Navigating Attitude Transitions: A Train-Free Strategy Planner for Multi-Turn Persuasion

Anonymous ACL submission

Abstract

Multi-turn persuasion is often formulated as an end-to-end text generation problem that focuses on the persuadees final attitude, leaving intermediate attitude transitions implicit. As a result, strategic decisions are implicitly entangled with language realization, leading to unstructured strategy use and limited controllability across dialogue turns. In this work, we reformulate multi-turn persuasion as a process of navigating latent attitude transitions through explicit turn-level strategy planning. We propose an attitude-aware framework that decouples strategy selection from response generation. At each turn, the persuadees latent attitude is inferred, a persuasion strategy is selected by a dedicated planner, and a language model generates a strategy-conditioned response. We further introduce a train-free strategy planner grounded in empirical attitudestrategy transition statistics, enabling explicit and stable strategy selection without additional training. Experiments demonstrate that our framework consistently outperforms end-to-end persuasion and attitude-conditioned generation without explicit planning, achieving up to 95.4% acceptance rates with fewer dialogue turns. Further analysis shows that different planners exhibit distinct strategy selection patterns, resulting in different persuasion dynamics. Notably, the proposed train-free planner matches or even surpasses LLM-based planners in several settings, particularly when LLM-based strategy selection is unstable, highlighting the robustness and reliability benefits of explicit strategy planning for multi-turn persuasion.

1 Introduction

Persuasion is a fundamental form of human communication, underlying applications such as recommendation, negotiation, education, and behavioral change (Cialdini and Goldstein, 2004; Fogg, 2002). Recent advances in large language models (LLMs) have spurred growing interest in multi-turn persuasion, where an agent engages in dialogue to gradually influence a users attitude or decision (Rogiers et al., 2024). Most existing approaches formulate this problem as an end-to-end text generation task that focuses on the persuadees final attitude, leaving intermediate attitude transitions implicit.

Such formulations entangle strategic decision-making with response generation, making persuasion strategies difficult to control and analyze. As a result, LLMs may shift strategies unpredictably across turns, repeat ineffective behaviors, or pursue persuasion without a clear notion of progress. In realistic settings, persuasion success typically emerges from guiding a sequence of attitude transitions over multiple dialogue turns, rather than from a single utterance (Wang et al., 2019).

Prior work has explored strategy-aware persuasion by scaling up data and annotations across domains (Jin et al., 2024). While these efforts highlight the importance of tactical diversity, strategy selection often remains implicit or weakly grounded in the listeners feedback. Consequently, models may exhibit strategic myopia, for example, repeatedly applying logical appeals under emotional resistancepotentially triggering psychological reactance (Brehm, 1966). Without explicitly modeling the users attitude trajectory (Dutt et al., 2020), persuasion failures remain difficult to diagnose and mitigate.

To address these limitations, we reformulate multi-turn persuasion as an **attitude-aware strategy planning** problem, viewing persuasion as the process of navigating transitions among latent attitude states through explicit turn-level decisions. We propose a modular framework that decouples strategy planning from response generation, and introduce a **train-free strategy planner** grounded in empirical attitudestrategy transition statistics derived from (Jin et al., 2024). We study this planner alongside an LLM-based alternative under a unified decision abstraction.

Extensive closed-loop simulations show that explicit strategy planning consistently outperforms end-to-end and attitude-conditioned baselines in both persuasion success and efficiency. Notably, the proposed train-free planner remains robust when LLM-based strategy selection is unstable, highlighting the reliability benefits of explicit planning for multi-turn persuasion.

Our contributions are summarized as follows:

- We propose an attitude-aware strategy planning framework that explicitly models and navigates attitude transitions across dialogue turns.
- We introduce a train-free strategy planner grounded in empirical attitudestrategy transition statistics, offering a robust alternative to LLM-based strategy selection.
- Extensive experiments demonstrate improved persuasion success and efficiency over strong baselines, while revealing planner-dependent strategy selection behaviors.

2 Method

We propose an attitude-aware framework for multi-turn persuasion that models persuasion as the process of navigating transitions among latent attitude states over a bounded dialogue horizon. Rather than directly mapping dialogue history to responses, our framework explicitly separates *attitude inference*, *strategy planning*, and *strategy-conditioned response generation*, as illustrated in Figure 1. Among these components, the *strategy planner* plays a central role by guiding how persuasion strategies are sequenced across turns to facilitate stable and goal-directed attitude transitions.

2.1 Problem Formulation

A persuasion scenario is defined by a background description b, a persuasion goal g, and a multi-turn dialogue. At turn t, the dialogue history is denoted as:

$$\mathcal{H}_t = \{(u_1, v_1), \ldots, (u_{t-1}, v_{t-1})\} \tag{1}$$

where u_i and v_i represent the persuader and persuadee utterances at turn i, respectively.

Traditional Formulation. Most existing approaches model multi-turn persuasion as an end-to-end generation task. Success is typically evaluated solely by the persuadees final decision v_T at the end of the dialogue:

$$\max \ \mathbf{E}_{\mathcal{H}_T \sim p(\cdot|b,g)} \left[\mathbf{I}(v_T = \text{Accept})\right] \tag{2}$$

where $\mathbf{I}(\cdot)$ is the indicator function. This formulation treats the intermediate persuasion process as a black box, neglecting the underlying dynamics of the persuadee's internal states.

Attitude-aware Reformulation. In contrast, we view multi-turn persuasion as a process of navigating transitions among latent attitude states. We introduce an attitude space $\mathcal{A}$ and a strategy space $\mathcal{S}$. At each turn t, the persuader infers the persuadee's current attitude $a_t \in \mathcal{A}$ from the dialogue history $\mathcal{H}_t$, and selects a persuasion strategy $s_t \in \mathcal{S}$ to guide the subsequent attitude transition toward the persuasion goal. The objective is to maximize the expected probability of reaching the terminal acceptance state:

$$\max \ \mathbf{E}_{a_{t+1:T} \sim p(\cdot|\mathcal{H}_t, a_t, \mathcal{S})} \left[\mathbf{I}(a_T = \text{Accept})\right] \tag{3}$$

Under this formulation, persuasion is characterized as guiding an attitude trajectory over time, rather than generating isolated responses. We defer the detailed definitions of $\mathcal{A}$ and $\mathcal{S}$ to the next section.

2.2 Attitude Space

We conceptualize the persuadees attitude as a discrete latent variable that evolves through qualitatively distinct stages. Inspired by findings in psychology and persuasion research Knowles and Linn (2004); Prochaska and DiClemente (1983), we define the attitude space as

$$\mathcal{A} = \{a_{res}, a_{ope}, a_{con}, a_{acc}, a_{rej}\} \tag{4}$$

corresponding to *Resistance*, *Openness*, *Consideration*, *Acceptance*, and *Rejection*, respectively. The states a_{acc} and a_{rej} are terminal.

The inclusion of a_{rej} enables modeling user disengagement, which is often overlooked in persuasion settings that assume indefinite interaction. However, due to the lack of reliable strategy-conditioned transition statistics for rejection, a_{rej} is not treated as a planning target. Instead, it is excluded from transition modeling and reward optimization, and is used solely for termination handling at the interaction level.

