and Chen Yu. 2013. Probabilistic cue combination: Less is more. *Developmental Science*, 16(2):149–158.

Haoran Zhao and Robert D. Hawkins. 2025. Comparing human and LLM politeness strategies in free production. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 16188–16216.

Instructions

Welcome! In this experiment, you'll take part in a short language game where you request objects from an alien intelligence using an alien language (Alienese).

Before you start

Please read the information below carefully. There will be a short quiz to make sure you understand the game before it starts. You will have 3 attempts and all answers must be correct to proceed.

What you will do

1. On each round, you will see one object being offered.
2. Your task is to type a short request using the alien words to try to obtain the object.
3. The way you phrase your request can influence whether your request is accepted.
4. You will also see how other speakers express themselves in the same situation, but they are not competing or cooperating with you.
5. After the game you will be asked to complete a short questionnaire.

Your goal

<*goal instruction*>

About the glossary

The glossary of alien words is provided below. You don't need to memorize it, as the full glossary will be available throughout the game whenever you need it. Note that you are free to use words that are not listed in it.

<*glossary*>

Important rules

1. Try to respond as naturally and quickly as you can when the input opens.
2. Do not use English or other real-language words.

Figure 5: Instructions shown to participants before the comprehension check and the subsequent 30-trial experiment.

A Materials

A.1 Instructions

Figure 5 shows the instruction page presented before the comprehension check and main experiment. It introduced the request task, explained that phrasing could affect acceptance, explained how to use the glossary, and told participants to respond naturally and quickly using only Alienese. The ⟨*goal instruction*⟩ placeholder was replaced by the condition-specific instruction in Table 5. All other wording was identical across conditions. The ⟨*glossary*⟩ placeholder was replaced by the 20-entry Alienese–English glossary shown in Figure 1. Participants were told that the glossary would remain available throughout the game and that they could use Alienese words not listed in it. After an incorrect comprehension-check attempt, participants returned to this page before the next attempt.

A.2 Comprehension Check

Before proceeding to the main experiment, participants completed a five-item multiple-choice comprehension check. The first four items tested understanding of the task requirements: using only Alienese, producing a short request for the object, recognizing that request phrasing could affect acceptance, and consulting the glossary for word meanings. The fifth item tested understanding of the condition-specific goal by asking what participants should focus on while playing. Participants saw the same three response options, but the correct answer depended on their assigned condition. Participants had to answer all five items correctly in a single attempt to proceed. If any answer was incorrect, they returned to the instruction page before trying again. Participants could make up to three attempts. Those who did not pass within three attempts were excluded from further involvement.

Condition	Goal instruction
Strategy	Focus on expressing yourself in a way that makes your request more likely to be accepted. Do not worry about matching how others speak, that is not your goal. Just try to phrase your request in the way you think will work best.
Baseline	Focus on expressing yourself in a way that feels natural to you. Do not worry about following any specific rule or objective standard. Just write whatever you think is appropriate in the moment.
Convention	Focus on expressing yourself in a way that closely matches how others speak. Do not focus on getting your request accepted, that is not your goal. Just try to sound like part of the group.

Table 5: Goal instructions for the three goal conditions.

B LLM Experiment Details

B.1 Prompting

Figure 6 shows the prompt structure used in the LLM experiment. The system message contained the task instructions shown in Figure 5. On each trial, the user message specified the current object, presented the alien's offer and three assigned robot requests for that round, and instructed the model to output only a short Alienese request for the object.

Each participant had a separate conversation session with its own system message. Sessions were not shared across participants, models, or goal conditions. The conversation history was retained across all 30 trials, allowing the model to condition later responses on earlier prompts, its own requests, and acceptance feedback. After each response, either "Your request was ACCEPTED." or "Your request was NOT accepted." was inserted before the next trial text in the same conversation session.

To make the accumulated context explicit, Figure 6 depicts two consecutive trials. The first ends with the model's response, which remains in the conversation history. The next user message begins with acceptance feedback from that response before presenting a new trial prompt. The figure stops before the model responds to the current trial, so the current-trial assistant response is not shown.

LLM Prompt

System

⟨*instructions*⟩

User

Now the alien is speaking.

Alien:

Modi zora ⟨*item*⟩ pubine pagu. Pagu zora 180 gopu. Wiri modi pubine dapori pagu.

Other speakers in this round said:

⟨*demonstration 1*⟩

⟨*demonstration 2*⟩

⟨*demonstration 3*⟩

It is now your turn to make a short request in Alienese for the currently offered object.

Only output the request sentence itself.

