THINK SOCIALLY VIA COGNITIVE REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

LLMs trained for logical reasoning excel at step-by-step deduction to reach verifiable answers. Yet, this paradigm is ill-suited for navigating social situations, which require an interpretive process for ambiguous cues that rarely yield a definitive outcome. To bridge this gap, we introduce Cognitive Reasoning, a paradigm modeled on human social cognition. It structures the interpretive process as an adaptive cognitive flow of interconnected cognitive units (e.g., *observation* or *attribution*) for effective social thinking and responses. We then propose CogFlow, a framework that instills this capability in LLMs. CogFlow first curates a dataset of cognitive flows through tree-structured simulation of the associative and progressive nature of human thought. After instilling the basic cognitive reasoning capability via supervised fine-tuning, CogFlow adopts reinforcement learning to enable the model to improve itself via trial and error, guided by a multi-objective reward that optimizes both cognitive flow and response quality. Extensive experiments show that CogFlow improves the social cognitive capabilities of LLMs and supports humans for more effective social decision-making, while generalizing promisingly to diverse reasoning tasks. Our repository is available at: `https://anonymous.4open.science/r/CogFlow-iclr2027/`.

1 INTRODUCTION

Social cognition, the core mental process of human social intelligence, governs how individuals perceive, interpret, and respond to social situations (Fiske & Taylor, 2020). This unique ability allows humans to navigate complex social dynamics wisely (Thorndike, 1920). As large language models (LLMs, (Guo et al., 2025; OpenAI, 2024b; Yang et al., 2025)) increasingly collaborate with humans, their capability of social intelligence is being actively examined (Chen et al., 2024; 2025a). Recent studies have revealed promising signs, including evidence of human-like social behaviors (Park et al., 2023) and lobe structure for social skills (Zhou et al., 2025a). Deeper cognitive analysis further suggests that LLMs spontaneously exhibit human-like cognitive features, e.g., reasoning patterns that mimic empathy (Dong et al., 2025), indicating a potential capacity for social cognition.

Despite the potential evidenced in the aforementioned observational studies, improving the social cognitive abilities of LLMs remains insufficiently explored. The root cause lies in the fundamental mismatch between the LLMs' currently implanted reasoning structures and the nature of social intelligence (Moore et al., 2025). Specifically, LLMs excel at complex tasks like math and coding (Ni et al., 2024; Shao et al., 2024), which rely on **step-by-step logical deduction** to arrive at a single verifiable solution (Zheng et al., 2025). In contrast, reasoning in social situations is an **interpretive process** that involves analyzing ambiguous cues that rarely yield a definitive answer (Gandhi et al., 2023; Xu et al., 2025). Not to mention LLMs, even when humans try to apply rigid logic rules to the fluid social domains, they risk falling into "*cognitive rumination*" (Marjanović et al., 2025; Wu et al., 2026), which is a state of over-analyzing simple cues, engaging in redundant reasoning cycles, and producing protracted internal monologues that lead to erroneous judgments or delayed responses (as shown in Fig. 1). This exposes a vital challenge in applying LLMs to social situations, and defining and implementing an effective LLM reasoning paradigm to close this gap is thus urgently needed.

To this end, we pioneer a complete learning framework to instill social cognition into LLM reasoning. Drawing from social cognitive theory (Bandura et al., 1986), we operationalize social cognition process using six theoretically grounded cognitive units that form social thinking: ***Observation, Attribution, Motivation, Regulation, Efficacy, and Behavior***. For example, in the social scene "*choosing the last member for a new robotics team*" shown in Figure 1, one can predict Carlos's

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

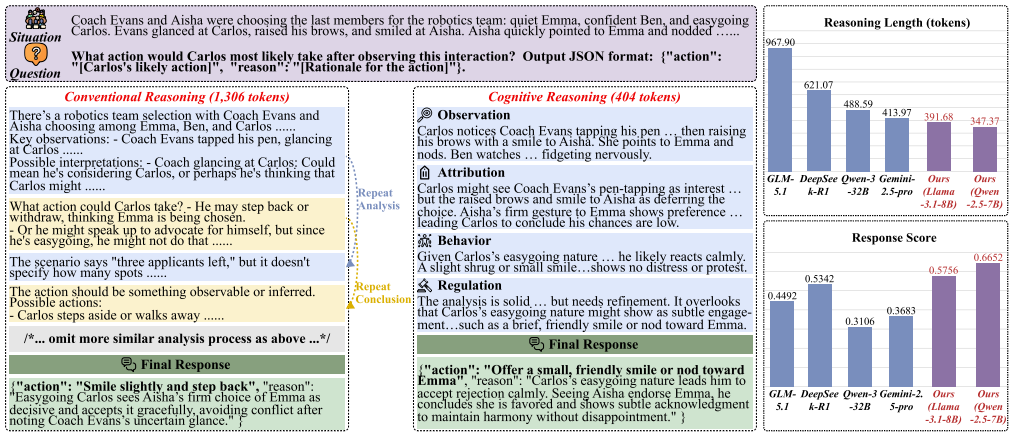


Figure 1: An example of conventional reasoning (DeepSeek-R1) falling into "cognitive rumination", while cognitive reasoning efficiently reaches a better response. Bar charts show the average reasoning length and comparative preference scores of LLMs on our test set (§4.3).

action by first *observing* the situation (e.g., *coach Evans smiles at Aisha*), making an *attribution* about that behavior (e.g., *the smile signals deference*), formulating Carlos's intended *behavior* (e.g., *a slight shrug*), applying *regulation* (e.g., *Carlos should show a friendly nod in accordance with his easygoing personality*), and finally leading to predicted *actions* (e.g., *offer a small, friendly smile or nod toward Emma*). These cognitive units flow adaptively among each other to create an effective, structured reasoning process. We define this process as **Cognitive Reasoning**, a paradigm for thinking and responding effectively in social situations. While cognitive reasoning provides a clear blueprint for social cognition, its implementation in LLMs presents two crucial challenges: **1) Reasoning paradigm shift**: shifting models' training objective from optimizing verifiable logic to guiding analytical reasoning that lacks definitive answers; **2) Cognitive flow control**: teaching the model to adaptively regulate its use of these cognitive units to avoid rumination.

To address these challenges, we teach LLMs to think socially in a form of cognitive flow. **First**, we curate a cognitive reasoning dataset via cognitive flow simulation. We prompt advanced LLMs to simulate human thoughts by crafting cognitive flows about a social situation. This process generates cognitive units sequentially, where each unit acts as a reasoning node that enables the planning of the next, mirroring the associative and progressive process of human cognition. The uncertainty in social situations allows these nodes to naturally branch into a cognitive reasoning tree, and each leaf node contains the response derived from the corresponding cognitive flow about the social situation (as shown in Figure 2). **Second**, given the absence of definitive answers, we design a comparative preference ranking principle to identify the most promising cognitive flows by the relative plausibility of the responses from all leaf nodes. We then prune the flows based on criteria derived from social cognitive theory – coherence, interpretability, predictability (Bandura et al., 1986) – to create high-quality data for supervised fine-tuning (SFT). **Finally**, after instilling basic cognitive reasoning capability via SFT, we empower the model to autonomously explore better reasoning paths using reinforcement learning (RL), guided by a multi-objective reward function: a) a comparative preference reward to steer the model toward flows that yield more plausible responses; and b) a cognitive flexibility reward to encourage adaptive regulation of the cognitive flow's diversity and depth. We name this training framework CogFlow.

Our contributions are summarized as follows: (1) We introduce cognitive reasoning, a paradigm designed to enable LLMs to think socially via structured interplay among cognitive units. (2) We propose CogFlow, a training framework that instills cognitive reasoning capability into LLM, using a combination of preference-based SFT and multi-objective RL. (3) Extensive experiments show that CogFlow improves the social cognitive capabilities of LLMs and supports humans for more effective social decision-making, while promisingly generalizing to diverse reasoning tasks.

2 PRELIMINARIES

Definition of Cognitive Reasoning Humans' social cognition is a dynamic process (Fiske & Taylor, 2020) where people navigate complex social situations by building and refining internal

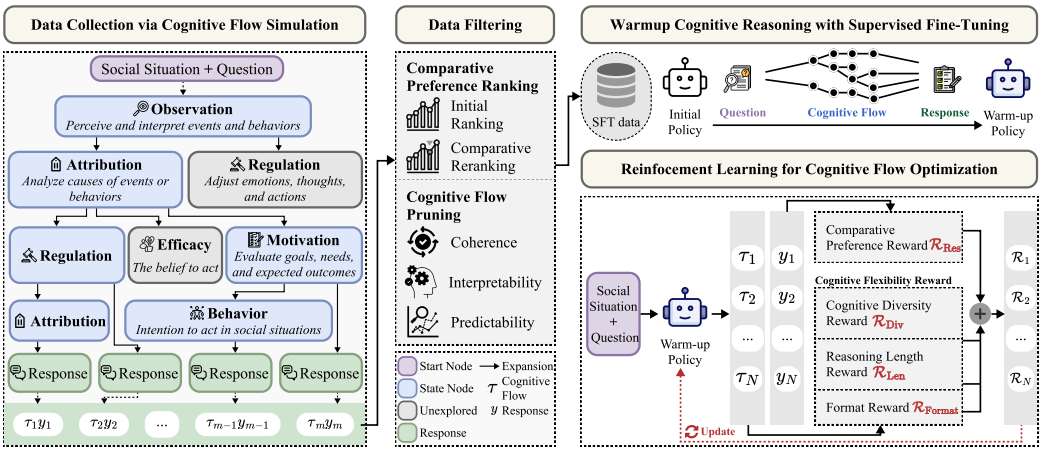


Figure 2: Overview of our training framework CogFlow, which crafts cognitive flows via tree-structured planning, uses the filtered data for supervised fine-tuning (SFT), and then employs multi-objective reinforcement learning (RL) for self-improvement.

cognitive maps (Tolman, 1948). This occurs in a feedback loop: people map social situations to their actions (Bandura & Walters, 1977), and the outcomes provide feedback that reshapes both their internal map and the external situations (Bandura et al., 1986). We formalize the mental activity within this loop as *cognitive reasoning*, and operationalize the cognitive map as a "cognitive flow", an adaptive adoption of several core cognitive units (Bandura et al., 1986), e.g., *attributions about a person's intent shape one's motivation to interact, and a strong sense of efficacy can enhance regulation strategies*. Definitions of the cognitive units are: (1) `Observation`: Perceiving and interpreting events and others' behaviors to form an initial cognitive appraisal of a situation (Lazarus, 1991). (2) `Attribution`: Analyzing the causes of events or behaviors (Heider, 2013). (3) `Motivation`: The expectations and value assessments of potential outcomes for self and others' behavior, which provide the drive to act (Vroom, 1964). (4) `Regulation`: Reflecting on and adjusting emotions, beliefs, and behaviors in pursuit of social goals (Carver & Scheier, 2012). (5) `Efficacy`: The belief to execute a specific social behavior, which influences motivational intensity (Bandura, 1997). (6) `Behavior`: Formulating an intention to act in response to social situations (Ajzen, 1991). Crucially, these units are not isolated components, but dynamically coupled elements of a reasoning process. For example, attributions about a person's intent can shape one's motivation to interact, and a strong sense of efficacy can support more effective regulation strategies. This flexible interplay equips cognitive reasoning with the analytical capacity needed to navigate ambiguous and fluid social situations, distinguishing it from traditional logical reasoning.

Task Formulation in Cognitive Reasoning Given a social situation $\mathcal{S}$ and a query $\mathcal{Q}$, the goal is to obtain a response y by first generating an explicit cognitive flow τ. Each reasoning step is materialized by a particular cognitive unit $r_i = (u_i, c_i)$, where u_i is the unit category (e.g., `Observation`) and c_i is the materialized text content of u_i. A complete flow τ is thus an ordered sequence of n reasoning steps: $\tau = \{r_1, r_2, \cdots, r_n\} = \{(u_1, c_1), (u_2, c_2), \cdots, (u_n, c_n)\}$. We define the input x as the concatenation of $\mathcal{S}$ and $\mathcal{Q}$, $x = [\mathcal{S}; \mathcal{Q}]$. Our goal is to learn a policy π_θ that maximizes the joint probability of generating flow τ and response y: $\pi_\theta(\tau, y|x) = \pi_\theta(y|\tau, x) \cdot \pi_\theta(\tau|x)$, where $\pi_\theta(\tau|x)$ depicts the generation of structured cognitive flow τ, $\pi_\theta(y|\tau, x)$ measures the correspondence between the cognitive reasoning content of τ and produced response y about the input x.