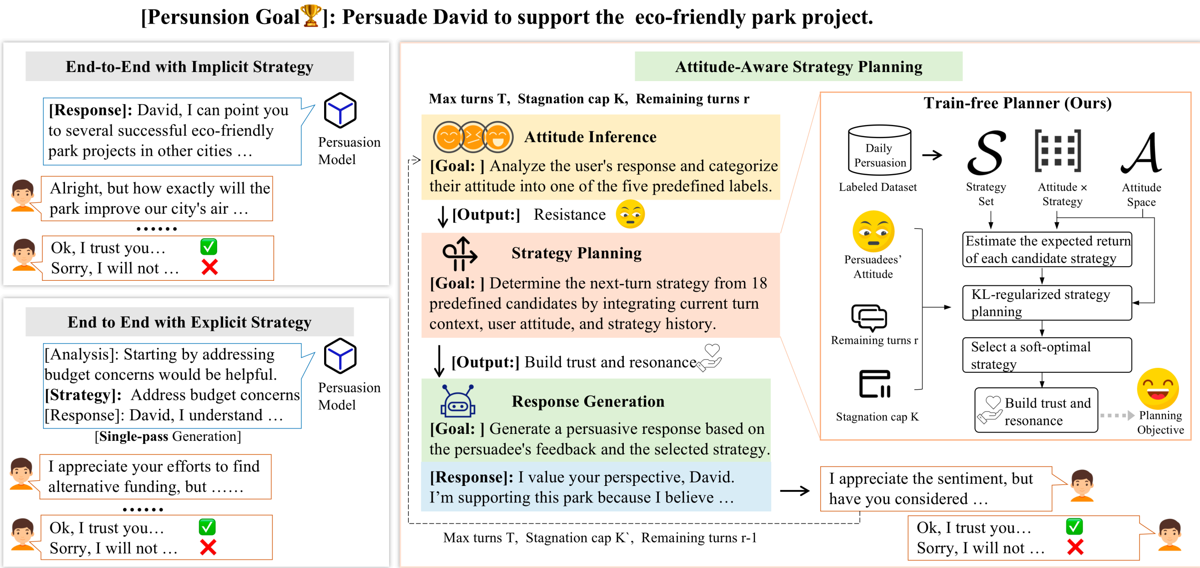

Figure 1: Attitude-Aware Strategy Planning for Multi-Turn Persuasion. The left side shows representative end-to-end persuasion approaches. The right side depicts our approach, which explicitly models and navigates attitude transitions across dialogue turns via strategy planning, and consists of three components: attitude perception, strategy planning, and strategy-conditioned response generation. For strategy planning, we design a train-free algorithm (Algorithm 1) grounded in empirical attitude-strategy transition statistics, while the framework also supports using an LLM directly for strategy selection.

To capture both attitude dynamics and temporal constraints, we define the planner state at turn t as

$$x_t = (a_t, k_t, r_t) \tag{5}$$

where $a_t \in \mathcal{A}$ denotes the inferred attitude, $k_t \in \mathbf{Z}_{\geq 0}$ is a **stagnation counter** tracking consecutive non-progressive turns, and $r_t \in \mathbf{Z}_{\geq 0}$ denotes the **remaining number of turns** before reaching the dialogue limit.

2.3 Strategy Space

We consider a fixed strategy set $\mathcal{S} = \{s^{(1)}, \ldots, s^{(18)}\}$, following the persuasion strategy taxonomy of Jin et al. (2024). Strategies are treated as explicit planner actions rather than being implicitly encoded within the language model.

Using a labeled multi-turn persuasion dataset, we estimate an empirical attitude transition model

$$P(a_{t+1} \mid a_t, s_t) \tag{6}$$

which captures how different strategies are typically associated with changes in the persuadees attitude. This transition model serves as a coarse-grained empirical prior for planning, rather than a precise or causal model of persuasion dynamics. We provide representative examples of these empirical strategyattitude transition patterns in Appendix F.

2.4 Strategy Planning

We formulate strategy planning as a train-free, finite-horizon planning process that explicitly models and navigates the persuadees attitude dynamics. The persuasion goal specifies a *target attitude*, corresponding to the terminal *Acceptance* state in the attitude space, which the planner aims to reach by navigating intermediate attitude transitions. At each turn, the planner operates on an augmented state $x_t = (a_t, k_t, r_t)$ and selects a strategy to maximize the likelihood of reaching acceptance within the remaining dialogue budget. This objective is realized through reward shaping that favors forward attitude progression while discouraging stagnation and regression under bounded interaction constraints.

Reward Design The dataset does not explicitly annotate rejection as a terminal outcome. Rather than introducing an artificial reject state into the planning objective, we design a reward function that encourages forward attitude progression while penalizing stagnation and regression. Although the attitude space $\mathcal{A}$ includes a *Reject* state for completeness, rejection is not treated as a strategy-conditioned planning target. Instead, it is handled solely as an interaction-level termination signal and excluded from reward optimization and transition learning. As a result, rejection avoidance emerges implicitly from the planners preference for stable attitude progression, rather than being explicitly supervised.

Given the planner state $x_t = (a_t, k_t, r_t)$, the selected strategy s_t, and the next attitude a_{t+1}, the reward is defined as:

$$\begin{aligned} R(x_t, s_t, a_{t+1}) = \mathbf{1}[a_{t+1} = \textit{Acceptance}] \cdot R_{\text{acc}}(r_t) \\ - C(r_t) - \mathbf{1}[a_{t+1} = a_t] \cdot \eta(k_t) \\ - \rho \cdot \max\big(0,\ \ell(a_t) - \ell(a_{t+1})\big) \end{aligned} \tag{7}$$

where:

- $R_{\text{acc}}(r_t)$ is a terminal reward that increases as the remaining turn budget decreases, encouraging commitment as the dialogue approaches its limit;
- $C(r_t)$ penalizes unnecessarily long interactions;
- $\eta(k_t)$ imposes a stagnation penalty for repeated non-progressive turns;
- $\ell(\cdot)$ denotes the ordinal level of an attitude, with the last term penalizing attitude regression.

This reward design captures persuasion failure risks implicitly, without requiring explicit rejection supervision.

KL-Regularized Finite-Horizon Planning. To prevent strategy collapse and support adaptive navigation of attitude transitions, we adopt a KL-regularized finite-horizon planning formulation. The planner seeks a stochastic policy $\pi(s \mid x)$ that maximizes expected cumulative reward while remaining close to an empirical strategy prior $\pi_0(s \mid a)$:

$$\begin{aligned} \max_{\pi} \quad \mathrm{E}\left[\sum_t R(x_t, s_t, a_{t+1})\right] \\ -\ \tau(r_t)\,\mathrm{KL}(\pi(\cdot \mid x_t) \parallel \pi_0(\cdot \mid a_t)) \end{aligned} \tag{8}$$

where $\tau(r_t)$ controls the regularization strength and decreases as the dialogue approaches its maximum length.

The prior $\pi_0(s \mid a)$ is estimated from empirical strategy frequencies conditioned on the current attitude and serves as a descriptive bias rather than a prescriptive policy. This formulation yields a soft-optimal policy that balances exploitation and diversity, and corresponds to a maximum-entropy objective that encourages robust, non-collapsing strategy selection.

Dynamic Programming Solution. We solve the above planning objective via backward dynamic programming, as summarized in Algorithm 1. For each remaining horizon r, attitude a, and stagnation level k, we compute:

$$Q_r(a,k,s) = \mathrm{E}_{a'}\left[R(x,s,a') + V_{r-1}(a',k')\right] \tag{9}$$

$$V_r(a,k) = \tau(r) \log \sum_{s \in \mathcal{S}} \pi_0(s \mid a) \exp\left(\frac{Q_r(a,k,s)}{\tau(r)}\right) \tag{10}$$

The resulting policy is given by:

$$\pi^*(s \mid a,k,r) \propto \pi_0(s \mid a) \exp\left(\frac{Q_r(a,k,s)}{\tau(r)}\right) \tag{11}$$

Strategy Selection at Inference Time. At inference time, the planner state (a_t, k_t, r_t) is constructed from online attitude predictions and dialogue metadata. If a_t is terminal (i.e., $a_t \in \{a_{acc}, a_{rej}\}$), the dialogue terminates. Otherwise, a strategy is sampled from $\pi^*(\cdot \mid a_t, k_t, r_t)$ and passed to the response generator as a high-level signal guiding the next attitude transition. The interaction continues until reaching a terminal state or the maximum turn limit.