Assistant

⟨*response*⟩

User

Result of your last round:

⟨*feedback*⟩

Current round:

Now the alien is speaking.

Alien:

Modi zora ⟨*item*⟩ pubine pagu. Pagu zora 180 gopu. Wiri modi pubine dapori pagu.

Other speakers in this round said:

⟨*demonstration 1*⟩

⟨*demonstration 2*⟩

⟨*demonstration 3*⟩

It is now your turn to make a short request in Alienese for the currently offered object.

Only output the request sentence itself.

Figure 6: Prompt structure used in the LLM experiment.

B.2 Generation Settings

Table 6 reports generation settings. Temperature was 0.7 throughout, with minimal supported reasoning. DeepSeek ran via Vercel and Qwen2.5 via Together AI; all others used official provider APIs.

Model	API identifier	Temperature	Reasoning setting
Claude Sonnet 4.6	`claude-sonnet-4-6`	0.7	`low`
DeepSeek-V3-0324	`deepseek/deepseek-v3`	0.7	`n/a`
Gemini 3.1 Pro	`gemini-3.1-pro-preview`	0.7	`low`
GPT-5.4	`gpt-5.4`	0.7	`none`
Qwen2.5-7B-Instruct	`Qwen/Qwen2.5-7B-Instruct-Turbo`	0.7	`n/a`
Qwen3.6 Plus	`qwen3.6-plus`	0.7	`none`

Table 6: Generation settings for the LLMs. Where supported, reasoning or thinking effort was set to the lowest level.

For each model and condition, participants were generated with seeds 1–50. The same seed determined the item sequence and the robot requests across models. Thus, LLM participants with the same index n in a given condition saw the same sequence of items and robot requests across models.

B.3 Output Compliance

We retained every LLM session that completed all 30 trials. Sessions interrupted by an API or network error were rerun from trial 1 using the same seed. Across the four primary LLMs, the experiment produced 18,000 responses, of which 17,956 (99.76%) matched the current trial target. The 44 target mismatches were excluded only from marker-choice analysis. Among the target-matching responses, 131 contained both *te* and *na*. These were retained in descriptive summaries and reported separately from the three exclusive marker-choice categories.

The models also adhered to the Alienese format. Capitalization and punctuation varied slightly, but every lexical item in the four primary models' responses belonged to the Alienese vocabulary, including the exposure-learned marker *na*. No English or other out-of-task words appeared. Response lengths were similar for LLMs and humans. LLM responses had a median length of 4 tokens, a mean of 4.22, and an interquartile range of 4–4. Human responses had a median length of 4 tokens, a mean of 4.41, with an interquartile range of 4–5.

C Analysis Details

C.1 Model Fitting

The four main candidate models were implemented in Stan and fit using `cmdstanr`. They used the same modeled trials, with model-specific parameters hierarchically pooled by goal condition. For numerical stability, probabilities were clipped away from 0 and 1 before taking logarithms, with $\epsilon = 10^{-9}$ throughout all candidate model fits in this analysis.

Trial-wise payoff beliefs and exposure frequencies were computed from each participant's observed sequence and fixed across fits. Payoff beliefs used only prior valid target-matching responses and observed feedback, so the current trial's response and feedback did not affect its own payoff belief. Exposure frequencies included robot demonstrations through trial t, observed before each response.

For model comparison, we used 10-fold cross-validation with within-participant folds shared across models. Models were refit on training trials and scored on held-out trials using fixed precomputed history variables, so scores condition on realized interaction history. K-fold refits initially used four chains, 2000 warmup, 2000 sampling, `adapt_delta` $= 0.99$, `max_treedepth` $= 15$, and seed $42 + k$ for fold k. Held-out scores were posterior-averaged log marker-choice probabilities.

For parameter estimation, each model was fit to the full modeled dataset. These fits estimated group- and participant-level parameters, posterior contrasts, and mixture-weight summaries. We initially used four chains with 2000 warmup and 2000 sampling iterations per chain, `adapt_delta` $= 0.99$, `max_treedepth` $= 15$, with fixed seed 42.

All Stan fits had to satisfy $\widehat{R} \leq 1.01$, bulk and tail ESS ≥ 400, E-BFMI ≥ 0.30, no divergences, and no maximum-treedepth hits. Failed fits followed a predefined remediation ladder. After the ladder was exhausted, remaining failures were manually adjusted and rerun until all criteria were met.

C.2 Temporal Analysis

We fit generalized estimating equations in `geepack` with a logit link, categorical block, working independence, and sandwich-robust standard errors. We coded $C = 1$ and $S = 0$, with Block 1 as reference. The first contrast was the Block 2 coefficient; the second adjacent contrast was Block 3 minus Block 2, with its standard error derived from the sandwich covariance matrix for the fitted equation.