3 Methodology

As shown in Figure 2, our training framework CogFlow begins by collecting cognitive flows via tree-structured planning, guided by carefully crafted instructions. We implement a dual-validated data filtering procedure: a comparative preference ranking module identifies cognitive flows that yield high-quality responses, which are then pruned under coherence, interpretability, and predictability criteria. After instilling such structured cognitive reasoning into an LLM via SFT, the model improves itself via RL, guided by a multi-objective reward that optimizes both cognitive flow and response quality. The instructions used in data collection and filtering are shown in Appendix G.

3.1 Data Collection via Cognitive Flow Simulation

Seed Data Collection To approach realistic and complex social situations, we collect seed data from Reddit. Unlike existing datasets, e.g., SocialIQA (Sap et al., 2019), which often feature simple situations with limited social dynamics, our collection focuses on complex multi-person interactions, presenting a more substantial challenge for LLMs. The pipeline is constructed as follows: **1) Situation curation**: We curate and anonymize Reddit posts, removing all sensitive content. Then we prompt Deepseek-R1 (hereafter R1) to distill them into concise situation descriptions ($\mathcal{S}$). **2) Question generation**: For each situation, we prompt R1 to extract detailed social cues and generate a corresponding question ($\mathcal{Q}$) that demands deep interpretation, analysis, and prediction. **3) Filtering**: To ensure high quality of our seed data, each generated situation-question pair ($\mathcal{S}$,$\mathcal{Q}$) is re-assessed by R1. We discard pairs rated low on the situation complexity or question relevance.

Cognitive Flow Simulation To obtain cognitively motivated flows, we simulate the associative and progressive process of human cognition (Bandura et al., 1986). The flows are crafted by prompting R1 to sequentially plan and materialize the content of cognitive units. Each unit acts as a reasoning node, and the preceding path supports planning for the next. As a single thought can lead to multiple continuations, this process naturally forms a tree-structured exploration of reasoning flows.

- **State**: Each node in the tree is a state s_k, denoting a partially generated cognitive reasoning path from the root. The root node, s_0, is the initial input $x = [S; Q]$.
- **Action**: At any state s_k, we prompt R1 to choose the next cognitive unit u_{k+1} from dynamic candidates, $A(s_k) \subseteq \{u_1, \cdots, u_6\}$. The initial state s_0 is constrained to be unit `Observation`.
- **Planning**: Planning begins from the root state s_0 and iteratively expands the tree by: *a) Generation*: For the selected unit u_{k+1}, we prompt R1 to generate the unit's text content c_{k+1} with respect to the current state s_k. This forms a new reasoning node $r_{k+1} = (u_{k+1}, c_{k+1})$ and expands the path to a new state $s_{k+1} = [s_k; r_{k+1}]$. *b) Prediction*: R1 is then prompted to analyze the reasoning path s_{k+1} and predict a set of relevant next cognitive units $A(s_{k+1})$ to explore. This step adaptively prunes the action space from all six possible units to only the most contextually appropriate ones. *c) Expansion*: Each candidate unit from the predicted set becomes a new node, expanding the tree with multiple parallel reasoning paths.
- **Completion**: This expansion process repeats until R1 determines that the reasoning has reached a terminal state, which is referred to as a leaf node. At this point, it generates a final response y using the fully constructed chain.

We define a complete cognitive flow from the root node to any leaf node as a rollout. By performing multiple rollouts for each seed instance, we collect a diverse set of cognitive flows $\{\tau_1,\cdots,\tau_m\}$ and their final responses $\{y_1,\cdots,y_m\}$.

Dual-Validation based Filtering In the absence of definitive answers, we design a two-step filtering procedure to ensure the quality of generated cognitive flows. We first identify flows landing on high quality responses via **two-stage Comparative Preference Ranking** (*CPRank*^{2}):

- **Comparison pool construction**: For each seed instance, we craft a candidate pool containing responses from our rollouts and a baseline response directly from R1. The pool size is set to 10, which we found to be satisfactory in our preliminary tests. If the size of valid rollouts is fewer than 10, we create variations by perturbing the generated flows (e.g., combining flow snippets).
- **Initial ranking**: We prompt R1 to generate situation-specific criteria and then use them to assign an initial score and critique for each response in the pool, i.e., LLM-as-a-judge (Liu et al., 2025).
- **Comparative reranking**: To mitigate scoring biases (e.g., positional bias in R1's initial scores), we then select the median-ranked response as an anchor. We ask R1 to perform a final comparative reranking of the pool against this anchor, yielding a more robust preference order.

Following the ranking, we conduct **cognitive flow pruning**. We select flows with responses scored higher than those from R1, designating them as high-quality candidates. The candidates are then pruned by R1 using the following criteria constructed based on social cognitive theories: *a) coherence*: logically sound and free of contradictions; *b) interpretability*: clearly explain the social dynamics; *c) predictability*: offer reasonable insight into the future evolution of social dynamics. Only cognitive flows satisfying all criteria are retained.

3.2 Warmup Cognitive Reasoning with SFT

To endow LLM with basic cognitive reasoning capability, we train it with the constructed cognitive flows via SFT. For each curated data instance (x, τ, y), we format it by concatenating all reasoning

steps $r_i = (u_i, c_i)$ within the cognitive flow τ into a continuous text sequence τ_{SFT}: $\tau_{\text{SFT}} = \oplus_{r_i \in \tau} \langle u_i \rangle c_i \langle / u_i \rangle$, where $\oplus$ is string concatenation. The cognitive unit tags $\langle u_i \rangle$ and $\langle / u_i \rangle$ (e.g., $\langle$`Observation`$\rangle$ and $\langle /$`Observation`$\rangle$) are added to LLM's vocabulary as new special tokens, allowing them to be directly embedded and enabling the LLM to learn the cognitive reasoning structure intrinsically. The policy π_θ is optimized by minimizing the standard SFT loss:

$$\mathcal{L}_{\text{SFT}}(\theta) = -\mathbb{E}_{(x,\tau,y)\sim\mathcal{D}_{\text{SFT}}}[\log(\pi_\theta(\tau_{\text{SFT}}, y|x))], \tag{1}$$

where $\mathcal{D}_{\text{SFT}}$ is our curated data for SFT. After warmup, π_θ is able to generate structured cognitive flows using these specialized tags without relying on manually crafted prompts.

3.3 RL for Cognitive Flow Optimization

To enable the model to progressively refine its cognitive flows, we adopt RL, where the policy model learns to improve its cognitive reasoning via trial and error. We use GRPO (detailed in Appendix E) to optimize policy π_θ, which is guided by our multi-objective reward as follows:

Comparative Preference Reward ($\mathcal{R}_{\text{Res}}$) While the preference ranking used for data filtering is well-suited for scenarios lacking definitive answers, it is not economically feasible for large-scale online training (which needs to execute R1 against each generated rollouts). We thus train a dedicated reward model RM_ϕ to predict pairwise preference. For each input x, RM_ϕ is trained to predict whether a candidate response y is preferred over a set of k reference responses $\{y^1_{\text{ref}}, \cdots, y^k_{\text{ref}}\}$ (we use $k = 3$). During RL, for each generated response y, we use the top-k responses from our curated data as the reference set and set the reward to be RM_ϕ's predicted probability of y is preferred:

$$\mathcal{R}_{\text{Res}}(y|x) = P_\phi(y \succ \{y^1_{\text{ref}}, \cdots, y^k_{\text{ref}}\}|x). \tag{2}$$

Cognitive Flexibility Reward Beyond response, we foster policy π_θ to regulate its thought process.

- **Cognitive diversity reward ($\mathcal{R}_{\text{Div}}$)**: To prevent the model from falling into simplistic or repetitive reasoning patterns, we introduce $\mathcal{R}_{\text{Div}}$ to encourage exploration of diverse cognitive flows. This design is inspired by human social cognition, where people flexibly adapt their cognitive strategies to situational nuances (Fiske & Taylor, 2020). The reward evaluates a cognitive flow τ by incentivizing the use of rarer cognitive units within a batch of rollouts, encouraging the model to avoid over-reliance on common reasoning steps. Given m rollouts for an input x, yielding flows $\{\tau_1, \ldots, \tau_m\}$, the reward for a chain τ containing v unique cognitive units $\{u_1, \ldots, u_v\}$ is:
$$\mathcal{R}_{\text{Div}}(\tau) = -\frac{1}{v}\sum_{j=1}^{v} \log(p(u_j)), \tag{3}$$
where $p(u_j)$ is the frequency of the cognitive unit u_j across all m sampled cognitive flows.
- **Reasoning length reward ($\mathcal{R}_{\text{Len}}$)**: While cognitive diversity is crucial, it must be balanced with conciseness to avoid cognitive rumination, i.e., overly long and unproductive reasoning. We therefore introduce $\mathcal{R}_{\text{Len}}$ to encourage focused yet comprehensive thought by penalizing cognitive flows that are either too short or too long. For each input x, we build a dynamic target length range $[L_{\min}, L_{\max}]$ derived from the top-k reference flows in our curated data. The reward is calculated using a soft bounding function created by multiplying two sigmoid functions. This forms a "reward window" that gently penalizes flows whose length is outside the desired range:
$$\mathcal{R}_{\text{Len}}(\tau) = \sigma\Big(\frac{|\tau| - (L_{\min})/2}{L_{\min}/8}\Big) \cdot \sigma\Big(\frac{L_{\max} + L_{\min} - |\tau|}{L_{\max}/8}\Big). \tag{4}$$
- **Structural format reward ($\mathcal{R}_{\text{Format}}$)**: To maintain structural integrity, we use a rule-based binary reward to encourage cognitive flows to follow $\langle u_i \rangle c_i \langle / u_i \rangle$ structure:
$$\mathcal{R}_{\text{Format}}(\tau) = \begin{cases} 1 & \text{if format of } \tau \text{ is valid,} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Weighted Reward Function The final reward for a rollout (τ, y) is a weighted sum of all rewards, with the format reward acting as a gate to discard structurally invalid flows:

$$\mathcal{R} = \mathcal{R}_{\text{Format}} \cdot (\omega_1 \cdot \mathcal{R}_{\text{Res}} + \omega_2 \cdot \mathcal{R}_{\text{Div}} + \omega_3 \cdot \mathcal{R}_{\text{Len}}). \tag{6}$$

4 Experiments

4.1 Experimental Setup

Datasets We collected 5,100 social situations from Reddit, spanning 5 major categories and 16 subcategories, all passing rigorous safety filtering (Kim et al., 2023). Each seed instance yielded 3.6

Table 1: Results of consistency between LLM evaluators and human judgments. The agreement ratio *kappa* $\in$ [0.41, 0.6] is moderate agreement.

Evaluators	Easy	Medium	Hard	Overall
Score-R1	0.5604	0.4938	0.4848	0.5141
Score-Q32B	0.5824	0.5219	0.5152	0.5405
Score-Claude-Sonnet-5	0.5282	0.5734	0.5359	0.5521
CPRank^1-Q32B	0.6374	0.5094	0.6515	0.5669
CPRank^1-R1	0.6374	0.5625	0.4697	0.5757
CPRank^1-Claude-Sonnet-5	0.6134	0.5442	0.5478	0.5629
RM$_\phi$-7B	0.5714	0.5781	**0.6667**	0.5863
CPRank^2-Q32B	**0.6648**	0.6031	0.5909	0.6215
CPRank^2-R1	0.6538	0.6312	0.6212	0.6373
CPRank^2-Claude-Sonnet-5	0.6540	**0.6474**	0.6390	**0.6476**
kappa	0.4534	0.4559	0.5041	0.4693

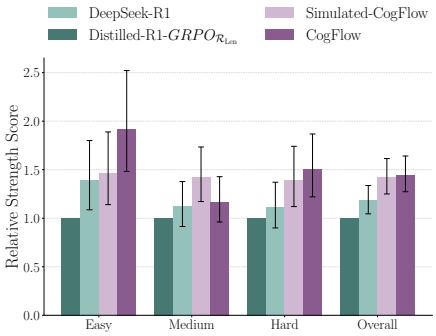


Figure 3: Pairwise results from Bradley-Terry model. The higher strength score, the better.

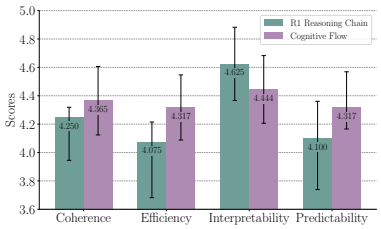
high-quality cognition flows, with 4 cognitive units per flow on average. We employ six domain experts (with Master's degrees or higher) to inspect 500 random instances, achieving a 96.8% pass rate that indicates the dataset's high quality. For model training, we allocate 1,000 seed instances (3,661 cognitive flows) for SFT and 3,600 instances for RL (3,200 for training and 400 for validation), and another 500 for final evaluation. Moreover, we extract 26,676 candidate-reference response pairs to train our comparative preference reward model. See Appendix F.1 for the dataset's more details.