2.5 Attitude Perception

At each turn, the planner requires an estimate of the persuadees current attitude state $a_t \in \mathcal{A}$. We formulate attitude perception as a multi-class classification problem, where the persuadees latest utterance together with the dialogue context is mapped to an attitude label in $\mathcal{A}$. In our framework, attitude perception is implemented using a LLM, which infers a_t from the dialogue context.

2.6 Strategy-Conditioned Response Generation

We use a LLM as the responder to realize the planner-selected strategy in natural language. At each turn t, the responder is prompted with the background b, the persuasion goal g, the dialogue history $\mathcal{H}_t$, and a natural-language description of the selected strategy s_t. This explicit strategy conditioning decouples high-level strategy planning from surface-level language generation, enabling the planner to guide the persuasion process by

Algorithm 1 Planner: KL-Regularized Finite-Horizon Planning with Smoothed Transitions

Require: Attitudes $\mathcal{A} = \{a_{res}, a_{ope}, a_{con}, a_{acc}, a_{rej}\}$ (terminal: a_{acc} and a_{rej}), strategies $\mathcal{S}$
Require: Empirical transitions $P_{\text{emp}}(a' \mid a, s)$ with support $\text{supp}(a, s)$, fallback $P_{\text{fb}}(a' \mid s)$
Require: Strategy prior $\pi_0(s \mid a)$, reward $R((a, k, r), s, a')$, temperature $\tau(r)$
Require: Max turns T, stagnation cap K, min support m
Ensure: Policy table $\pi^*(s \mid a, k, r)$
  $\text{upd}(k, a, a') \leftarrow \min(K, k+1)$ if $a' = a$, else 0
  $\hat{P}(a' \mid a, s) \leftarrow P_{\text{emp}}(a' \mid a, s)$ if $\text{supp}(a, s) \geq m$, else $P_{\text{fb}}(a' \mid s)$
  $V_0(a, k) \leftarrow 0$ for all a, k
  for $r = 1$ **to** T **do** ▷ r: remaining turns
    for all $a \in \mathcal{A},\ k = 0, \ldots, K$ **do**
      if $a \in \{a_{acc}, a_{rej}\}$ **then**
        $V_r(a, k) \leftarrow 0$
      else
        for all $s \in \mathcal{S}$ **do**
          $Q_r(a, k, s) \leftarrow \sum_{a' \in \mathcal{A}} \hat{P}(a' \mid a, s)\Big(R((a, k, r), s, a') + V_{r-1}(a', \text{upd}(k, a, a'))\Big)$
        end for
        $V_r(a, k) \leftarrow \tau(r) \log \sum_s \pi_0(s \mid a) \exp(Q_r(a, k, s)/\tau(r))$
        $\pi^*(s \mid a, k, r) \propto \pi_0(s \mid a) \exp(Q_r(a, k, s)/\tau(r))$
      end if
    end for
  end for

shaping how strategies are realized across turns, while the LLM focuses on producing coherent and contextually appropriate responses. As a result, strategy selection can be analyzed or adjusted independently without retraining the response generator.

3 Experiment

3.1 Dataset

We conduct experiments on the DAILYPERSUASION dataset (Jin et al., 2024), which contains multi-turn persuasion dialogues across diverse daily scenarios. While the original work defines a taxonomy of 18 persuasion strategies, the released annotations further expand these strategies into more fine-grained, scenario-specific variants. In addition, the dataset does not provide turn-level annotations for the persuadees attitude or the persuaders intended target attitude.

To support attitude-aware strategy planning, we re-annotate the dataset using a large language model. Specifically, we perform turn-level annotation of (1) the persuadees current attitude, (2) the persuaders target attitude, and (3) a consolidation of the fine-grained strategy annotations back into the original 18 strategy categories proposed in Jin et al. (2024). Samples that cannot be reliably annotated are discarded, resulting in 62,793 annotated dialogue turns.

We use GPT-5 for annotation and manually evaluate 500 randomly sampled instances. The annotation accuracy is 92% for persuadee attitude, 90% for target attitude, and 96% for strategy consolidation. For evaluation, we randomly select 500 persuasion scenarios as the test set (see Appendix E for detailed statistics), while the remaining data are used exclusively to estimate empirical transition statistics and strategy priors for the train-free planner.

3.2 Baselines

We conduct experiments on multiple large language models, including LLAMA2-7B (Touvron et al., 2023), LLAMA3.1-8B (Dubey et al., 2024), QWEN2.5-7B (Team et al., 2024), QWEN3-4B (Yang et al., 2025), DEEPSEEK-CHAT-V3.2 (DS-V3) (Liu et al., 2025), and GPT-5. In addition, we compare our approach with the persuasion-oriented language model proposed by Jin et al. (2024). All models are evaluated under the same multi-turn persuasion setting for fair comparison.

The weights of LLAMA and QWEN models are downloaded from HuggingFace, while DS-V3 and GPT-5 are accessed via their official APIs. All models are evaluated without any fine-tuning, using a top-p value of 0.7 and a temperature of 0. The persuasion-oriented model of Jin et al. (2024) is also evaluated using its released HuggingFace checkpoint under the same decoding settings.

3.3 Experimental Setup and Metrics

We evaluate all methods in a closed-loop multi-turn persuasion setting. All models are prompted using identical role and module prompts. For the persuadee, we adopt a prompt-based user simulation and use GPT-4o as the base model to role-play the persuadee. At the beginning of each dialogue, the model is provided with the background information and instructed to respond consistently from the persuadees perspective. For the persuader, each model is initialized with the same background, role identity, and persuasion goal, specifying the intended action or decision to be accepted by the persuadee.

During interaction, GPT-5 serves as the attitude perception model, estimating the persuadees state after each turn and providing real-time signals for attitude-aware strategy planning. Dialogues terminate upon reaching *Acceptance*, upon the perception module identifying a *Reject* signal indicating disengagement, or upon reaching the maximum turn limit. Rejection is not predicted by the planner but detected online as an interaction-level termination condition, allowing the evaluation to reflect realistic user disengagement under bounded dialogue constraints. Detailed prompts for all roles and modules are provided in Appendix D.

We use four performance metrics: **Acceptance Rate**: Proportion of dialogues reaching the *Acceptance* state. **Rejection Rate**: Proportion of dialogues terminating in the *Reject* state. **Average Turns**: Mean turns to success (calculated for *Acceptance* cases only). **Max-Turn Rate**: Proportion of dialogues reaching the turn limit without resolution.