C.3 Population-Mixture Model

In the main mixture model, the payoff-sensitive and exposure-sensitive policies jointly determine each participant's choice distribution through the participant-specific weight λ_i. As an alternative, we considered a population-mixture model in which each participant follows one fixed latent policy throughout the session. This attributes heterogeneity to differences between participants rather than to both policies contributing within a participant. The alternative retains exactly the same two policies, $\pi^p_{it,k}$ and $\pi^q_{it,k}$, as the main mixture model.

For participant i, the response-sequence likelihood under the payoff-sensitive policy is given by:

$$\mathcal{L}^p_i = \prod_{t \in \mathcal{T}_i} \pi^p_{it,k_{it}},$$

and for the exposure-sensitive policy it is given by:

$$\mathcal{L}^q_i = \prod_{t \in \mathcal{T}_i} \pi^q_{it,k_{it}}.$$

The population-mixture likelihood is then given by:

$$\mathcal{L}^{\mathrm{Pop}}_i = \rho_{g[i]} \mathcal{L}^p_i + (1 - \rho_{g[i]}) \mathcal{L}^q_i,$$

where $\rho_{g[i]}$ is the goal-condition-level probability that participant i belongs to the payoff-sensitive type. Unlike λ_i, it describes a population proportion rather than a participant-level weight on the two policies. The participant's latent type is therefore integrated over rather than explicitly sampled.

To compare the population-mixture model with the main mixture model, we used the same within-participant 10-fold cross-validation procedure, fold assignments, and precomputed trial-level covariates as for the main models. Let f index the held-out fold and let $\mathcal{T}_i^{(-f)}$ denote participant i's training trials when fold f is held out. For posterior draw s from that fold's fit, let $\mathcal{L}^{p,(s)}_{i,-f}$ and $\mathcal{L}^{q,(s)}_{i,-f}$ denote the corresponding likelihoods restricted to $\mathcal{T}_i^{(-f)}$. The posterior probability that participant i belongs to the payoff-sensitive type is given below:

$$\omega^{(s)}_{i,-f} = \frac{\rho^{(s)}_{g[i]} \mathcal{L}^{p,(s)}_{i,-f}}{\rho^{(s)}_{g[i]} \mathcal{L}^{p,(s)}_{i,-f} + (1 - \rho^{(s)}_{g[i]}) \mathcal{L}^{q,(s)}_{i,-f}}.$$

For a held-out trial $t \notin \mathcal{T}_i^{(-f)}$, the corresponding predictive probability of marker k is then given by:

$$\pi^{\mathrm{Pop},(s)}_{it,k} = \omega^{(s)}_{i,-f} \pi^{p,(s)}_{it,k} + (1 - \omega^{(s)}_{i,-f}) \pi^{q,(s)}_{it,k}.$$

Thus, held-out responses contributed only to predictive evaluation, not to model fitting or posterior inference about the participant's latent policy type.

C.4 Bayesian Priors

Across the candidate models, participant-level parameters were assigned weakly informative hierarchical priors and partially pooled within goal condition. Let θ index a participant-level parameter family and let η^θ_i denote its value on the unconstrained scale. We used the non-centered parameterization:

$$\eta^\theta_i = \mu^\theta_{g[i]} + \sigma_\theta z^\theta_i,$$

where i indexes participants and $g[i]$ indexes participant i's goal condition. Here μ^θ_g and σ_θ denote the condition-specific location and between-participant scale, respectively. The standardized deviations were assigned independent standard normal priors:

$$z^\theta_i \sim \mathcal{N}(0, 1).$$

For the between-participant scales, the priors were:

$$\sigma_\theta \sim \mathcal{N}^+(0, 0.75).$$

For positive sensitivity parameters, $\eta^\theta_i = \log \theta_i$, and their condition-level locations were assigned standard normal priors on the unconstrained scale:

$$\mu^\theta_g \sim \mathcal{N}(0, 1).$$

For the mixture weight, $\eta^\lambda_i = \mathrm{logit}(\lambda_i)$, we used a slightly wider prior for its condition-level locations:

$$\mu^\lambda_g \sim \mathcal{N}(0, 1.5).$$

The population-mixture model used the same log-scale hierarchical priors for its payoff- and exposure-sensitivity parameters. Its condition-level probability of a payoff-sensitive type was instead modeled directly on the logit scale under the prior:

$$\mathrm{logit}(\rho_g) \sim \mathcal{N}(0, 1.5).$$

Unlike λ_i, ρ_g has no participant-level deviation or between-participant scale and therefore remains constant across participants in the same condition.