Baselines and LLM Evaluators We compared baselines: **(1) Tuning-free Reasoning LLMs**: OpenAI o3, GPT-5.6-Sol, GLM-5.1 (GLM et al., 2025), Qwen-3-32B (Yang et al., 2025), DeepSeek-R1, Gemini-2.5-pro/flash&3.8-flash. **(2) Fine-tuned Open-source LLMs**: using Llama-3.1-8B and Qwen-2.5-7B backbones, we train: a) *Direct-SFT/GRPO*: directly on curated responses of our dataset without cognitive flow, where GRPO is vanilla and relies solely on our $\mathcal{R}_{\text{Res}}$ reward (same use below). b) *Distilled-R1-SFT/GRPO/GRPO*$_{\mathcal{R}_{\text{Len}}}$: on distilled R1's reasoning chains via SFT, GRPO, and GRPO using both $\mathcal{R}_{\text{Res}}$ and $\mathcal{R}_{\text{Len}}$. c) *CogFlow-SFT/GRPO*: on our dataset with cognitive flows. We use R1, cost-effective Qwen-3-32B, Claude-Sonnet-5 as our LLM evaluators to perform two-stage comparative preference ranking, called *CPRank*^{2}*-R1/Q32B/Claude-Sonnet-5*. For each test instance, each generated response is ranked within a comparison pool containing the pre-curated reference responses. The rank is normalized to a score by $(M - rank)/(M - 1)$, M=10. A model's performance is its average score across all test instances. See Appendices F.2, F.3, F.4, and F.5 for more details about baseline implementation, LLM evaluators, and baseline results.

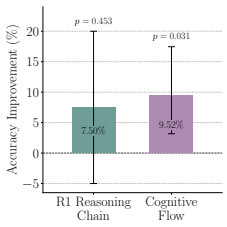
4.2 Main Results by Human Evaluations

LLM Evaluators' Consistency with Human Judgment We evaluate 10 LLM evaluators: 1) *CPRank*^{2}*-Q32B&R1&Claude-Sonnet-5* and their single-stage variation without reranking (*CPRank*^{1}), 2) reward model *RM*$_\phi$ trained on Qwen-2.5-7B, 3) prompt-based direct scoring (*Score-Q32B&R1&Claude-Sonnet-5*). Six experts perform pairwise comparisons on all 500 seed instances in our test set, each with 4 distinct responses from 4 models: *CogFlow* (trained on Llama, same use below experiments), *Distilled-R1-GRPO*$_{\mathcal{R}_{\text{Len}}}$, *DeepSeek-R1*, and *Simulated-CogFlow* (top-ranked response from cognitive flow simulation). To balance workload, 500 instances are split into two sets, each assigned to 3 experts. Experts provide *win/tie/loss* judgments for all response pairs and label difficulty of each instance (*easy, medium, hard*). The final preference label and the difficulty label are determined by majority vote among experts. We measure consistency between the pairwise orderings of LLM evaluators and the aggregated human judgments. Results in Table 1 show: *(1) CPRank*^{2} *series achieves the highest overall consistency.* Against their single-stage *CPRank*^{1}, the reranking step is crucial for improving alignment by mitigating initial scoring biases. *(2) our trained reward model RM*$_\phi$ *is an effective proxy,* outperforming all direct scoring and single-stage ranking baselines and offering a cost-effective substitute for expensive LLM judges during online training.

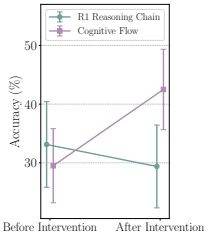
Results of Expert Pairwise Comparison We convert experts' pairwise *win/tie/loss* judgments into scalar scores using the Bradley-Terry model (Bradley & Terry (1952), detailed in Appendix F.6). The results in Figure 3 show that *CogFlow* surpasses its teacher model (*DeepSeek-R1*), showing our framework enables a smaller model to effectively learn and use cognitive reasoning to produce high-quality responses. More importantly, *CogFlow* performs on par with *Simulated-CogFlow*, while *Distilled-R1-GRPO*$_{\mathcal{R}_{\text{Len}}}$ remains inferior to its teacher. This clearly suggests that cognitive reasoning is consistently beneficial for social responding, and cognitive reasoning is effectively learnable.



(a) Results of reasoning quality rated by humans with four criteria on a 1-5 scale.



(b) Results of helpfulness for human decision-making.



(c) Results of cognitive intervention for humans.

Figure 4: Quantitative analysis of two types of reasoning (cognitive reasoning vs. R1 reasoning).

Helpfulness of Cognitive Flow for Humans Beyond LLM evaluation, we assess cognitive flow as a cognitive aid for human decision-making. We hired 6 additional annotators to assess 100 multiple-choice instances from the expert-curated dataset. For each instance, the golden response from expert consensus is explained by one of two reasoning styles: our ***cognitive flow*** or ***R1 reasoning chain***, shown in a balanced frequency. Annotators are evenly split to perform two tasks: *(1) Quality ratings* on a 1-5 scale for *coherence*, *efficiency* (conciseness of content and logic), *interpretability*, and *predictability*; *(2) Value test*: Annotators first select their preferred response from four responses and are then shown the reasoning behind the expert-preferred response. Finally, they are asked to make their final decision. We define "*Helpfulness*" as the average accuracy improvement before and after exposing the reasoning. Results in Figure 4a and 4b reveal that our cognitive flows are better received by annotators, as they are considered more coherent, efficient, and have higher predictability, making them better tools to understand social situations. We can also notice that there is a trade-off between interpretability and efficiency: R1's reasoning often includes exhaustive self-reflection that, while boosting its interpretability by listing all social cues, does so at the cost of lower efficiency. Most importantly, cognitive flows are more helpful for human decision-making, yielding a higher accuracy improvement (*p=0.031*), showing their potential to augment human social intelligence.

Cognitive Intervention for Humans We conduct a preliminary intervention study to further examine cognitive reasoning's effects on humans. We compare two forms of guidance: **1) *Cognitive flow*-style**: emphasize key social cues and analytical transitions among cognitive units; and **2) *R1 reasoning chain*-style**: provide a chain-of-thought summary. To create the guidance, we convert the cognitive flow/R1 reasoning chain into hints that expose the thought process without revealing the final answer (see examples in Appendix F.7). We hire 20 volunteers who are randomly assigned to experimental (*cognitive flow-style*) or control group (*R1 reasoning chain-style*). Before intervention, each completes 10 tasks without guidance, and then an ANOVA test (Fisher, 1970) confirms no statistically significant differences exist between the two groups (p=0.42). During the intervention, participants sequentially complete 25 instances without guidance and 25 with guidance, allowing us to measure the change in decision-making accuracy. Results in Figure 4c show the ***cognitive flow*-style guidance significantly improves participants' social decision-making accuracy** (*p=0.017, Cohen's d=0.92*), while the *R1 reasoning chain*-style shows a slight downward trend. This reveals the potential of structured cognitive reasoning to improve human social thinking.

4.3 Results on Automated Evaluations

Main Results For each model, we generate 4 responses per test instance and report their average scores in Table 2. Results reveal *CogFlow* outperforms all baselines across both backbones and evaluators, showing the effectiveness of combining structured cognitive reasoning with RL. Cognitive reasoning has a clear edge to unstructured reasoning in improving model learning, e.g., *CogFlow* vs. *Distilled-R1-GRPO*$_{\mathcal{R}_{\text{Len}}}$. Across both model family and reasoning style, models tuned with RL clearly outperform their SFT counterparts, showing RL's ability to effectively refine the reasoning strategies. Another finding is that models trained on cognitive flows produce shorter with better performance, showing the capability of cognitive reasoning to reduce reasoning costs.

Ablation Study of Rewards Results in Table 2 reveal: (1) $\mathcal{R}_{\text{Div}}$ promotes exploration but requires constraints. Removing $\mathcal{R}_{\text{Len}}$ sharply degrades performance and nearly doubles reasoning length (e.g., 391.68 vs. 725.77 tokens on Llama), showing that unconstrained diversity leads to inefficient reasoning. (2) $\mathcal{R}_{\text{Len}}$ regularizes reasoning efficiency and quality. Without $\mathcal{R}_{\text{Div}}$, performance still exceeds the GRPO baseline while reasoning remains concise (282.73 vs. 314.03 tokens on Llama), showing that $\mathcal{R}_{\text{Len}}$ effectively guides concise, high-quality reasoning.

Table 2: Automatic evaluation results with three CPRank^2 judges. Best results of two model families are **bold**. w/o is reward/unit ablation, blackboard-bold letters are unit initials (e.g., $\mathbb{A}$ is Attribution).