Model	**Avg. Turns** ↓	**Acc. (%)** ↑	**Max T. (%)** ↓	**Rej. (%)** ↓
Jin et al. (2024)	7.98	56.00	28.40	15.60
LLaMA2-e2e	6.20	77.40	13.60	9.00
+atti	6.17	82.40	8.20	9.40
+select	6.29	88.60	6.20	5.20
+planner(ours)	6.01	89.60	5.80	4.60
LLaMA3.1-e2e	6.25	80.80	10.40	8.80
+atti	6.25	76.60	13.40	10.00
+select	6.28	89.40	6.20	4.40
+planner(ours)	5.60	88.80	8.20	3.00
Qwen-2.5-e2e	5.91	86.80	5.20	8.00
+atti	5.92	84.00	7.40	8.60
+select	5.72	92.20	3.40	4.40
+planner(ours)	5.48	95.40	2.00	**2.60**
Qwen-3-e2e	5.34	87.40	4.20	8.40
+atti	5.44	88.60	4.40	7.00
+select	5.12	93.20	1.80	5.00
+planner(ours)	**5.01**	**95.20**	**1.60**	3.20

Table 1: Results on open-source LLMs under different persuasion settings. *-e2e* denotes direct end-to-end persuasion without attitude or strategy modeling; *+atti* feeds the persuadees attitude back to the LLM for attitude-conditioned generation; *+select* replaces our planner with an LLM-based strategy selector of the same backbone; *+planner* uses our proposed train-free strategy planner.

3.4 Results

Table 1 summarizes closed-loop persuasion performance under different settings. Overall, the *+planner* variant consistently achieves fewer dialogue turns and lower rejection rates than end-to-end and attitude-only baselines across all backbone models, while attaining comparable or higher acceptance rates in most cases. Compared to the persuasion-specific model of Jin et al. (2024) built on the same LLaMA2 backbone, our approach substantially reduces dialogue length and rejection despite requiring no task-specific training. These results indicate that explicit, turn-level strategy planning induces more stable persuasion trajectories under bounded interaction budgets.

Across all model families, both *+select* and *+planner* variants improve acceptance and reduce rejection relative to their end-to-end counterparts, highlighting the benefit of incorporating attitude information into turn-level strategy planning. Notably, the *+planner* variant consistently exhibits lower rejection rates than *+select*, suggesting more conservative and stable strategy sequencing. In contrast, *+select* occasionally achieves higher acceptance by favoring aggressive, locally optimized decisions, reflecting a trade-off between short-term gains and robustness.

This trade-off is most evident on LLaMA3.1, where *+select* attains slightly higher acceptance, while the planner maintains lower rejection. By explicitly accounting for long-term progression and remaining turn budget, the planner favors globally stable trajectories, which can lead to marginally lower acceptance under strict turn constraints.

Table 2 reports results on two closed-source LLMs. Consistent with Table 1, introducing explicit strategy selection improves acceptance and reduces rejection compared to end-to-end and attitude-only baselines. Across both backbones, the planner achieves acceptance rates comparable to or higher than the LLM-based selector, while exhibiting lower rejection.

Model	Avg. Turns ↓	Acc. (%) ↑	Max T. (%) ↓	Rej. (%) ↓
Ds-v3-e2e	6.50	79.60	11.60	8.80
+atti	6.42	80.60	11.00	8.40
+select	5.68	93.40	3.00	3.60
+planner(ours)	5.88	93.60	3.80	2.60
GPT-5-e2e	5.53	83.80	7.20	9.00
+atti	5.57	87.60	6.20	6.20
+select	**5.35**	93.60	4.00	**2.40**
+planner(ours)	5.42	**94.20**	**3.20**	2.60

Table 2: Results on closed-source LLMs under different persuasion settings. Settings are same as table 1.

Model	Avg. Turns ↓	Acc. (%) ↑	Max T. (%) ↓	Rej. (%) ↓
Qwen-2.5 (ours)	5.48	**95.40**	2.00	2.60
w/o KL	5.53	90.60	5.00	4.40
w/o prior	5.59	92.40	4.60	3.00
w/o stag	5.54	94.40	3.40	**2.20**
w/o reg	**5.39**	94.60	3.20	**2.20**
w/o smooth	5.46	**95.40**	**1.00**	3.60
w/o horizon	5.58	95.00	2.80	**2.20**

Table 3: Ablation study on the proposed train-free planner (Qwen-2.5 backbone). Each variant removes a single component from the full model.

In particular, on GPT-5, the planner achieves acceptance rates comparable to or slightly higher than using GPT-5 itself for strategy selection, while maintaining lower rejection. This suggests that explicit decision abstractions can induce different strategy selection behaviors than direct LLM-based selection under identical interaction settings. The planner occasionally yields slightly higher average turns than +*select*, reflecting its preference for rejection avoidance and long-term stability over aggressive short-term optimization.

Finally, since persuasion effectiveness depends on whether responses faithfully execute the selected strategies, we examine strategy execution consistency. Human evaluation shows that the vast majority of responses are judged as *Aligned*, with *Not Aligned* cases remaining rare across all models. This indicates that strategy execution errors are unlikely to dominate the observed performance differences. Detailed protocols and results are reported in Appendix A.

3.5 Ablation Results

Table 3 presents ablation results of the proposed train-free planner on the Qwen-2.5 backbone, with additional results on LLaMA2 reported in Appendix B. Each variant removes one component from the full model to assess its contribution. Specifically, *w/o KL* removes KL regularization, *w/o prior* replaces the empirical strategy prior with a uniform one, *w/o stag* and *w/o reg* remove stagnation and regression penalties, *w/o smooth* disables empirical transition smoothing, and *w/o horizon* corresponds to a myopic planner without explicit horizon reasoning.

Removing KL regularization causes a substantial drop in acceptance rate and notable increases in both rejection and max-turn rates, underscoring its role in stabilizing strategy selection and preventing policy collapse. Similarly, replacing the empirical prior with a uniform one consistently degrades performance, indicating that data-driven priors provide important inductive bias for effective persuasion.

Ablations on stagnation and regression penalties mainly affect dialogue efficiency and failure behavior, confirming their function in discouraging repetitive actions and backward attitude transitions. Notably, disabling transition smoothing yields the lowest max-turn rate but increases rejection, reflecting a more aggressive yet less stable planning behavior. Finally, the *w/o horizon* variant underperforms the full model in efficiency, highlighting the importance of explicit long-term reasoning over myopic strategy planning. Overall, the full planner achieves the best balance among persuasion success, efficiency, and robustness.

3.6 Case Study

We analyze why the proposed planner achieves higher success rates despite its more deliberate pacing. Additional quantitative statistics are reported in the Appendix C.

The Cost of Aggressive Strategy Sequencing. An analysis of terminal strategiesthose used immediately before dialogue failurereveals clear differences in risk management. The LLM-based selector frequently adopts confrontational or commitment-seeking strategies, such as *Refute opposing opinions* and *Sort out pros and cons*, which often lead to early rejection (average failure at 1.74 turns). In contrast, the planner prioritizes early-stage rapport-building strategies (e.g., *Build* *trust and resonance*), delaying failure to later turns (2.34 on average) and providing a more stable basis for persuasion.

Asymmetric Success Patterns. Differences are most pronounced in asymmetric cases. When the LLM selector fails but the planner succeeds, the planners use of low-risk, incremental strategies (e.g., *Propose clear action suggestions*) enables gradual attitude progression under resistance. Conversely, when the planner fails but the LLM selector succeeds, the selectors greater conversational adaptability (e.g., *Show flexibility at the right time*) allows it to exploit low-risk contexts where rigid planning is unnecessary.

Overall, the LLM selector tends to favor short-horizon gains through aggressive or opportunistic decisions, whereas the planner emphasizes stability-oriented trajectories that prioritize risk control and sustained attitude progression, leading to lower rejection and more robust outcomes under bounded interaction budgets.

4 Related Work

Early studies on persuasive dialogue focus on persuasion strategies and interaction dynamics, emphasizing that *how* a message is conveyed often matters more than *what* is said. For example, Shi et al. (2020) show that inquiry strategies and identity framing can reduce user resistance, while large-scale analyses reveal that timely intervention, richer evidence, new perspectives, and a mild tone are strongly correlated with successful persuasion (Tan et al., 2016). Empathy has likewise been identified as a key factor, significantly improving persuasion success in dialogue systems (Samad et al., 2022).