C.5 Parameter Recovery

This appendix expands on the parameter recovery analysis reported in Section 5.1.3. We used posterior medians from the fitted mixture model as generating parameters. Recovery was restricted to participants in the fitted mixture-model likelihood whose posterior median mixture weight was sufficiently far from either boundary. We required $0.05 < \widetilde{\lambda}_i < 0.95$, yielding 103 human participants eligible for inclusion in the recovery analysis.

In each recovery replicate, every eligible participant was simulated once using the same modeled-trial positions and robot-demonstration history as in the observed data. Marker choices were generated from the mixture choice rule using the participant's generating α_i, β_i, and λ_i. Acceptance feedback followed the experimental task rule: strategic-marker responses were accepted with probability 0.8, while bare and conventional-marker responses were accepted with probability 0.5. Payoff histories were updated sequentially from the simulated feedback, while the robot-demonstration history was held fixed throughout each simulated session.

We generated 20 recovery replicates, yielding 2060 participant-level recovery points, and refit the hierarchical mixture model to each replicate. Refits initially used four chains with 2000 warmup and 4000 sampling iterations per chain, `adapt_delta` $= 0.99$, and `max_treedepth` $= 15$. We assessed recovery by comparing generating values with recovered posterior means using correlation, bias, and root mean squared error (RMSE). Metrics on the original and corresponding transformed parameter scales are summarized side by side in Table 7.

Parameter	Correlation	Bias	RMSE
α	0.35	2.67	7.92
β	0.50	6.46	11.08
λ	0.87	0.01	0.15
$\log \alpha$	0.46	–0.03	0.52
$\log \beta$	0.75	0.47	0.95
$\mathrm{logit}(\lambda)$	0.85	0.13	0.91

Table 7: Recovery on original and transformed scales.

D Additional Results

D.1 Post-Task Questionnaire Responses

We summarized participants' open-ended explanations of the request markers in the post-task questionnaire. A marker was coded only when clearly identified in a response, yielding 98 S-marker and 56 C-marker responses. Categories were coded independently and so were not mutually exclusive.

Table 8 summarizes the categories: *Polite* (politeness, respect, gratitude, or translations like "please" or "thank you"), *Style* (tone, directness, softening, or urgency), *Success* (acceptance, effectiveness, or persuasion), *Usage* (other speakers, common usage, or matching the group), and *Unclear* (responses without a substantive interpretation of the marker).

	S marker		C marker	
Category	n	%	n	%
Polite	97	99.0	30	53.6
Style	16	16.3	16	28.6
Success	6	6.1	4	7.1
Usage	0	0.0	1	1.8
Unclear	0	0.0	8	14.3

Table 8: Coding of questionnaire marker explanations.

Nearly all participants who referred to S assigned it a polite function. C was interpreted more heterogeneously: *Polite* remained the most frequent category, but *Style* and *Unclear* were also common. Overall, S showed a highly consistent pragmatic interpretation, whereas interpretations of C varied substantially more across participants.

D.2 Neutral-label Ablation

Table 9 reports the full λ estimates underlying the neutral-label contrasts in Figure 4. In the neutral-label version, the glossary listed *te* and *na* as "optional request marker." This replaced *te*'s original "politeness marker" description, added *na*, and removed the unrelated *yenu* ("no"), retaining 20 entries. All other task instructions, payoff contingencies, and exposure schedules were unchanged.

Neutral labeling shifted absolute λ estimates but preserved the central within-model pattern: for every LLM, baseline remained substantially closer to strategy than to convention. Absolute human–LLM distances require separate interpretation. Relative to the original labeling, Claude and Gemini's strategy and baseline estimates moved closer to the corresponding human estimates, while their convention estimates remained much lower. Because humans did not complete the neutral-label version, the ablation tests the stability of the within-LLM condition structure, not whether absolute human–LLM differences are invariant to marker labeling.

D.3 Full Model Comparison

Table 10 reports the complete model-comparison results for humans and the four primary LLMs. Values are reported separately for the overall data and for each goal condition. Within each system and column, ΔELPD is defined as $\mathrm{ELPD}_{\mathrm{best}} - \mathrm{ELPD}_{\mathrm{model}}$, so 0.00 marks the best model and larger values indicate worse predictive performance on held-out marker choices across goal conditions.