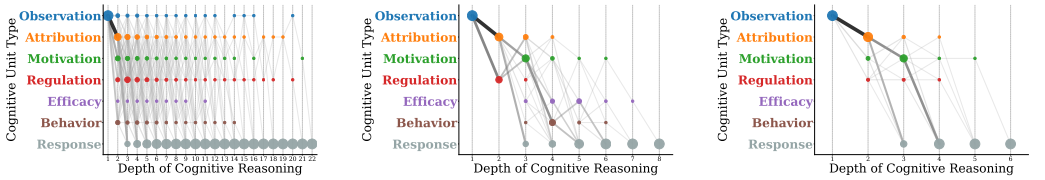
Models	CPRank^2-R1 (↑)		CPRank^2-Q32B (↑)		CPRank^2-Claude-Sonnet-5 (↑)		Reasoning Length (tokens, ↓)
	Overall	Easy/Medium/Hard	Overall	Easy/Medium/Hard	Overall	Easy/Medium/Hard	
Tuning-free Reasoning LLMs							
o3	0.2933	0.4191/0.2497/0.2214	0.4096	0.4567/0.4106/0.3659	0.4151	0.4552/0.3331/0.4626	163.08
Qwen3-32B	0.3106	0.3696/0.3709/0.1912	0.4243	0.5184/0.3893/0.3875	0.3110	0.3750/0.3164/0.2417	488.59
Gemini-2.5-flash	0.3111	0.4316/0.2557/0.2460	0.4383	0.4713/0.4419/0.3977	0.2989	0.3166/0.2848/0.2954	360.42
Gemini-2.5-pro	0.3683	0.4347/0.2818/0.3883	0.5037	0.5525/0.4862/0.4835	0.4437	0.4189/0.4055/0.5068	413.97
GLM-5.1	0.4397	0.4455/0.4296/0.4517	0.5638	0.5099/0.5927/0.5691	0.4765	0.4843/0.4731/0.4746	967.90
DeepSeek-R1	0.5342	0.5220/0.4990/0.5816	0.6578	0.6267/0.6485/0.7067	0.4800	0.5054/0.4351/0.5159	621.07
Gemini-3.8-flash	0.3273	0.3791/0.2873/0.3445	0.4236	0.4552/0.4125/0.4103	0.4289	0.4043/0.3978/0.5103	–
GPT-5.6-Sol	0.4661	0.5182/0.4548/0.3864	0.6684	0.6587/0.6444/0.6831	0.4933	0.5202/0.4458/0.5501	–
Tuned Llama-3.1-8B-Instruct Series							
Direct-SFT	0.3545	0.3962/0.3490/0.3193	0.5407	0.6236/0.5144/0.5011	0.3223	0.4208/0.2870/0.2830	–
Direct-GRPO	0.5041	0.5154/0.5764/0.4208	0.7196	0.7751/0.7332/0.6380	0.3498	0.4288/0.3585/0.2542	–
Distilled-R1-SFT	0.3213	0.4530/0.2202/0.2941	0.4508	0.5036/0.4383/0.4181	0.3408	0.3613/0.3578/0.3110	554.76
Distilled-R1-GRPO	0.5157	0.5603/0.5601/0.4279	0.7310	0.8127/0.7400/0.6305	0.3488	0.3641/0.3570/0.3183	568.90
Distilled-R1-GRPO$_{\mathcal{R}_{\text{Len}}}$	0.5017	0.6167/0.4438/0.4474	0.7519	0.8080/0.7423/0.7108	0.4111	0.5532/0.3837/0.3141	444.90
CogFlow-SFT	0.4024	0.4827/0.3443/0.3821	0.5999	0.6472/0.5772/0.5916	0.4385	0.5110/0.4217/**0.3952**	451.14
CogFlow-GRPO	0.5564	**0.6974**/0.5501/0.3193	0.7420	0.7534/0.7441/**0.7265**	0.4580	0.5606/0.4759/0.3125	314.03
CogFlow (ours)	**0.5756**	0.6645/0.5350/**0.5294**	**0.7828**	**0.8271/0.7908**/0.7232	**0.4960**	0.6086/**0.5079**/0.3631	391.68
CogFlow (w/o $\mathcal{R}_{\text{Len}}$)	0.5525	0.6199/0.5665/0.4727	0.7069	0.7271/0.7359/0.6347	0.4468	0.5459/0.4504/0.3374	725.77
CogFlow (w/o $\mathcal{R}_{\text{Div}}$)	0.5702	0.5783/**0.6250**/0.5073	0.7574	0.8176/0.7431/0.7202	0.4878	**0.6111**/0.4701/0.3926	282.73
CogFlow (w/o $\mathbb{REB}$ units)	0.4327	0.5192/0.4489/0.3137	0.6510	0.6479/0.6676/0.6250	0.3079	0.3475/0.3042/0.2733	298.40
CogFlow (w/o $\mathbb{EB}$ units)	0.4594	0.6217/0.4379/0.3279	0.6760	0.7480/0.6502/0.6463	0.3930	0.5297/0.3760/0.2801	332.19
CogFlow (w/o $\mathbb{AM}$ units)	0.3649	0.5240/0.3346/0.2520	0.5111	0.6305/0.4885/0.4602	0.2911	0.3269/0.2871/0.2614	312.45
Tuned Qwen-2.5-7B-Instruct Series							
Direct-SFT	0.3144	0.3451/0.3239/0.2742	0.5113	0.5862/0.5016/0.4500	0.2963	0.3231/0.3299/0.2321	–
Direct-GRPO	0.6148	0.7147/0.6407/0.4914	0.7630	0.8221/0.7605/0.7062	0.3006	0.3332/0.3125/0.2465	–
Distilled-R1-SFT	0.1751	0.2083/0.1101/0.1984	0.3776	0.4086/0.3689/0.3606	0.3518	0.3951/0.3557/0.3295	711.80
Distilled-R1-GRPO	0.5261	0.6109/0.4682/0.5013	0.7061	0.7171/0.7395/0.6355	0.3696	0.4394/0.3815/0.2772	955.16
Distilled-R1-GRPO$_{\mathcal{R}_{\text{Len}}}$	0.5298	0.5727/0.5970/0.4206	0.7458	0.8038/0.7398/0.6962	0.3985	0.5208/0.3915/0.2857	437.06
CogFlow-SFT	0.3672	0.4186/0.4032/0.2810	0.5567	0.5838/0.5564/0.5291	0.4582	0.5208/0.4400/0.4285	368.41
CogFlow-GRPO	0.5988	**0.6750**/0.5871/0.5361	0.7542	0.7971/0.7526/0.7124	0.5074	0.5752/0.5237/0.4097	237.11
CogFlow (ours)	**0.6652**	0.6404/**0.7531/0.6015**	**0.7956**	**0.8248/0.7963/0.7641**	**0.5727**	**0.6452/0.5858/0.4769**	347.37
CogFlow (w/o $\mathcal{R}_{\text{Len}}$)	0.6142	0.6462/0.6133/0.5840	0.7568	0.7784/0.7610/0.7269	0.5316	0.6418/0.5496/0.3865	502.30
CogFlow (w/o $\mathcal{R}_{\text{Div}}$)	0.6084	0.6660/0.6249/0.5356	0.7824	0.7902/0.7930/0.7555	0.5202	0.5681/0.5208/0.4717	277.24
CogFlow (w/o $\mathbb{REB}$ units)	0.3085	0.4286/0.3061/0.1879	0.3824	0.5387/0.2833/0.3894	0.4484	0.5641/0.4237/0.3720	331.90
CogFlow (w/o $\mathbb{EB}$ units)	0.3558	0.4647/0.3549/0.2435	0.5256	0.6875/0.4683/0.4503	0.3515	0.4141/0.3466/0.2953	352.72
CogFlow (w/o $\mathbb{AM}$ units)	0.4499	0.5618/0.4440/0.3434	0.5792	0.7068/0.6399/0.3417	0.3215	0.3506/0.3357/0.2660	298.45

Ablation Study of Cognitive Units We craft representative ablations upon the empirical frequencies of cognitive units in Figure 5, avoiding exhaustive combinations: $\mathbb{REB}$ and $\mathbb{EB}$ remove low-frequency units, while $\mathbb{AM}$ removes high-frequency units (letters denote unit initials, e.g., $\mathbb{AM}$ for Attribution and Motivation). Rather than masking units at inference, which may introduce distribution shift, we retrain each variant by selecting SFT flows without the ablated units, regenerate trajectories under the reduced unit space to maintain a training set comparable in size to CogFlow, and rerun SFT+RL with identical settings. Table 2 reveals all variants underperform full CogFlow on all three evaluators, showing that restricting the cognitive-unit space degrades response quality.

4.4 In-depth Analysis

Transition Patterns of Cognitive Units To dissect the structure of cognitive flow, we plot the frequency of each cognitive unit at different reasoning depths (denoted by node size) and the transition probability between the units (denoted by edge thickness). Figure 5 reveals that: (1) *Cogflow* can guide the model to learn more ordered cognitive strategies. Against diffuse reasoning patterns from *CogFlow-SFT*, *CogFlow* exhibits a more structured and hierarchical cognitive flow, which also leads to higher-quality responses, as supported by Table 2. (2) The diversity reward $\mathcal{R}_{\text{Div}}$ is critical for preventing pattern collapse. *CogFlow (w/o $\mathcal{R}_{\text{Div}}$)* shows a stark collapse into a monotonous and rigid reasoning path (e.g., *Observation→Attribution→Motivation*). This highlights the importance of $\mathcal{R}_{\text{Div}}$ for maintaining cognitive flexibility, supporting the results of ablation study.

Generalization Capabilities of CogFlow To test whether CogFlow serves as a transferable reasoning scaffold rather than a domain-specific heuristic, we evaluate it on three out-of-domain benchmarks: MATH-500 (math reasoning, (Lightman et al., 2024)), SocialIQA (social commonsense), and HiToM (theory of mind, (Wu et al., 2023)). Using Llama-3.1-8B as the backbone, we compare CogFlow with the strong Distilled-R1-GRPO$_{\mathcal{R}_{Len}}$ on two settings: **1) Direct Task Transfer**, which evaluates zero-shot performance of models trained on social situations when applied directly to new tasks. **2) Task Adaptation**, which initializes from CogFlow-SFT (trained only on cognitive flows) and applied our RL method (§3.3) to MATH (Hendrycks et al., 2021) and SocialIQA without any



(a) Patterns in *CogFlow-SFT* (b) Patterns in *CogFlow* (c) Patterns in *CogFlow (w/o* $\mathcal{R}_{\text{Div}}$*)*

Figure 5: The transition patterns of cognitive units in reasoning. Depth is the sequential position of a unit within a cognitive flow. The proportion of units at different depths is denoted by node size, and the transition probability between units is denoted by edge thickness.

Table 3: Results of CogFlow's generalization capability, evaluated on three out-of-domain benchmarks using accuracy (Acc, ↑) and reasoning length (Len, ↓) metrics.

Models	MATH-500		SocialIQA		HiToM	
	Acc	Len	Acc	Len	Acc	Len
Tuning-free Backbone Model						
Llama-3.1-8B	42.2	-	70.7	-	19.4	-
Llama-3.1-8B (CoT)	44.0	308.18	67.6	249.13	33.4	722.64
Direct Task Transfer						
Distilled-R1-GRPO$_{\mathcal{R}_{Len}}$	44.2	1034.54	66.3	399.27	35.7	1209.72
CogFlow (Ours)	44.6	500.27	70.4	227.76	**36.3**	412.99
Task Adaptation						
Distilled-R1-adapt	49.0	449.88	75.7	348.26	-	-
CogFlow-adapt (Ours)	**53.4**	304.86	**78.7**	166.28	-	-

domain-specific SFT data. This yields **CogFlow-adapt**, compared with **Distilled-R1-adapt** initialized from Distilled-R1-SFT, testing whether SFT-learned cognitive reasoning can generalize to new tasks via RL. Table 3 reports accuracy and reasoning length, with tuning-free Llama-3.1-8B and its CoT variant as the baselines. Results reveal that direct task transfer yields only marginal gains under domain shift, whereas task adaptation (CogFlow-adapt) shows strong generalization, significantly outperforming standard baselines and Distilled-R1 trained on the same data (*p=0.015 on SocialIQA and p=0.040 on MATH-500*). These results support CogFlow as a robust reasoning paradigm beyond social reasoning. ***More in-depth analysis is presented in Appendix F.8.***

5 Related Work

Recent works have revealed parallels between LLMs and human social behavior (Park et al., 2023), social skills (Zhou et al., 2025a), and cognitive habits in reasoning (Dong et al., 2025; Kargupta et al., 2025), driving to integrate cognitive theories (Chen et al., 2024; 2025a) like simulation theory (Wilf et al., 2024; Sarangi et al., 2025), to guide LLM reasoning (Wang et al., 2024a; AlKhamissi et al., 2025). Yet, a substantial line of work rely on prompting to enforce specific strategies (Wang et al., 2024a; Park et al., 2025; Aytes et al., 2025; Zhou et al., 2025b). While externally shaping LLMs' reasoning, it encourages format-level imitation without necessarily internalizing adaptive reasoning. (Zhou et al., 2023). A practical solution is to internalize reasoning into LLM's parameters via training (Magister et al., 2023; Paliotta et al., 2025), adapting prompt-based CoT (Wei et al., 2023; Yao et al., 2023) for reasoning models (OpenAI, 2024a; Guo et al., 2025). This has proven effective for tasks like math (Shao et al., 2024), which rely on step-by-step logical deductions to reach verifiable outcomes (Zheng et al., 2025). Yet, it can induce over-thinking (Cuadron et al., 2025; Chen et al., 2025b; Kumar et al., 2025), a state of repetitive thought cycling (Gandhi et al., 2025), prompting efforts to improve efficiency (Wang et al., 2024b; 2025). Critically, such logical reasoning is ill-suited for social situations, which involve an interpretive process of analyzing ambiguous cues that rarely yield a definitive answer (Gandhi et al., 2023; Xu et al., 2025; Moore et al., 2025; Kong et al., 2025). Using this deductive paradigm in the fluid social domain risks "cognitive rumination", i.e., over-analysis simple social cues (Hou et al., 2025; Marjanović et al., 2025; Mathur et al., 2025; Wu et al., 2026; Kim et al., 2026). Thus, we introduce cognitive reasoning to bridge this gap.

6 Conclusions

In this paper, we introduce cognitive reasoning, a paradigm that models human social cognition by formulating it into a structured cognitive flow of interconnected cognitive units. We then propose CogFlow, a framework that instills the cognitive reasoning capability in LLMs using a combination of preference-based SFT and multi-objective RL. Extensive experiments show CogFlow significantly enhances the social cognitive capabilities of LLMs, yielding more effective social decision-making and superior task generalization. Human intervention trials also reveal that structured cognitive flow is promising for augmenting human social intelligence.

REFERENCES

Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

Badr AlKhamissi, C. Nicolò De Sabbata, Zeming Chen, Martin Schrimpf, and Antoine Bosselut. Mixture of cognitive reasoners: Modular reasoning with brain-like specialization, 2025. URL https://arxiv.org/abs/2506.13331.

Simon A. Aytes, Jinheon Baek, and Sung Ju Hwang. Sketch-of-thought: Efficient LLM reasoning with adaptive cognitive-inspired sketching. *CoRR*, abs/2503.05179, 2025. doi: 10.48550/ARXIV.2503.05179. URL https://doi.org/10.48550/arXiv.2503.05179.

Albert Bandura. *Self-efficacy: The exercise of control*. Macmillan, 1997.

Albert Bandura and Richard H Walters. *Social learning theory*, volume 1. Prentice hall Englewood Cliffs, NJ, 1977.

Albert Bandura et al. Social foundations of thought and action. *Englewood Cliffs, NJ*, 1986(23-28): 2, 1986.

Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

Charles S Carver and Michael F Scheier. *Attention and self-regulation: A control-theory approach to human behavior*. Springer Science & Business Media, 2012.

Ruirui Chen, Weifeng Jiang, Chengwei Qin, and Cheston Tan. Theory of mind in large language models: Assessment and enhancement, 2025a. URL https://arxiv.org/abs/2505.00026.

Xingyu Chen, Jiahao Xu, Tian Liang, Zhiwei He, Jianhui Pang, Dian Yu, Linfeng Song, Qiuzhi Liu, Mengfei Zhou, Zhuosheng Zhang, Rui Wang, Zhaopeng Tu, Haitao Mi, and Dong Yu. Do not think that much for 2+3=? on the overthinking of o1-like llms, 2025b. URL https://arxiv.org/abs/2412.21187.