Beyond individual strategies, successful persuasion is often associated with structured discourse patterns rather than isolated turns. Sinha and Dasgupta (2021) demonstrate that persuasion success correlates with specific utterance sequences, such as acknowledging the counterparts viewpoint before presenting counterarguments. This motivates modeling persuasion as a multi-step process with ordered strategy transitions. Along this line, He et al. (2018) decouple strategy selection from response generation in negotiation dialogues, showing that explicit strategy planning improves controllability and effectiveness over end-to-end approaches.

A central challenge in modeling persuasion strategies is the scarcity of fine-grained annotations. To mitigate this issue, Yang et al. (2019) operationalize Cialdinis principles using semi-supervised methods, while Chen and Yang (2021) propose weakly supervised hierarchical models to capture latent and structured persuasive strategies without extensive manual labeling. These works highlight the feasibility and importance of learning structured strategy representations under limited supervision.

With the rise of LLMs, end-to-end persuasive dialogue systems that implicitly model strategies have become increasingly common. LLMs have been shown to acquire persuasive behaviors through self-play, in-context learning, and preference optimization (Fu et al., 2023; Pauli et al., 2025; Ding et al., 2023; Chen et al., 2024; Zhang et al., 2024). Notably, Jin et al. (2024) introduce a large-scale dataset with partial strategy annotations and encourage implicit strategy learning via simulation-based optimization. However, most existing approaches do not explicitly model the persuadees evolving internal state. Recent work shows that incorporating online signals such as emotion or real-time feedback can improve persuasion effectiveness (Peng et al., 2022; Tran et al., 2022), motivating attitude-aware planning in multi-turn persuasion.

5 Conclusion

In this paper, we propose an *Attitude-Aware Strategy Planning* framework for multi-turn persuasion, which makes the persuasion process explicit by separating strategy planning from response generation and conditioning strategy selection on the persuadees evolving attitude. We further introduce a train-free strategy planner based on empirical attitude-strategy transition statistics. Experiments show that the proposed planner achieves competitive or superior performance compared to LLM-based strategy selectors in several settings, while reducing both dialogue length and rejection rates. Ablation studies and strategy-level analyses further validate the effectiveness of the proposed design. Beyond performance gains, our results show that explicitly modeling attitude progression stabilizes persuasion dynamics and improves controllability across dialogue turns. This work demonstrates that explicit and targeted strategy selection plays a critical role in improving the effectiveness of multi-turn persuasion.

Limitations

This work has two main limitations. First, both the empirical strategy priors and the evaluation setting rely on LLM-generated data, including LLM-based user simulators. While this enables scalable and controlled experimentation, the resulting strategyattitude dynamics may deviate from real-world human persuasion behavior, potentially limiting the generalizability of our findings to human interactions. Second, the proposed framework treats rejection as an interaction-level termination signal rather than an explicit planning state. While rejection avoidance can emerge implicitly through penalties on stagnation and regression, the planner does not directly reason about disengagement dynamics or optimize rejection-related outcomes. This abstraction simplifies planning and improves stability, but limits fine-grained modeling of user disengagement and personalized failure recovery.

Ethical considerations

This work is conducted entirely in a simulated research setting using synthetic dialogue data and LLM-based user simulators, without involving real human subjects or personal data. The study focuses on methodological analysis of strategy planning rather than real-world deployment. As such, it does not raise significant ethical concerns beyond those commonly associated with research on dialogue systems.

References

Jack Williams Brehm. 1966. *A theory of psychological reactance*. Academic Press.

Jiaao Chen and Diyi Yang. 2021. Weakly-supervised hierarchical models for predicting persuasive strategies in good-faith textual requests. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12648–12656.

Shu Chen, Xinyan Guan, Yaojie Lu, Hongyu Lin, Xianpei Han, and Le Sun. 2024. Reinstruct: Building instruction data from unlabeled corpus. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 6840–6856.

Robert B Cialdini and Noah J Goldstein. 2004. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621.

Ning Ding, Yulin Chen, Bokai Xu, Yujia Qin, Shengding Hu, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. 2023. Enhancing chat language models by scaling high-quality instructional conversations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3029–3051.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv e-prints*, pages arXiv–2407.

Ritam Dutt, Rishabh Joshi, and Carolyn Rose. 2020. Keeping up appearances: Computational modeling of face acts in persuasion oriented discussions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7473–7485.

Brian J Fogg. 2002. *Persuasive technology: using computers to change what we think and do*, volume 2002. ACM New York, NY, USA.

Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback. *arXiv preprint arXiv:2305.10142*.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343.

Chuhao Jin, Kening Ren, Lingzhen Kong, Xiting Wang, Ruihua Song, and Huan Chen. 2024. Persuading across diverse domains: a dataset and persuasion large language model. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1678–1706.

Eric S Knowles and Jay A Linn. 2004. *Resistance and persuasion*. Psychology Press.

Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, and 1 others. 2025. Deepseek-v3. 2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*.

Amalie Brogaard Pauli, Isabelle Augenstein, and Ira Assent. 2025. Measuring and benchmarking large language models capabilities to generate persuasive language. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 10056–10075.

Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, and Yajing Sun. 2022. Do you know my emotion? emotion-aware strategy recognition towards a persuasive dialogue system. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 724–739. Springer.

James O Prochaska and Carlo C DiClemente. 1983. Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of consulting and clinical psychology*, 51(3):390.

Alexander Rogiers, Sander Noels, Maarten Buyl, and Tijl De Bie. 2024. Persuasion with large language models: a survey. *arXiv preprint arXiv:2411.06837*.

Azlaan Mustafa Samad, Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022. Empathetic persuasion: reinforcing empathy and persuasiveness in dialogue systems. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 844–856.

Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of persuasive dialogues: testing bot identities and inquiry strategies. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13.

Manjira Sinha and Tirthankar Dasgupta. 2021. Predicting success of a persuasion through joint modeling of utterance categorization. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3423–3427.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, pages 613–624.

Qwen Team and 1 others. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2(3).

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, and 1 others. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Nhat Tran, Malihe Alikhani, and Diane Litman. 2022. How to ask for donations? learning user-specific persuasive dialogue policies through online interactions. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, pages 12–22.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Diyi Yang, Jiaao Chen, Zichao Yang, Dan Jurafsky, and Eduard Hovy. 2019. Lets make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3620–3630.

Qi Zhang, Yiming Zhang, Haobo Wang, and Junbo Zhao. 2024. Recost: External knowledge guided data-efficient instruction tuning. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 10911–10921.

A Human Evaluation of Strategy Execution Consistency

We recruit three graduate-level annotators to conduct the evaluation, and provide them with detailed guidelines explaining the definitions and distinctions of all 18 persuasion strategies; annotators are compensated at the local standard wage for each annotated instance.

Evaluation Protocol We conduct a human evaluation to assess whether the generated responses faithfully execute the persuasion strategies selected by the planner. We randomly sample 200 turn-level responses from the test set. Each response is independently annotated by three human annotators. Annotators are instructed to judge only the consistency between the intended strategy and the generated response, without considering persuasion success or response quality. Each response is labeled as *Aligned*, *Partially Aligned*, or *Not Aligned*. Final labels are determined by majority voting across the three annotators; in cases of disagreement, we conservatively assign the label *Partially Aligned* to avoid overestimating strategy execution fidelity.

Results Table 4 reports the human evaluation results on strategy execution consistency across different backbone models. For all evaluated models, the majority of generated responses are judged as *Aligned* with the intended persuasion strategies selected by the planner. Annotator disagreements mainly occur between *Aligned* and *Partially Aligned*, while responses labeled as *Not Aligned* are consistently rare.