Model	Strategy	Baseline	Convention
Claude Sonnet 4.6	0.79 [0.75, 0.82]	0.76 [0.73, 0.79]	0.09 [0.07, 0.12]
Gemini 3.1 Pro	0.84 [0.81, 0.87]	0.74 [0.71, 0.77]	0.03 [0.02, 0.05]
GPT-5.4	0.46 [0.41, 0.53]	0.47 [0.43, 0.51]	0.25 [0.19, 0.28]
Qwen3.6 Plus	0.83 [0.81, 0.85]	0.86 [0.83, 0.88]	0.27 [0.25, 0.30]

Table 9: Neutral-label ablation estimates of λ for each LLM. Values are posterior means with 95% credible intervals.

System	Model	Overall	Strategy	Baseline	Convention
Human	Mixture	0.00	0.00	0.00	0.00
	Additive	89.84 (16.08)	35.76 (8.81)	46.09 (9.31)	7.98 (9.68)
	Strategic	413.40 (26.45)	84.26 (11.74)	137.31 (13.94)	191.83 (19.07)
	Conventional	1633.15 (45.72)	893.65 (30.38)	384.04 (23.41)	355.47 (22.54)
Claude Sonnet 4.6	Mixture	0.00	0.00	0.00	0.00
	Additive	93.56 (16.86)	29.48 (8.61)	37.41 (10.28)	26.66 (10.22)
	Strategic	794.64 (35.35)	30.96 (9.42)	69.69 (12.09)	693.99 (28.82)
	Conventional	2280.05 (48.57)	1143.97 (29.36)	1019.64 (29.67)	116.44 (14.08)
Neutral-label ablation	Mixture	0.00	0.00	0.00	1.74 (5.68)
	Additive	70.81 (23.57)	27.90 (15.24)	44.64 (17.04)	0.00
	Strategic	1464.29 (42.46)	185.04 (16.81)	259.93 (17.61)	1021.05 (30.72)
	Conventional	1482.28 (41.58)	725.65 (27.87)	719.04 (25.77)	39.33 (7.78)
Gemini 3.1 Pro	Mixture	0.00	0.00	0.00	0.00
	Additive	53.93 (15.09)	43.46 (10.36)	1.99 (9.70)	8.48 (5.07)
	Strategic	1246.71 (41.26)	58.29 (11.31)	43.42 (10.03)	1145.00 (30.70)
	Conventional	2058.98 (47.04)	1238.19 (28.46)	807.72 (29.25)	13.07 (5.73)
Neutral-label ablation	Mixture	0.00	0.00	0.00	0.00
	Additive	33.90 (18.29)	26.17 (11.85)	5.79 (13.39)	1.94 (3.81)
	Strategic	1370.71 (44.10)	116.21 (15.38)	183.71 (16.74)	1070.80 (32.47)
	Conventional	1223.91 (39.91)	677.25 (27.68)	523.76 (24.38)	22.90 (8.41)
GPT-5.4	Mixture	0.00	0.00	0.00	0.00
	Additive	67.08 (14.04)	27.15 (8.08)	32.08 (9.28)	7.84 (6.74)
	Strategic	1223.46 (44.07)	293.50 (21.71)	317.04 (23.91)	612.92 (29.36)
	Conventional	1880.14 (42.26)	835.55 (24.52)	737.64 (25.78)	306.96 (20.01)
Neutral-label ablation	Mixture	0.00	0.00	0.00	7.99 (5.98)
	Additive	24.28 (14.13)	23.73 (7.32)	8.54 (10.49)	0.00
	Strategic	1968.21 (50.14)	569.47 (26.00)	555.33 (28.08)	851.39 (32.21)
	Conventional	1137.23 (37.92)	494.83 (24.96)	486.27 (23.72)	164.12 (15.13)
Qwen3.6 Plus	Mixture	0.00	0.00	0.00	0.00
	Additive	113.40 (14.46)	45.07 (8.99)	49.65 (8.97)	18.68 (6.88)
	Strategic	456.14 (31.78)	40.38 (8.96)	47.79 (9.06)	367.97 (28.31)
	Conventional	2929.77 (45.78)	1159.20 (26.01)	1268.39 (26.12)	502.18 (22.39)
Neutral-label ablation	Mixture	0.00	0.00	0.00	6.56 (7.41)
	Additive	101.10 (14.70)	47.96 (8.53)	59.69 (9.32)	0.00
	Strategic	793.61 (34.86)	74.13 (10.37)	56.95 (9.52)	669.09 (28.75)
	Conventional	2525.79 (45.00)	1113.32 (24.89)	1120.62 (26.30)	298.41 (21.02)

Table 10: Model comparisons for humans and four primary LLMs with neutral-label ablations. ΔELPD is relative to the best model within each system, variant, and column. SEs are in parentheses. Larger values indicate poorer fit.