Zhuang Chen, Jincenzi Wu, Jinfeng Zhou, Bosi Wen, Guanqun Bi, Gongyao Jiang, Yaru Cao, Mengting Hu, Yunghwei Lai, Zexuan Xiong, and Minlie Huang. Tombench: Benchmarking theory of mind in large language models, 2024. URL https://arxiv.org/abs/2402.15052.

Alejandro Cuadron, Dacheng Li, Wenjie Ma, Xingyao Wang, Yichuan Wang, Siyuan Zhuang, Shu Liu, Luis Gaspar Schroeder, Tian Xia, Huanzhi Mao, Nicholas Thumiger, Aditya Desai, Ion Stoica, Ana Klimovic, Graham Neubig, and Joseph E. Gonzalez. The danger of overthinking: Examining the reasoning-action dilemma in agentic tasks, 2025. URL https://arxiv.org/abs/2502.08235.

Jianshuo Dong, Yujia Fu, Chuanrui Hu, Chao Zhang, and Han Qiu. Towards understanding the cognitive habits of large reasoning models. *CoRR*, abs/2506.21571, 2025. doi: 10.48550/ARXIV.2506.21571. URL https://doi.org/10.48550/arXiv.2506.21571.

Ronald Aylmer Fisher. Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution*, pp. 66–70. Springer, 1970.

Susan T Tufts Fiske and Shelley E Taylor. Social cognition: From brains to culture. 2020.

Kanishk Gandhi, Jan-Philipp Fränken, Tobias Gerstenberg, and Noah D. Goodman. Understanding social reasoning in language models with language models, 2023. URL https://arxiv.org/abs/2306.15448.

Kanishk Gandhi, Ayush Chakravarthy, Anikait Singh, Nathan Lile, and Noah D. Goodman. Cognitive behaviors that enable self-improving reasoners, or, four habits of highly effective stars, 2025. URL https://arxiv.org/abs/2503.01307.

GLM, Aohan Zeng, Xin Lv, Qinkai Zheng, Zhenyu Hou, Bin Chen, Chengxing Xie, Cunxiang Wang, Da Yin, Hao Zeng, Jiajie Zhang, Kedong Wang, Lucen Zhong, Mingdao Liu, Rui Lu, Shulin Cao, Xiaohan Zhang, Xuancheng Huang, Yao Wei, Yean Cheng, Yifan An, Yilin Niu, Yuanhao Wen, Yushi Bai, Zhengxiao Du, Zihan Wang, Zilin Zhu, Bohan Zhang, Bosi Wen, Bowen Wu, Bowen Xu, Can Huang, Casey Zhao, Changpeng Cai, Chao Yu, Chen Li, Chendi Ge, Chenghua Huang, Chenhui Zhang, Chenxi Xu, Chenzheng Zhu, Chuang Li, Congfeng Yin, Daoyan Lin, Dayong Yang, Dazhi Jiang, Ding Ai, Erle Zhu, Fei Wang, Gengzheng Pan, Guo Wang, Hailong Sun, Haitao Li, Haiyang Li, Haiyi Hu, Hanyu Zhang, Hao Peng, Hao Tai, Haoke Zhang, Haoran Wang, Haoyu Yang, He Liu, He Zhao, Hongwei Liu, Hongxi Yan, Huan Liu, Huilong Chen, Ji Li, Jiajing Zhao, Jiamin Ren, Jian Jiao, Jiani Zhao, Jianyang Yan, Jiaqi Wang, Jiayi Gui, Jiayue Zhao, Jie Liu, Jijie Li, Jing Li, Jing Lu, Jingsen Wang, Jingwei Yuan, Jingxuan Li, Jingzhao Du, Jinhua Du, Jinxin Liu, Junkai Zhi, Junli Gao, Ke Wang, Lekang Yang, Liang Xu, Lin Fan, Lindong Wu, Lintao Ding, Lu Wang, Man Zhang, Minghao Li, Minghuan Xu, Mingming Zhao, Mingshu Zhai, Pengfan Du, Qian Dong, Shangde Lei, Shangqing Tu, Shangtong Yang, Shaoyou Lu, Shijie Li, Shuang Li, Shuang-Li, Shuxun Yang, Sibo Yi, Tianshu Yu, Wei Tian, Weihan Wang, Wenbo Yu, Weng Lam Tam, Wenjie Liang, Wentao Liu, Xiao Wang, Xiaohan Jia, Xiaotao Gu, Xiaoying Ling, Xin Wang, Xing Fan, Xingru Pan, Xinyuan Zhang, Xinze Zhang, Xiuqing Fu, Xunkai Zhang, Yabo Xu, Yandong Wu, Yida Lu, Yidong Wang, Yilin Zhou, Yiming Pan, Ying Zhang, Yingli Wang, Yingru Li, Yinpei Su, Yipeng Geng, Yitong Zhu, Yongkun Yang, Yuhang Li, Yuhao Wu, Yujiang Li, Yunan Liu, Yunqing Wang, Yuntao Li, Yuxuan Zhang, Zezhen Liu, Zhen Yang, Zhengda Zhou, Zhongpei Qiao, Zhuoer Feng, Zhuorui Liu, Zichen Zhang, Zihan Wang, Zijun Yao, Zikang Wang, Ziqiang Liu, Ziwei Chai, Zixuan Li, Zuodong Zhao, Wenguang Chen, Jidong Zhai, Bin Xu, Minlie Huang, Hongning Wang, Juanzi Li, Yuxiao Dong, and Jie Tang. Glm-4.5: Agentic, reasoning, and coding (arc) foundation models, 2025. URL https://arxiv.org/abs/2508.06471.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638, 2025.

Fritz Heider. *The psychology of interpersonal relations*. Psychology Press, 2013.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In Joaquin Vanschoren and Sai-Kit Yeung (eds.), *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, virtual*, 2021.

Guiyang Hou, Wenqi Zhang, Zhe Zheng, Yongliang Shen, and Weiming Lu. Scaling LLMs' social reasoning: Sprinkle cognitive "aha moment" into fundamental long-thought logical capabilities. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 3126–3138, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-256-5. doi: 10.18653/v1/2025.findings-acl.162. URL https://aclanthology.org/2025.findings-acl.162/.

David R. Hunter. MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics*, 32(1):384 – 406, 2004. doi: 10.1214/aos/1079120141. URL https://doi.org/10.1214/aos/1079120141.

Priyanka Kargupta, Shuyue Stella Li, Haocheng Wang, Jinu Lee, Shan Chen, Orevaoghene Ahia, Dean Light, Thomas L. Griffiths, Max Kleiman-Weiner, Jiawei Han, Asli Celikyilmaz, and Yulia Tsvetkov. Cognitive foundations for reasoning and their manifestation in llms, 2025. URL https://arxiv.org/abs/2511.16660.

Eunsu Kim, Junyeong Park, Juhyun Oh, Kiwoong Park, Seyoung Song, A. Seza Doğruöz, Alice Oh, and Najoung Kim. Are they lovers or friends? evaluating LLMs' social reasoning in English and Korean dialogues. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 23431–23451, San Diego, California, United States, July 2026. Association for Computational Linguistics. ISBN 979-8-89176-390-6. doi: 10.18653/v1/2026.acl-long.1074. URL https://aclanthology.org/2026.acl-long.1074/.

Hyunwoo Kim, Jack Hessel, Liwei Jiang, Peter West, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, and Yejin Choi. Soda: Million-scale dialogue distillation with social commonsense contextualization, 2023. URL https://arxiv.org/abs/2212.10465.

Aobo Kong, Wentao Ma, Shiwan Zhao, Yongbin Li, Yuchuan Wu, Ke Wang, Xiaoqian Liu, Qicheng Li, Yong Qin, and Fei Huang. SDPO: Segment-level direct preference optimization for social agents. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12409–12423, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.607. URL https://aclanthology.org/2025.acl-long.607/.

Abhinav Kumar, Jaechul Roh, Ali Naseh, Marzena Karpinska, Mohit Iyyer, Amir Houmansadr, and Eugene Bagdasarian. Overthink: Slowdown attacks on reasoning llms, 2025. URL https://arxiv.org/abs/2502.02542.

Richard S Lazarus. *Emotion and adaptation*. Oxford University Press, 1991.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=v8L0pN6EOi.

Zijun Liu, Peiyi Wang, Runxin Xu, Shirong Ma, Chong Ruan, Peng Li, Yang Liu, and Yu Wu. Inference-time scaling for generalist reward modeling, 2025. URL https://arxiv.org/abs/2504.02495.

Lucie Charlotte Magister, Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. Teaching small language models to reason, 2023. URL https://arxiv.org/abs/2212.08410.

Sara Vera Marjanović, Arkil Patel, Vaibhav Adlakha, Milad Aghajohari, Parishad BehnamGhader, Mehar Bhatia, Aditi Khandelwal, Austin Kraft, Benno Krojer, Xing Han Lù, Nicholas Meade, Dongchan Shin, Amirhossein Kazemnejad, Gaurav Kamath, Marius Mosbach, Karolina Stańczak, and Siva Reddy. Deepseek-r1 thoughtology: Let's think about llm reasoning, 2025. URL https://arxiv.org/abs/2504.07128.

Leena Mathur, Marian Qian, Paul Pu Liang, and Louis-Philippe Morency. Social genome: Grounded social reasoning abilities of multimodal models. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 24868–24891, Suzhou, China, November 2025. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.1264. URL https://aclanthology.org/2025.emnlp-main.1264/.

Meta. Llama 3 model card. 2024. URL https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md.

Jared Moore, Ned Cooper, Rasmus Overmark, Beba Cibralic, Nick Haber, and Cameron R. Jones. Do large language models have a planning theory of mind? evidence from mindgames: a multi-step persuasion task, 2025. URL https://arxiv.org/abs/2507.16196.

Ansong Ni, Miltiadis Allamanis, Arman Cohan, Yinlin Deng, Kensen Shi, Charles Sutton, and Pengcheng Yin. Next: Teaching large language models to reason about code execution, 2024. URL https://arxiv.org/abs/2404.14662.

Lixing Niu, Jiapeng Li, Xingping Yu, Shu Wang, Ruining Feng, Bo Wu, Ping Wei, Yisen Wang, and Lifeng Fan. Rˆ 3-vqa:" read the room" by video social reasoning. *arXiv preprint arXiv:2505.04147*, 2025.

Keane Ong, Wei Dai, Carol Li, Dewei Feng, Hengzhi Li, Jingyao Wu, Jiaee Cheong, Rui Mao, Gianmarco Mengaldo, Erik Cambria, and Paul Pu Liang. Human behavior atlas: Benchmarking unified psychological and social behavior understanding, 2026. URL https://arxiv.org/abs/2510.04899.

OpenAI. Openai o1 system card, 2024a. URL https://cdn.openai.com/o1-system-card.pdf.

OpenAI. Gpt-4o system card, 2024b. URL https://arxiv.org/abs/2410.21276.

Daniele Paliotta, Junxiong Wang, Matteo Pagliardini, Kevin Y. Li, Aviv Bick, J. Zico Kolter, Albert Gu, François Fleuret, and Tri Dao. Thinking slow, fast: Scaling inference compute with distilled reasoners, 2025. URL https://arxiv.org/abs/2502.20339.

Pranshu Pandya, Vatsal Gupta, Agney S Talwarr, Tushar Kataria, Dan Roth, and Vivek Gupta. Ntsebench: Cognitive reasoning benchmark for vision language models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pp. 3680–3708, 2025.

Eunkyu Park, Wesley Hanwen Deng, Gunhee Kim, Motahhare Eslami, and Maarten Sap. Cognitive chain-of-thought: Structured multimodal reasoning about social situations, 2025. URL https://arxiv.org/abs/2507.20409.

Joon Sung Park, Joseph C. O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In Sean Follmer, Jeff Han, Jürgen Steimle, and Nathalie Henry Riche (eds.), *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST 2023, San Francisco, CA, USA, 29 October 2023- 1 November 2023*, pp. 2:1–2:22. ACM, 2023. doi: 10.1145/3586183.3606763. URL https://doi.org/10.1145/3586183.3606763.

Maarten Sap, Hannah Rashkin, Derek Chen, Ronan Le Bras, and Yejin Choi. Socialiqa: Commonsense reasoning about social interactions. *CoRR*, abs/1904.09728, 2019. URL http://arxiv.org/abs/1904.09728.

Sneheel Sarangi, Maha Elgarf, and Hanan Salam. Decompose-ToM: Enhancing theory of mind reasoning in large language models through simulation and task decomposition. In Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert (eds.), *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 10228–10241, Abu Dhabi, UAE, January 2025. Association for Computational Linguistics. URL https://aclanthology.org/2025.coling-main.682/.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024. URL https://arxiv.org/abs/2402.03300.

Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. *arXiv preprint arXiv: 2409.19256*, 2024.

EL Thorndike. Intelligence and its uses. *Harper's magazine*, 1920.

Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.

Victor H Vroom. Work and motivation. *John Willey & Sons*, 1964.

Chenlong Wang, Yuanning Feng, Dongping Chen, Zhaoyang Chu, Ranjay Krishna, and Tianyi Zhou. Wait, we don't need to "wait"! removing thinking tokens improves reasoning efficiency, 2025. URL https://arxiv.org/abs/2506.08343.

Xin Wang, Boyan Gao, Yi Dai, Lei Cao, Liang Zhao, Yibo Yang, and David Clifton. Cognition chain for explainable psychological stress detection on social media, 2024a. URL https://arxiv.org/abs/2412.14009.

Xinyi Wang, Lucas Caccia, Oleksiy Ostapenko, Xingdi Yuan, William Yang Wang, and Alessandro Sordoni. Guiding language model reasoning with planning tokens, 2024b. URL https://arxiv.org/abs/2310.05707.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023. URL https://arxiv.org/abs/2201.11903.

Alex Wilf, Sihyun Lee, Paul Pu Liang, and Louis-Philippe Morency. Think twice: Perspective-taking improves large language models' theory-of-mind capabilities. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8292–8308, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.451. URL https://aclanthology.org/2024.acl-long.451/.

Jincenzi Wu, Yuxuan Lei, Jianxun Lian, Yitian Huang, Lexin Zhou, Haotian Li, Xing Xie, and Helen Meng. Social-r1: Towards human-like social reasoning in llms, 2026. URL https://arxiv.org/abs/2603.09249.

Yufan Wu, Yinghui He, Yilin Jia, Rada Mihalcea, Yulong Chen, and Naihao Deng. Hi-tom: A benchmark for evaluating higher-order theory of mind reasoning in large language models. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, pp. 10691–10706. Association for Computational Linguistics, 2023. doi: 10.18653/V1/2023.FINDINGS-EMNLP.717. URL https://doi.org/10.18653/v1/2023.findings-emnlp.717.

Zixiang Xu, Yanbo Wang, Yue Huang, Jiayi Ye, Haomin Zhuang, Zirui Song, Lang Gao, Chenxi Wang, Zhaorun Chen, Yujun Zhou, Sixian Li, Wang Pan, Yue Zhao, Jieyu Zhao, Xiangliang Zhang, and Xiuying Chen. Socialmaze: A benchmark for evaluating social reasoning in large language models, 2025. URL https://arxiv.org/abs/2505.23713.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025. URL https://arxiv.org/abs/2505.09388.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models, 2023. URL https://arxiv.org/abs/2305.10601.

Chujie Zheng, Zhenru Zhang, Beichen Zhang, Runji Lin, Keming Lu, Bowen Yu, Dayiheng Liu, Jingren Zhou, and Junyang Lin. Processbench: Identifying process errors in mathematical reasoning, 2025. URL https://arxiv.org/abs/2412.06559.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, Bangkok, Thailand, 2024. Association for Computational Linguistics. URL http://arxiv.org/abs/2403.13372.

Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Libiao Peng, Jiaming Yang, Xiyao Xiao, Sahand Sabour, Xiaohan Zhang, Wenjing Hou, Yijia Zhang, Yuxiao Dong, Jie Tang, and Minlie Huang. Characterglm: Customizing chinese conversational AI characters with large language models. *CoRR*, abs/2311.16832, 2023. doi: 10.48550/ARXIV.2311.16832. URL https://doi.org/10.48550/arXiv.2311.16832.

Jinfeng Zhou, Yuxuan Chen, Yihan Shi, Xuanming Zhang, Leqi Lei, Yi Feng, Zexuan Xiong, Miao Yan, Xunzhi Wang, Yaru Cao, Jianing Yin, Shuai Wang, Quanyu Dai, Zhenhua Dong, Hongning Wang, and Minlie Huang. Socialeval: Evaluating social intelligence of large language models. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, pp. 30958–31012. Association for Computational Linguistics, 2025a. URL https://aclanthology.org/2025.acl-long.1496/.

Xuhui Zhou, Jiarui Liu, Akhila Yerukola, Hyunwoo Kim, and Maarten Sap. Social world models, 2025b. URL https://arxiv.org/abs/2509.00559.

A LIMITATIONS

Cultural Limitations Social norms are inherently subjective and culturally dependent. Our model is trained on data derived from Reddit, which may disproportionately reflect English-speaking social norms and demographics. Consequently, the social logic learned by the model may not generalize to other cultural contexts and could inadvertently reinforce specific social biases present in the training corpus. We encourage future research to explore cognitive reasoning across diverse cultural settings to mitigate these biases.

Teacher Model Dependence in Cognitive Flow Construction In our work, we use DeepSeek-R1 as the primary teacher model to construct social situations, generate cognitive-flow rollouts, and perform automatic filtering. This choice reflects a practical trade-off between construction quality, reproducibility, token cost, and annotation time, rather than a requirement of the CogFlow framework itself. Since large-scale cognitive-flow construction is expensive, using a single strong teacher enables a controlled and repeatable pipeline for evaluating the feasibility of cognitive reasoning. Nevertheless, this empirical instantiation may introduce teacher-specific biases in generated situations, reasoning flows, and filtering decisions. We mitigate this risk through multi-stage filtering (§3.1), expert inspection of sampled instances (§4.1), human pairwise evaluation (Figure 3), and evaluation across multiple backbone model families, e.g., Qwen-3-32B and Claude-Sonnet-5 (Table 2). These safeguards provide evidence that the observed gains are not solely artifacts of a single teacher model, and they may not fully eliminate teacher-induced bias. Importantly, this limitation concerns the current experimental setup rather than the algorithmic design of CogFlow. CogFlow is model-agnostic: the teacher used for cognitive flow construction can be replaced by another capable LLM, an ensemble of LLMs, or human-written process supervision. Future work could scale to multi-teacher construction, cross-teacher agreement filtering, and human-adjudicated process labels to further reduce dependence on any single teacher while preserving the same overall framework.

Preference-based Evaluation and Reward Modeling Social reasoning often admits multiple plausible interpretations rather than a single definitive answer. Therefore, our data filtering, automatic evaluation, and reinforcement learning stages rely on comparative preference ranking and a learned reward model as scalable proxies for response quality. A specific concern is that DeepSeek-R1 serves simultaneously as the teacher model for cognitive flow construction, the primary automatic evaluator, and a key comparison baseline. This overlap creates a potential circularity: an R1-based judge may systematically favor reasoning styles that resemble R1's own outputs, which could inflate the apparent advantage of CogFlow over non-R1-distilled baselines and deflate it relative to R1 itself. To reduce the risk of relying solely on automatic judgments, we conduct extensive human validation (§4.2): expert annotations are used to assess the alignment between different LLM-based evaluators (e.g., Qwen-3-32B and Claude-Sonnet-5) and human preferences, and additional human pairwise evaluations are used to verify whether the conclusions suggested by automatic evaluation are consistent with human judgments. These analyses show that our two-stage comparative ranking

aligns better with expert judgments than direct scoring, and that its overall agreement with humans reaches a moderate level (Table 1). The human evaluation results further support the main trends observed in automatic evaluation (Figure 4). Critically, if R1-based evaluation carries a stylistic bias toward R1-like reasoning chains, the observed gains of CogFlow over Distilled-R1 baselines would be understated rather than overstated, since cognitive flows are structurally dissimilar to R1's chain-of-thought format. The human pairwise evaluations in §4.2, which are independent of any LLM judge, are therefore the most direct evidence for CogFlow's advantage. Since the preference labels used for automatic evaluation and reward-model training are partly generated through LLM-based ranking, the evaluators and the reward model may still inherit biases, such as length bias and stylistic bias. As a result, we view LLM-based evaluators and reward models as empirically validated but still imperfect proxies for human social judgment, rather than substitutes for direct human evaluation. Future work could further strengthen this evaluation framework by collecting larger-scale and more diverse human preference data, adding process-level annotations of cognitive flows, and training reward models directly from human judgments.

Text-only Formulation of Social Cognition Our current framework focuses on social situations described in text. However, real-world social cognition often relies on multimodal cues, including facial expressions, gaze, gestures, tone, timing, and surrounding visual context. Textual descriptions can only approximate these signals and may omit important non-verbal information. Therefore, our claims should be interpreted within the scope of text-based social reasoning, and extending cognitive reasoning to multimodal LLMs is an important direction for future work (Ong et al., 2026; Niu et al., 2025; Pandya et al., 2025; Park et al., 2025).

B Ethics Statement

This work aims to endow LLMs with the capability to perceive, interpret, and reason about complex social dynamics. While our primary goal is to enhance the transparency and effectiveness of AI in social decision-making, we recognize that this technology carries significant ethical implications and potential societal risks that must be carefully managed.

Data Sourcing and Privacy. The seed data for our research was sourced from public Reddit posts. To uphold the principle of respecting privacy and avoiding harm, we implemented a strict data processing pipeline. This pipeline included (1) the complete anonymization of all posts, removing any personally identifiable information, usernames, or sensitive content, and (2) the application of rigorous safety filters as described by (Kim et al., 2023) to eliminate potentially harmful or offensive content. Following Kim et al. (Kim et al., 2023), we applied a rigorous multi-stage process: *a) Canary Filter:* First, we employed Canary, a safety classification model, to remove social situations marked as needing caution (e.g., those involving crime). *b) LLM Filter:* Next, we prompted DeepSeek-R1 to identify and remove any remaining social situations that exhibited unsafe content or tendencies. *c) Expert Review:* Finally, during the data validation process, 6 domain experts were also tasked with reviewing the safety of the social situations, which resulted in a 100% pass rate. The resulting dataset consists of distilled, non-personal social situations intended solely for academic research.

Human Participant Engagement. Our study received approval from the Institutional Review Board (IRB) and involved human participation in several evaluation stages: 6 domain experts (different individuals with a master's degree or higher) for data validation, 6 annotators for evaluating reasoning chains, and 20 volunteers for a cognitive intervention trial. The 6 domain experts were recruited from the psychology department of a top-tier university in the country. The 6 annotators and 20 volunteers were recruited through a crowdsourcing platform. We balanced the demographics (e.g., gender) of the experts, annotators, and volunteers, and their ages ranged from 18 to 35. All of them were familiar with the concept of LLMs but were not otherwise involved in this research project. To prevent bias, participants were not informed of the specific experimental design, such as the source of the different example instances. They were only informed of the nature of their task (i.e., "evaluate the quality of reasoning" or "make a decision in a social situation"). For all human-involved experiments, we adhered to the following: (1) All participants were fully informed about the nature and purpose of the study, the type of tasks they would perform, and how their data would be used. (2) All data collected from participants were anonymized to protect their privacy. (3) All

participants were fairly compensated for their time and contribution based on the market price. (4) All participants were given full autonomy to exit the experiments at any time without any penalty.

Societal Benefits and Risks. The proposed framework, CogFlow, demonstrates the potential to augment human social intelligence, as evidenced by our intervention trials where structured cognitive reasoning assisted human participants in making more accurate social judgments. Applications of this research could extend to social skills training, educational tools for individuals with social cognitive deficits, and more empathetic human-AI collaborative systems. However, the ability to model social cognition also introduces dual-use risks. A model capable of analyzing social vulnerabilities and planning persuasive behaviors (e.g., via cognitive units) could theoretically be misused for social manipulation or to generate convincing misinformation. Furthermore, there is a risk of over-reliance/automation bias, where users might defer to the model's judgment in high-stakes interpersonal or professional situations. We explicitly state that CogFlow is a foundational research artifact. It is not designed to replace professional human judgment, nor is it suitable for use in therapeutic, crisis intervention, or high-stakes negotiation contexts without expert oversight.

C Reproducibility Statement

To ensure the reproducibility of our findings, we will release our full implementation of the CogFlow framework, which includes full data and code for data collection, SFT, and RL. The prompts used for data generation are provided in Appendix G. Crucial hyperparameters for training our models and the baselines are documented in Appendix F.2. The complete source code and model checkpoints will be made publicly available. We provide a temporary anonymized repository at `https://anonymous.4open.science/r/CogFlow-iclr2027/`.

D AI Use Statement

During paper writing, we used LLMs as an assistive tool to enhance the quality of the presentation. We employed LLMs to provide suggestions for grammatical corrections and polishing of the manuscript. The core ideas, scientific arguments, and the overall structure of the paper were developed exclusively by the authors. All suggestions generated by LLMs were carefully reviewed, edited, and approved by the authors to ensure they accurately reflect our meaning. The authors take full responsibility for all content presented in this paper, including any parts that were refined with the assistance of an LLM.