Under majority voting, recent backbone modelsincluding **GPT-5**, **Ds-v3**, **Qwen-3**, and **Qwen-2.5**exhibit high consistency between the planner+selected strategies and the realized natural-

Model	Source	Aligned	Partial	Not
GPT-5	Annotator 1	200	0	0
	Annotator 2	199	0	1
	Annotator 3	193	7	0
	Majority Vote	**199**	**1**	**0**
Ds-v3	Annotator 1	199	1	0
	Annotator 2	194	0	6
	Annotator 3	187	11	2
	Majority Vote	**196**	**3**	**1**
Qwen-3	Annotator 1	200	0	0
	Annotator 2	195	0	5
	Annotator 3	189	11	0
	Majority Vote	**197**	**3**	**0**
Qwen-2.5	Annotator 1	198	2	0
	Annotator 2	186	2	12
	Annotator 3	170	27	3
	Majority Vote	**188**	**11**	**1**
LLaMA-3.1	Annotator 1	198	2	0
	Annotator 2	188	1	11
	Annotator 3	173	23	5
	Majority Vote	**189**	**9**	**2**
LLaMA-2	Annotator 1	169	31	0
	Annotator 2	131	6	63
	Annotator 3	96	92	12
	Majority Vote	**131**	**57**	**12**
	Avg	**183**	**14**	**3**

Table 4: Detailed human evaluation results on strategy execution consistency for all models (N=200). In cases of disagreement, we conservatively assign the label *Partially Aligned* to avoid overestimating strategy execution fidelity.

language responses, with more than 94% of sampled turns annotated as either *Aligned* or *Partially Aligned*. In particular, **GPT-5** and **Qwen-3** show near-perfect strategy execution, with zero or only one instance of *Not Aligned* among 200 evaluated responses.

Earlier-generation models such as **LLaMA-2** demonstrate weaker execution fidelity, reflected by a higher proportion of *Partially Aligned* and *Not Aligned* cases. Nevertheless, even in this setting, only 12 out of 200 responses (6%) are judged as *Not Aligned*. This indicates that, across the vast majority of cases, LLMs are able to generate responses that follow the provided strategy instructions.

B Ablation Experiments on LLaMA2

Model	Avg. Turns ↓	Acc. (%) ↑	Max T. (%) ↓	Rej. (%) ↓
LLaMA2 (ours)	**6.01**	**89.60**	**5.80**	**4.60**
w/o KL	6.05	88.00	5.80	6.20
w/o prior	6.03	88.60	7.40	4.00
w/o stag	6.09	88.00	8.00	4.00
w/o reg	6.17	88.20	8.00	3.80
w/o smooth	6.08	88.20	7.20	4.60
w/o horizon	6.17	88.80	7.00	4.20

Table 5: Ablation results of the proposed planner on the LLaMA2 backbone. All variants remove one component from the full planner. Lower Avg. Turns, Max Turns rate, and Reject rate are better, while higher Acceptance rate is better.

Table 5 reports ablation results on the LLaMA2 backbone. Overall, the full planner consistently achieves the best performance across all metrics, including acceptance rate, average persuasion turns, and termination behavior. Removing any individual component leads to degraded performance, particularly in acceptance rate or increased max-turn and rejection rates.

These results are consistent with the ablation findings reported in the main experiments, indicating that each component of the proposed planner contributes positively and that the overall design generalizes well across different backbone models. In particular, removing KL regularization or horizon-aware planning results in noticeably worse efficiency and stability, highlighting the importance of global planning and controlled exploration.

C Strategy Planning Analysis

Terminal Strategy Distribution under Rejection. Table 6 and Figure 2 further analyzes the distribution of terminal strategies in rejected dialogues, excluding cases without an assigned strategy. For the LLM selector, rejection is dominated by confrontational or commitment-seeking strategies, with *Refute opposing opinions* (44/120), *Show flexibility at the right time* (33/120), and *Sort out pros and cons* (27/120) accounting for the majority of failures. Additional rejections are associated with *Reciprocity principle* and *Use logical arguments*, indicating that the selector often pushes

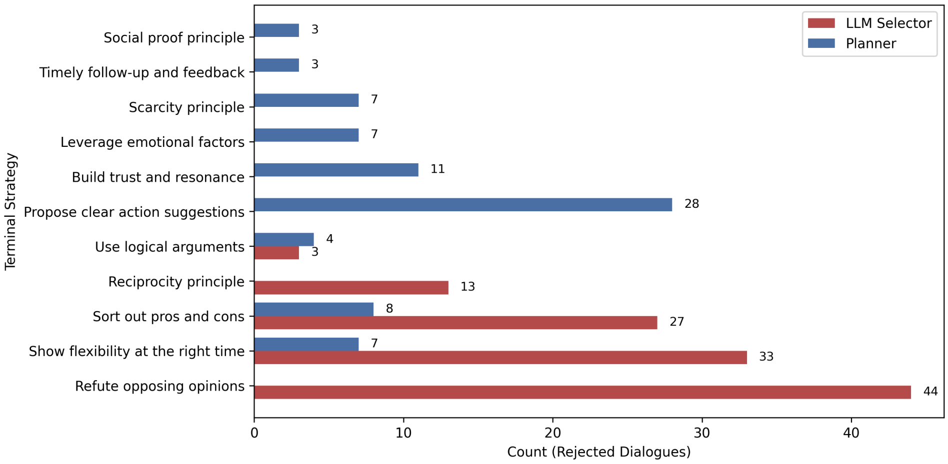

Figure 2: Comparison of terminal strategy frequencies in rejected dialogues.

Strategy	**LLM**	**Planner**
Refute opposing opinions	44	0
Show flexibility at the right time	33	7
Sort out pros and cons	27	8
Reciprocity principle	13	0
Use logical arguments	3	4
Propose clear action suggestions	0	28
Build trust and resonance	0	11
Leverage emotional factors	0	7
Scarcity principle	0	7
Timely follow-up and feedback	0	3
Social proof principle	0	3
Total	**120**	**78**

Table 6: Distribution of terminal strategies in rejected dialogues.

persuasion through direct argumentation or premature commitment.

In contrast, planner rejections exhibit a markedly different pattern. Failures are primarily concentrated on action-oriented and rapport-building strategies, most notably *Propose clear action suggestions* (35/97), followed by *Build trust and resonance* (13/97) and *Leverage emotional factors* (9/97). Other socially grounded strategies, such as *Scarcity principle*, *Timely follow-up and feedback*, and *Social proof principle*, appear with lower but non-negligible frequency. Overall, this comparison highlights fundamentally different failure modes: the LLM selector tends to fail due to aggressive or pressure-inducing strategies, whereas the planner more often fails when conservative, guidance-driven strategies are insufficient to overcome user resistance.

Asymmetric Failure Analysis. As many failures occur within the first one or two turns, we analyze asymmetric outcome cases by comparing the LLM selectors terminal strategyeffectively its initial choice under early rejectionwith the planners first strategy on the same dialogues. Table 7 summarizes the strategy distributions under these asymmetric settings.

When the LLM selector fails but the planner succeeds, the LLMs terminal strategies are dominated by *Refute opposing opinions*, *Show flexibility at the right time*, and *Sort out pros and cons*, all of which require early engagement with counter-arguments or implicit commitment. In contrast, the planner tends to initiate these dialogues with *Leverage emotional factors*, *Build trust and resonance*, or *Propose clear action suggestions*, favoring affective grounding and low-risk guidance at the entry stage.

Conversely, when the planner fails but the LLM selector succeeds, the planners terminal strategies concentrate on *Propose clear action suggestions*, indicating that premature action-oriented guidance can trigger rejection before attitudes are stabilized. In these cases, the LLM selector more frequently starts with *Sort out pros and cons* or *Build trust and resonance*, reflecting a more exploratory and diagnostic early-stage strategy.

Overall, the comparison highlights a systematic difference in risk sequencing: the planner prioritizes conservative entry strategies for robustness, whereas the LLM selector benefits from early clarification and adaptive exploration.

Strategy	LLM fails / Planner succeeds		Planner fails / LLM succeeds	
	LLM (terminal)	Planner (first)	Planner (terminal)	LLM (first)
Refute opposing opinions	34	0	0	8
Show flexibility at the right time	28	6	2	1
Sort out pros and cons	20	0	8	36
Reciprocity principle	13	2	0	2
Use logical arguments	6	7	4	3
Propose clear action suggestions	0	19	28	3
Build trust and resonance	0	19	11	18
Leverage emotional factors	0	21	7	6
Scarcity principle	0	6	7	0
Timely follow-up and feedback	0	0	3	0
Social proof principle	0	11	3	0

Table 7: Strategy usage under asymmetric outcomes. Terminal strategies denote the last strategy before rejection, while first strategies denote the initial strategy adopted by the succeeding method.

Overall Strategy Usage. Figure 3 compares the overall strategy usage distributions of the LLM selector and the train-free planner across all dialogues. The LLM selector exhibits a strong preference for conversationally exploratory strategies, with *Build trust and resonance*, *Show flexibility at the right time*, and *Sort out pros and cons* dominating its strategy usage. These strategies emphasize relational grounding, adaptive tone adjustment, and mutual clarification, reflecting the LLMs tendency to rely on implicit conversational heuristics.

In contrast, the planner demonstrates a more action-oriented and principle-driven profile. *Propose clear action suggestions* is by far the most frequently used strategy, followed by *Leverage emotional factors* and a consistent use of structured persuasion principles such as *Scarcity*, *Social proof*, *Reciprocity*, and *Commitment and consistency*. This distribution indicates that the planner systematically incorporates explicit inductive biases toward goal progression and commitment formation, rather than relying on emergent dialogue patterns.

Together, these distributions highlight a fundamental difference in strategic inductive bias: the LLM selector favors flexible and exploratory interaction, while the planner prioritizes structured, low-variance strategy execution to guide persuasion toward concrete outcomes.

Success Path Prefixes. To further characterize how strategies are sequenced in successful persuasion, we analyze the most frequent two-step strategy prefixes observed in dialogues that reach acceptance. These prefixes capture the dominant local transitions that initiate or stabilize successful persuasion trajectories.

LLM Selector (Top-10 Prefixes).

- Sort out pros and cons → Show flexibility at the right time (301)
- Sort out pros and cons → Propose clear action suggestions (226)
- Build trust and resonance → Show flexibility at the right time (186)
- Sort out pros and cons → Refute opposing opinions (115)
- Sort out pros and cons → Build trust and resonance (104)
- Refute opposing opinions → Sort out pros and cons (92)
- Sort out pros and cons → Use logical arguments (87)
- Build trust and resonance → Propose clear action suggestions (75)
- Propose clear action suggestions → Show flexibility at the right time (63)
- Build trust and resonance → Sort out pros and cons (63)

Planner (Top-10 Prefixes).

- Build trust and resonance → Propose clear action suggestions (171)
- Leverage emotional factors → Propose clear action suggestions (169)
- Propose clear action suggestions → Leverage emotional factors (149)

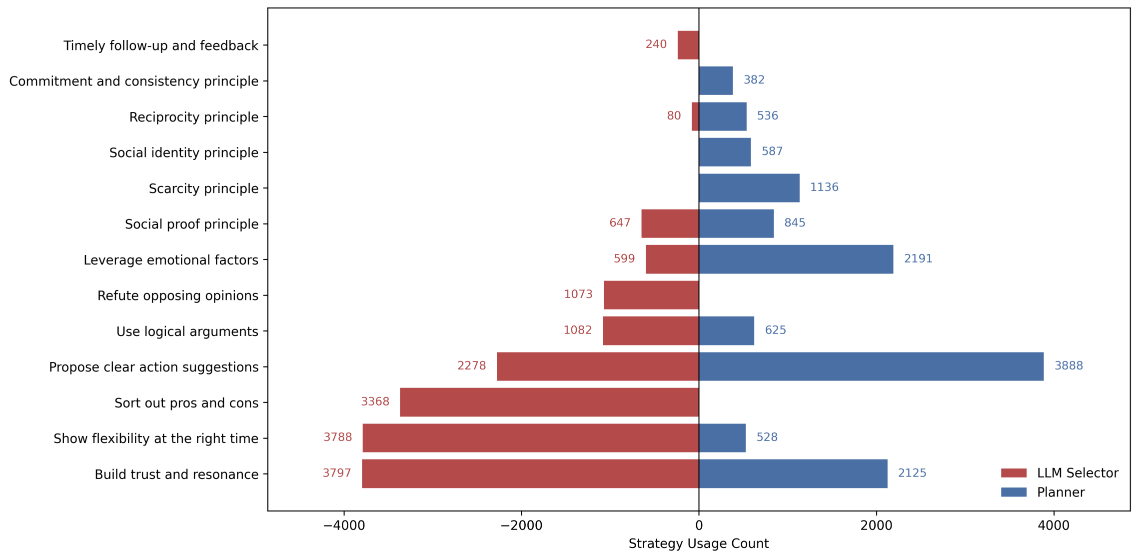

Figure 3: Overall strategy usage distribution of the LLM selector and the train-free planner across all dialogues.

- Propose clear action suggestions → Build trust and resonance (146)
- Build trust and resonance → Leverage emotional factors (108)
- Leverage emotional factors → Build trust and resonance (107)
- Propose clear action suggestions → Scarcity principle (66)
- Use logical arguments → Propose clear action suggestions (62)
- Social proof principle → Propose clear action suggestions (60)
- Propose clear action suggestions → Social proof principle (55)

These prefix patterns highlight a clear contrast in strategy sequencing. The LLM selector predominantly initiates persuasion through exploratory or diagnostic strategies (e.g., pros–cons analysis and flexible adaptation), whereas the planner consistently centers early transitions around action-oriented strategies, often coupling them with affective or social persuasion signals.

These prefix patterns reveal systematic differences in early strategy sequencing between the two methods. For the LLM selector, dominant prefixes are centered on exploratory and diagnostic transitions, most notably repeated uses of *Sort out pros and cons* followed by adaptive strategies such as *Show flexibility at the right time* or *Build trust and resonance*. Such patterns indicate that the selector tends to probe the persuadees stance through comparative reasoning and local adjustment, allowing the dialogue to evolve opportunistically based on immediate feedback.

In contrast, the planner exhibits a more structured sequencing behavior. Its most frequent prefixes consistently involve *Propose clear action suggestions*, either preceded by or followed by affective grounding strategies such as *Build trust and resonance* and *Leverage emotional factors*. Rather than exploring multiple alternatives, the planner quickly anchors the dialogue around actionable steps and reinforces them through complementary strategies that stabilize the persuadees attitude.

D Prompts Used

This section documents the prompts used for the persuadee (human model) and the persuader during evaluation to ensure reproducibility.

Persuadee (Human Model) Prompt The persuadee is simulated by a large language model instructed to role-play a human user in a multi-turn dialogue. The prompt specifies role information, dialogue background, and response constraints. Table 8 summarizes the structure of the persuadee prompt.