E Group Relative Policy Optimization (GRPO)

We adopt the GRPO (Shao et al., 2024) algorithm to optimize π_θ with respect to our reward function $\mathcal{R}$. Let $\pi_{\theta_{old}}$ denote the behavior policy from the previous iteration. For each input x, GRPO samples a group of cognitive flows $G = \{o_1, o_2, \cdots, o_N\}$, where each flow $o_i = (\tau_i, y_i)$. It then computes a relative advantage for each flow by normalizing its reward:

$$A_i = \frac{\mathcal{R}(o_i|x) - \underset{o_j \in G}{\text{mean}}(\mathcal{R}(o_j|x))}{\underset{o_j \in G}{\text{std}}(\mathcal{R}(o_j|x))} \tag{7}$$

GRPO then optimizes the following objective (denote $p_{i,j} = \frac{\pi_\theta(o_{i,j}|x, o_{i,<j})}{\pi_{\theta_{old}}(o_{i,j}|x, o_{i,<j})}$):

$$\mathbb{E}_{G \sim \pi_{\theta_{old}}} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{|o_i|} \sum_{j=1}^{|o_i|} \left(\min\left(p_{i,j} A_i, \text{clip}_\epsilon\left(p_{i,j}\right) A_i \right) - \beta \mathrm{D}_{KL}[\pi_\theta || \pi_{ref}] \right) \right] \tag{8}$$

F Experiments

F.1 More Details of Our Dataset

Detailed Information of Reddit Data We use anonymized Reddit[1] posts as our seed situations. The subreddits we used are as follows: FriendshipAdvice, LifeAdvice, Advice, AskWomenOver30, emotionalsupport, family, relationship_advice, confessions, socialskills, AmItheAsshole, AskMenOver30, AskMen, DecidingToBeBetter, mentalhealth, Anxiety, AskWomen, SocialEngineering, and familyadvice.

Distribution of Social Situations The distribution of social situation categories in our dataset is listed in Table 4.

Table 4: Distribution of social situations. Percentages represent the proportion of the total dataset.

Category / Subcategory	Count	Percentage
Romance	**819**	**16.06%**
Dating & Courtship	108	2.12%
Romantic Challenges	620	12.16%
Long-term Partnership	91	1.78%
Family	**1,414**	**27.73%**
Parent-Child Interaction	351	6.88%
Major Family Events & Issues	663	13.00%
Extended Family Relations	295	5.78%
Household & Logistics	105	2.06%
Public	**690**	**13.53%**
Stranger Encounters	184	3.61%
Community Life	422	8.27%
Service Interactions	84	1.65%
Friendship	**1,265**	**24.80%**
Intimate Friendship	723	14.18%
Group Activities & Events	383	7.51%
Casual Hangouts	159	3.12%
Professional	**912**	**17.88%**
Professional/Academic Challenges	396	7.76%
Professional Relationships	274	5.37%
Task-Oriented Collaboration	242	4.75%
Total	**5,100**	**100.00%**

Distribution of Test Set Difficulty Following experts' annotations, the test set instances were classified into three levels of difficulty: Easy (137), Medium (232), and Hard (131).

Details about Data-quality Validation and Data Annotation We detail the data-quality validation and data annotation for performance evaluation below. **(1) Data-quality validation** As described in §4.1, we recruited 6 domain experts to inspect 500 generated instances from our data. Their training and annotation procedure is as follows:

- **Training and qualifications**: Each expert first completed a training session of 50 pilot samples, learning to apply the criteria used in our automatic filtering (coherence, interpretability, and predictability) to judge cognitive flow quality. We provided feedback to help them calibrate the annotation criteria. They then took a qualification test of 50 samples, and only experts whose judgments aligned closely with our gold-standard annotations proceeded to the main task.
- **Annotation task**: The task was to provide a pass/fail judgment on whether a cognitive flow satisfied all three criteria. Agreement protocol: We divided six experts into two groups, each with 250 instances, with each instance judged by three experts. The pass/fail label of a cognitive flow was determined by majority vote among the three experts.

(2) Data annotation for performance evaluation. In our expert pairwise comparison (§4.2), the same six experts evaluated 500 seed instances, each with reasoning chains and responses from four models (CogFlow, Simulated-CogFlow, Distilled-R1-GRPO$_{R_{Len}}$, DeepSeek-R1). Pairwise combinations yield six pairs per instance (3,000 pairs total). The experts were again split into two groups of 250 instances each (1,500 pairs per group), and each pair received a win/tie/loss preference judgment. The final preference label per pair was set by majority vote and then converted into continuous strength scores via the Bradley-Terry model (Appendix F.6). In the same pass, each instance was also labeled as easy/medium/hard. To keep this judgment consistent across experts rather than purely subjective, they rated situational complexity against three observable cues: the number of interacting characters, the subtlety of the social signals to be interpreted (explicit vs. implicit), and the depth of inference required to reach a prediction, and assigned a level accordingly. The final difficulty label was also set by majority vote. Finally, these preference and difficulty labels drive the subsequent human study and automatic evaluation.

F.2 Implementation Details of our Models and Baselines

Experimental Setup All experiments were conducted on 8x NVIDIA H20 GPUs, using Llama-3.1-8B-Instruct (Meta, 2024) and Qwen-2.5-7B-Instruct (Yang et al., 2024) as base models. For the training pipeline, we employed the LLaMA-Factory framework[2] (Zheng et al., 2024) for the SFT and the veRL[3](Sheng et al., 2024) engine for RL.

[1]https://www.reddit.com

[2]https://github.com/hiyouga/LLaMA-Factory

[3]https://github.com/volcengine/verl

Details about Training Hyperparameters For the **SFT** stage, the model was trained for 2 epochs with a batch size of 8, a learning rate of 5×10^{-5}, and a context length of 8,192. The **preference reward model**, based on Qwen-2.5-7B-Instruct, was augmented with a linear classifier head to predict reward scores. It was trained for 1 epoch with a batch size of 8 and other hyperparameter settings the same as SFT. We split 26,676 candidate-reference pairs (§4.1) into train/validation/test sets at an 8:1:1 ratio, and the preference reward model achieved 0.85 test accuracy. During the **RL** stage, we generated trajectories by performing 6 rollouts for each of the 24 social situations in a training batch. The policy was subsequently updated using a mini-batch size of 4 situations, resulting in 6 gradient updates per collection batch. We use a learning rate of 10^{-6}, and a KL divergence coefficient to 10^{-3}. By default, we set $\omega_1 = 1, \omega_2 = 0.05$, and $\omega_3 = 0.1$.

To obtain the reward weights, in practice, we treat the Response Quality reward R_{Res} as the primary anchor ($w_1 = 1$) and tune the structural constraints accordingly. We performed a grid search over 0.01, 0.05, 0.1, 0.2 for the remaining weights, and selected $w_2 = 0.05$ for Cognitive Diversity R_{Div} and $w_3 = 0.1$ for Reasoning Length R_{Len}. Our sensitivity analysis shows: 1) with w_2 fixed, reducing w_3 makes reasoning substantially longer and harder to control; 2) with w_3 fixed, increasing w_2 inflates transition pattern diversity and makes the flow more fragmented and less aligned with normal human reasoning. This shows that w_2 balances flexible exploration and logical coherence, while w_3 controls reasoning length and prevents rumination.

Selection of Tuning-free Reasoning Models. Our main tuning-free baselines include OpenAI o3, GPT-5.6-Sol, GLM-5.1, Qwen3-32B, DeepSeek-R1, Gemini-2.5-pro/flash, and Gemini-3.8-flash. A practical consideration is that reasoning visibility varies substantially across recent models and API providers: some models expose explicit or quantifiable reasoning information, whereas others return only the final response or a summarized/opaque representation of their internal reasoning. We therefore include recent models whenever possible for response-quality comparison, while emphasizing models with accessible reasoning outputs or reasoning-token statistics when analyzing reasoning behavior and efficiency. Accordingly, reasoning length is reported only when a comparable token count is available from the API; otherwise, it is marked as unavailable.

Training Details about CogFlow Models Our primary models are fine-tuned from the base model using cognitive flow.

- **CogFlow-SFT**: The base model fine-tuned on our SFT dataset.
- **CogFlow**: The final model, fine-tuned from CogFlow-SFT using our complete RL reward function.
- **CogFlow-GRPO**: An ablation of our method, fine-tuned from CogFlow-SFT using only the response quality reward ($\omega_1 = 1, \omega_2 = 0, \omega_3 = 0$). This is equivalent to the GRPO algorithm.
- **CogFlow (w/o $\mathcal{R}_{\text{Div}}$)**: An ablation fine-tuned from CogFlow-SFT, excluding the diversity reward component ($\omega_1 = 1, \omega_2 = 0, \omega_3 = 0.1$).
- **CogFlow (w/o $\mathcal{R}_{\text{Len}}$)**: An ablation fine-tuned from CogFlow-SFT, excluding the length penalty component ($\omega_1 = 1, \omega_2 = 0.05, \omega_3 = 0$).

Tuning-free LLMs For LLMs that do not have a native long chain-of-thought ability, such as GPT-4o and DeepSeek-V3, we employed a zero-shot Chain-of-Thought (CoT) prompting strategy to elicit step-by-step reasoning. Specifically, we appended the following instruction to the end of each input prompt: `Let's think step by step, and use <FINAL RESPONSE> before you give the final answer.` Evaluation results about more LLMs are provided in Table 5.

Training Details about "Direct-" Model Series These models are trained to generate the final response directly, without any explicit reasoning process.

- **Direct-SFT**: It is fine-tuned from the base model using the CogFlow SFT dataset, but with the reasoning process removed.
- **Direct-GRPO**: Fine-tuned from Direct-SFT using GRPO. It only used response quality reward $\mathcal{R}_{\text{Res}}$.

Training Details about "Distilled-R1-" Model Series These models are designed to emulate the R1-style reasoning format, effectively serving as distilled versions of R1.

- **Distilled-R1-SFT**: Fine-tuned from the base model using all the social situations of CogFlow's SFT data, but DeepSeek-R1 directly generates the reasonings and responses.

- **Distilled-R1-GRPO**: Fine-tuned from Distilled-R1-SFT using GRPO. The reward function combines response quality with a format-checking reward, $\mathcal{R}'_{\text{Format}}$, which verifies the presence of `<think>` and `</think>` tags. The total reward is:

$$\mathcal{R} = \mathcal{R}'_{\text{Format}} \cdot \mathcal{R}_{\text{Res}} \tag{9}$$

- **Distilled-R1-GRPO**$_{\mathcal{R}_{\text{Len}}}$: This configuration is identical to Distilled-R1-GRPO but incorporates our reasoning length reward, $\mathcal{R}_{\text{Len}}$, to encourage more concise reasoning paths. Total reward is:

$$\mathcal{R} = \mathcal{R}'_{\text{Format}} \cdot (\omega_1 \cdot \mathcal{R}_{\text{Res}} + \omega_3 \cdot \mathcal{R}_{\text{Len}}) \tag{10}$$

F.3 Implementation Details of LLM Evaluators

We detail the LLM-based and reward model-based evaluators used in Tables 1, 2, and 5 below. Due to the high evaluation cost of Claude-Sonnet-5, we omit it from Table 5. As shown in Tables 1 and 2, its evaluations are consistent with human judgment and exhibit trends similar to the other two evaluators.

Prompt-Based Direct Scoring Evaluators These evaluators generate a direct score for each response individually. They use R1 (Guo et al., 2025) (Score-R1), Qwen3-32B (Yang et al., 2025) (Score-Q32B) and Claude-Sonnet-5 (Score-Claude-Sonnet-5) as the evaluators, both prompted with the template from Prompt 1.

Comparative Preference Ranking Evaluators These evaluators generate scores for a batch of responses simultaneously using comparative ranking methods. We apply two distinct methodologies to R1(Guo et al., 2025), Qwen3-32B(Yang et al., 2025), and Claude-Sonnet-5 models:

- **CPRank**[1]: This is a direct comparison method. The evaluator is prompted (using Prompt 10) to rank all responses within a given batch from best to worst in a single pass, thereby establishing a complete preference order at once. It results in CPRank[1]-R1, CPRank[1]-Q32B, and CPRank[1]-Claude-Sonnet-5.
- **CPRank**[2]: This method, as described in the main text, involves two steps to refine the evaluation. First, for **initial ranking**, the model generates situation-specific criteria and uses them to assign an initial score and critique for each response. Second, for **comparative reranking**, it selects the median-ranked response as an anchor to mitigate scoring biases (e.g., positional bias). The model then performs a final comparative reranking of the entire pool against this anchor, yielding a more robust preference order. It results in CPRank[2]-R1, CPRank[2]-Q32B, and CPRank[2]-Claude-Sonnet-5.
- **RM**$_\phi$: This evaluator is the reward model specifically trained to score responses described in 3.3. The implementation details are described in F.2. During RL training, we observed two main failure modes of the reward model: length bias, favoring longer responses, and formatting bias, favoring responses with more bold markers. To reduce reward hacking, we designed a reasoning-length reward and masked paired bold markers during RM scoring.