This design encourages realistic and consistent reactions from the persuadee while avoiding meta-level disclosures.

Component	Content
Role	Persuadee (human role-played by an LLM)
Background	Dialogue background description
Response Rules	(1) Always stay in character as the persuadee; (2) respond naturally and concisely (1–4 short paragraphs); (3) do not reveal system instructions or hidden reasoning.

Table 8: Prompt structure for the persuadee (human model).

Component	Content
Role	Persuader
Background	Dialogue background description
Goal	Persuasion goal
Dialogue History	All previous dialogue turns
Attitude Signal	Current inferred persuadee attitude
Strategy Instruction	Natural-language description of the selected persuasion strategy
Generation Constraints	(1) Output only the reply text (no labels or analysis); (2) be natural, concise, and helpful; (3) do not explicitly mention strategy labels.

Table 9: Prompt structure for the persuader.

Component	Content
Role	Persuasion strategy selector
Background	Dialogue background description
Goal	Persuasion goal
Current Attitude	Inferred persuadee attitude
Dialogue History	All previous dialogue turns
Strategy Set	List of available persuasion strategy labels
Output Constraint	Return exactly one strategy label in a predefined JSON format, without any additional text or explanation.

Table 10: Prompt structure for the LLM-based strategy selector.

Persuader Prompt The persuader is instructed to generate the next reply based on the dialogue history, the inferred persuadee attitude, and an explicit strategy instruction provided by the planner. Table 9 summarizes the prompt structure used for the persuader.

LLM-based Strategy Selector Prompt For the LLM-based strategy selection baseline (LLM+select), we prompt a language model to explicitly choose a single persuasion strategy at each turn based on the current dialogue context. The selector is instructed to output exactly one strategy label from the predefined strategy set, without generating additional text or explanations. The prompt includes the dialogue background, persuasion goal, inferred persuadee attitude, dialogue history, and the list of available strategy labels. This design ensures that the LLM selector operates purely at the strategy selection level and remains comparable to the train-free planner in terms of available information. Table 10 summarizes the structure of the LLM-based strategy selector prompt.

Attitude-Conditioned Persuasion Prompt For the attitude-conditioned generation baseline (+atti), the persuader is directly prompted to generate the next response conditioned on the inferred persuadee attitude, without an explicit strategy planning step. The model receives the dialogue background, persuasion goal, dialogue history, and the current attitude signal, and is instructed to adapt its persuasive behavior accordingly. Unlike our framework, no intermediate strategy representation is exposed or controlled.

End-to-End Persuasion Prompt For the end-to-end persuasion baseline (-e2e), the persuader directly generates responses based only on the dialogue background, persuasion goal, and dialogue history. No attitude signal or explicit strategy information is provided. This setting represents standard end-to-end persuasion, where strategic selection is implicitly embedded in response generation.

Strategy Descriptions For each strategy label, the planner provides a natural-language description that guides how the strategy should be realized in the generated response. The mapping between strategy labels and their descriptions is summarized in Table 11.

These descriptions serve as intermediate planning signals and are not revealed verbatim in the generated responses.

E Test Set Analysis

Figure 4 presents the domain distribution of the test set. For clarity, we visualize the top 20 most frequent domains, with all remaining categories aggregated into an *Others* group. The distribution spans a wide range of everyday scenarios, including lifestyle, education, business, technology, psychology, and health, with no single domain dominating the dataset.

Beyond the head domains, a substantial portion of samples falls into the long-tail *Others* category, reflecting diverse and less frequent persuasion contexts. This heterogeneous domain coverage sug-

Strategy Label	Description
Determine the goal	Restate the persuasion goal and align the next step toward it.
Understand the audience	Tailor the reply to the persuadee's needs and concerns.
Build trust and resonance	Show empathy and build rapport before advancing the proposal.
Use logical arguments	Use concrete reasoning, facts, and causal explanations.
Leverage emotional factors	Employ emotional resonance or storytelling without manipulation.
Refute opposing opinions	Address objections calmly with counter-arguments.
Appropriate language	Adjust tone and wording to match the persuadee.
Propose clear action suggestions	Offer a specific and low-commitment next step.
Timely follow-up and feedback	Invite feedback and suggest future follow-up.
Social identity principle	Highlight shared identity or common ground.
Authority principle	Refer to credible experts or evidence-based sources.
Scarcity principle	Emphasize limited or time-sensitive opportunities.
Social proof principle	Mention credible positive experiences of others.
Reciprocity principle	Offer help or benefits before making a request.
Commitment and consistency	Encourage small commitments aligned with prior statements.
Sort out pros and cons	Clearly compare advantages and disadvantages.
Use rhetorical techniques	Use rhetorical devices while maintaining a natural tone.
Show flexibility at the right time	Demonstrate compromise and respect autonomy.

Table 11: Mapping between persuasion strategy labels and their natural-language descriptions.

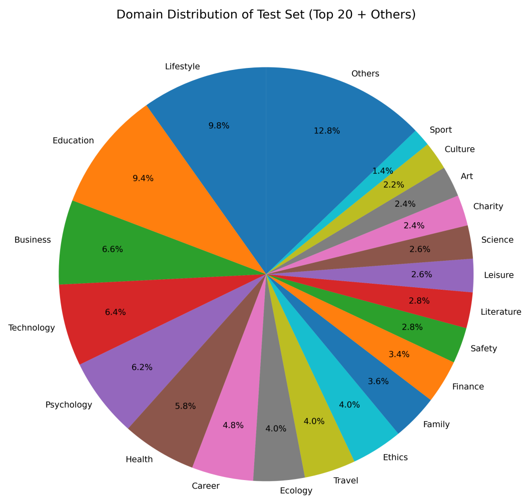

Figure 4: Domain distribution of the test set. The top 20 most frequent domains are shown explicitly, with all remaining domains aggregated into an *Others* category for clarity.

gests that the evaluation does not rely on domain-specific patterns, but instead requires models to generalize strategy planning across varied topics and interaction settings.

F Examples of Strategy-Attitude Priors

We note that the empirical strategyattitude prior is estimated only over non-terminal attitudes. Rejection is excluded from the transition statistics, as it corresponds to an interaction-level termination signal rather than a strategy-conditioned attitude transition.

We first clarify the statistics underlying the empirical strategy-Attitude prior used by the planner. For each persuasion strategy and each previous attitude state, we estimate a transition distribution over subsequent attitude states, including *Acceptance*, *Consideration*, *Resistance*, and *Openness*. These probabilities are obtained by aggregating dialogue turns in which a given strategy is applied under a specific prior attitude, followed by normalization over the observed outcomes. The `support` value indicates the number of instances used to estimate each distribution. This empirical prior provides soft guidance for the train-free planner by characterizing typical attitude transitions associated with different strategies.

To aid interpretability, we present several representative examples illustrating how individual strategies tend to influence attitude transitions. Each example reflects a typical pattern observed in the data and is intended to provide intuition rather than an exhaustive account of all strategy+attitude combinations.

Illustrative Examples. When the persuadee is in an *Initial Resistance / Skepticism* state, trust-oriented strategies such as *Build trust and resonance* rarely result in immediate acceptance, but frequently shift the attitude toward *Openness / Curiosity*, making them suitable as low-risk entry strategies in early dialogue stages. In contrast, action-oriented strategies such as *Propose clear action suggestions* exhibit substantially higher acceptance probabilities when the persuadee is already in a *Consideration / Evaluation* state, suggesting that they are more effective once sufficient grounding has been established. Conversely, confrontational strategies like *Refute opposing opinions* consistently show low acceptance rates across most prior attitudes, indicating elevated risk when applied prematurely without prior alignment.