F.4 Detailed Information of Models Used in Human Evaluation

We detail the models mentioned in the pairwise comparison here:

- **CogFlow**: Llama-3.1-8B-Instruct fine-tuned using the whole CogFlow pipeline.
- **Distilled-R1-GRPO**$_{\mathcal{R}_{\text{Len}}}$: Llama-3.1-8B-Instruct fine-tuned using the Distilled-R1-GRPO$_{\mathcal{R}_{\text{Len}}}$ method.
- **Simulated-CogFlow**: The pruned results of cognitive flow simulation (crafted by prompting R1).
- **DeepSeek-R1**: The native DeepSeek-R1 model.

F.5 More Baseline Model Performance

More baseline results (GPT-4o, DeepSeek-V3, Qwen-2.5-7B/72B-Instruct, and LLama-3.1-8B/70B-Instruct using the CoT strategy stated in F.2 are shown in Table 5. We do not include the concurrent work Social-R1 (Wu et al., 2026) as a baseline, as its code and data have not been publicly released at the time of submission, preventing a controlled and reproducible comparison under our experimental protocol. While both CogFlow and Social-R1 target process-level social reasoning, they differ

```
**[Task]**
Given a [User Input] and its corresponding [Answer], please provide a comprehensive score
between 1 and 10 based on the quality of the answer, where a higher score indicates a better
answer. A score of 5 indicates that the answer is basically correct but may be incomplete,
unclear, or partially inaccurate.

Scoring Criteria Explanation (for reference; please make a comprehensive judgment):

- 10: Perfect answer. Entirely accurate, informative, well-structured, and appropriately
worded. Effectively addresses the user's query, potentially even exceeding expectations.
- 8-9: Excellent answer. Accurate and complete in information, logically clear, fluently
expressed, fully satisfying the user's needs.
- 6-7: Good answer. Basically correct and relevant, but may lack depth in certain details or
contain minor inaccuracies.
- 5: Passable answer. Generally correct but potentially incomplete, somewhat unclear, or
containing individual errors that do not severely impact understanding.
- 3-4: Insufficient answer. Partially relevant but missing key information, containing
significant errors, or failing to address the core issue.
- 1-2: Poor answer. Severely off-topic, containing incorrect information, or entirely
unhelpful.

When evaluating, you may comprehensively consider the following dimensions (not all are
required):
- Accuracy: Whether the answer is factually correct and non-misleading.
- Completeness: Whether it covers the key points of the user's question.
- Relevance: Whether the answer stays closely aligned with the user's question without
deviating from the topic.
- Clarity: Whether the expression is clear, easy to understand, and well-organized.
- Practicality: Whether it offers practical help to the user and is actionable (if
applicable).

**[User Input]**
{user_input}

**[Answer]**
{answer}

**[Output Format]**
Output in JSON format, with the answer given one integer score in 1-10.
'''json
{{
    "score": evaluated score
}}
'''
```

Prompt 1: The prompt template for direct scoring response. {user_input}, {answer} are placeholders. {answer} is the answer to be evaluated. {user_input} is the social situation with question and format constraint.

substantially in reasoning representation, data construction pipeline, and training strategy, making direct numerical comparison less informative than the methodological distinction itself.

To further distinguish our training-based approach from advanced tuning-free strategies, we additionally introduced a **"Cognitive Reasoning Prompting" (CRP)** baseline in Table 5. Unlike standard CoT, CRP explicitly embeds the definitions of our six cognitive units and the logic of cognitive flow directly into the system prompt, instructing base models to mimic the reasoned social response without parameter updates. The specific prompt template used for CRP is provided in 2. It can be shown that our trained CogFlow model significantly outperforms these CRP-based baselines.

These results, combined with the main text analysis (where CogFlow models achieves parity with the strong teacher-prompted reference Simulated-CogFlow), suggest that training enables more effective acquisition of cognitive reasoning patterns than prompting alone.

F.6 Performance of Pairwise Comparison

We show the precise pairwise results from the experts' pairwise evaluation in Table 6.

To quantitatively assess the relative strength of our models from pairwise comparison data, we employ the Bradley-Terry (BT) model(Bradley & Terry, 1952), a statistical method for converting pairwise preferences into a continuous capability scale. The results are provided in Figure 3.

Table 5: Results of automatic evaluation for more tuning-free models with CoT strategy. "-" denotes cases where the number of reasoning tokens is unavailable from the API response. CPRank^2-Claude-Sonnet-5 is omitted due to its high evaluation cost.

Models	CPRank^2-R1 (↑)				CPRank^2-Q32B (↑)				Reasoning Length (tokens, ↓)
	Overall	Easy	Medium	Hard	Overall	Easy	Medium	Hard	
Tuning-free Models (CoT)									
Llama-3.1-70B (CoT)	0.0619	0.1016	0.0577	0.0264	0.1256	0.2016	0.1173	0.0607	231.33
Qwen-2.5-7B (CoT)	0.0820	0.1208	0.0849	0.0403	0.1340	0.2178	0.1255	0.0612	184.90
Llama-3.1-8B (CoT)	0.0885	0.1000	0.0882	0.0772	0.1679	0.2467	0.1575	0.1039	253.12
Qwen-2.5-72B (CoT)	0.1401	0.2010	0.1703	0.0465	0.1487	0.2588	0.1338	0.0598	242.75
DeepSeek-V3 (CoT)	0.2443	0.3077	0.2390	0.1862	0.3591	0.3854	0.3662	0.3192	927.40
GPT-4o (CoT)	0.2542	0.3366	0.2407	0.1876	0.3489	0.3861	0.3521	0.3035	918.55
Tuning-free Models (CRP)									
Llama-3.1-70B (CRP)	0.0933	0.1802	0.0733	0.0264	0.1786	0.3207	0.1509	0.0783	305.04
Qwen-2.5-72B (CRP)	0.1119	0.2340	0.0852	0.0167	0.2509	0.3969	0.2213	0.1506	279.67
DeepSeek-V3 (CRP)	0.1641	0.3070	0.1400	0.0451	0.3161	0.4198	0.3038	0.2294	269.75
o3 (CRP)	0.1822	0.3038	0.1469	0.0959	0.3049	0.4148	0.2834	0.2281	-
Qwen-3-32B (CRP)	0.2713	0.3258	0.2533	0.2347	0.5418	0.5828	0.5449	0.4935	671.04
gemini-2.5-pro (CRP)	0.3798	0.5528	0.3129	0.273	0.4922	0.6073	0.4677	0.4153	564.14
glm-4.5 (CRP)	0.4024	0.5052	0.3509	0.3392	0.6521	0.7337	0.6344	0.5844	770.18
DeepSeek-R1 (CRP)	0.5271	0.5705	0.5027	0.5032	0.6564	0.6912	0.6448	0.6351	995.53

```
You are an advanced AI assistant specializing in cognitive reasoning. Your task is to analyze
a complex situation and answer a related question.

You must use a specific "Cognitive Reasoning Framework" to structure your thought process.
This framework consists of six core cognitive units.

**[Cognitive Reasoning Framework Definitions]**
You must use the following six cognitive units (and their definitions) to build your
reasoning. Note that you may use these units in any order and as many times as necessary to
best support your reasoning:
- <Behavior>: Formulating an intention to act in response to social situations.
- <Observation>: Perceiving and interpreting events and others' behaviors to form an initial
cognitive appraisal of a situation.
- <Attribution>: Analyzing the causes of events or behaviors.
- <Motivation>: The expectations and value assessments of potential outcomes for self and
others' behavior, which provide the drive to act.
- <Regulation>: Reflecting on and adjusting emotions, beliefs, and behaviors in pursuit of
social goals.
- <Efficacy>: The belief to execute a specific social behavior, which influences motivational
intensity.

**[Strict Output Format]**
You must adhere strictly to the following output format.

- First, provide your complete thought process.
- The entire thought process must be enclosed in a single pair of <think>...</think> tags.
- Inside the <think> tags, each step of your reasoning must use one of the cognitive unit
tags defined above (e.g., <Observation>...</Observation> or <Attribution>...</Attribution>).
- You may use these tags in any order and as many times as necessary to best support your
reasoning.
- After the closing </think> tag, provide your "Final Response" on a new line, following the
requirements by the user.
```

Prompt 2: The system prompt for Cognitive Reasoning Prompting.

Objective Function The core assumption of the BT model is that each model i possesses an unobserved strength parameter $p_i \in \mathbb{R}_{>0}$. The probability of model i winning against model j is given by:

$$P(i \text{ defeats } j) = \frac{p_i}{p_i + p_j}. \tag{11}$$

Given the observed number of model i defeats model jW_{ij}, our objective is to find the set of strength parameters $p = \{p_1, p_2, ..., p_n\}$ that maximizes the log-likelihood of the observed outcomes. The

Table 6: Results of pairwise comparison. The three numbers are the percentage of *win/tie/loss* for the paired models.

Models	**Compared Models**	**Easy** (%)	**Medium** (%)	**Hard** (%)	**Overall** (%)
CogFlow vs.	Simulated-CogFlow	56.9 / 0.8 / 42.3	43.1 / 0.9 / 55.9	50.6 / 1.9 / 47.4	49.1 / 1.2 / 49.7
	DeepSeek-R1	55.5 / 0.0 / 44.5	49.7 / 1.0 / 49.2	52.2 / 1.6 / 46.2	51.9 / 1.0 / 47.0
	Distilled-R1-GRPO$_{\mathcal{R}_{\text{Len}}}$	64.6 / 4.4 / 31.0	54.7 / 3.0 / 42.3	62.6 / 3.7 / 33.7	59.9 / 3.6 / 36.5
Simulated-CogFlow vs.	DeepSeek-R1	47.1 / 4.3 / 48.6	55.2 / 1.7 / 43.1	54.9 / 4.7 / 40.4	52.9 / 3.6 / 43.6
	Distilled-R1-GRPO$_{\mathcal{R}_{\text{Len}}}$	61.1 / 1.6 / 37.3	56.1 / 1.7 / 42.2	54.8 / 1.3 / 43.9	56.9 / 1.6 / 41.5
DeepSeek-R1 vs.	Distilled-R1-GRPO$_{\mathcal{R}_{\text{Len}}}$	54.0 / 0.7 / 45.3	52.2 / 0.0 / 47.8	49.2 / 1.6 / 49.2	51.6 / 0.8 / 47.6

total log-likelihood function is:

$$\mathcal{L}(\mathbf{p}) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} W_{ij} \log \left(\frac{p_i}{p_i + p_j} \right) \tag{12}$$

$$= \sum_{i=1}^{n} \left(\left(\sum_{j \neq i} W_{ij} \right) \log(p_i) - \sum_{j \neq i} W_{ij} \log(p_i + p_j) \right). \tag{13}$$

Optimization Method Since a closed-form solution for maximizing this likelihood is not available, we use an iterative algorithm to find the Maximum Likelihood Estimates (MLE) for the parameters p. The update rule for each parameter p_i at each iteration is derived from the likelihood equations, resulting in the following fixed-point iteration scheme:

$$p_i^{(\text{new})} = \frac{\sum_{j \neq i} W_{ij}}{\sum_{j \neq i} \frac{W_{ij} + W_{ji}}{p_i^{(\text{old})} + p_j^{(\text{old})}}}. \tag{14}$$

The proof of its convergence and optimality is detailed in (Hunter, 2004). Here, we initialize all p_i to 1 and then perform iterative updates. Each update step first applies Eq. 14 for $i = 1, ..., n$, and then normalizes the resulting parameter vector $\left(p_1^{(\text{new})}, ..., p_n^{(\text{new})} \right)$. This process is repeated until the parameters converge, defined as when the L_2 norm of the parameter vector change is below a tolerance of 10^{-6}. Finally, to ensure a unique solution, the parameters are normalized such that the weakest model has a score of 1.

F.7 Prompts and Examples for Cognitive Intervention for Humans

We use R1 to translate reasoning chains into natural language interventions, following the Prompt 3. To illustrate this, Table 7 presents two intervention examples derived from two reasoning styles.

F.8 In-depth Analysis of CogFlow

Mechanism of Generalization As shown in Table 3, it may seem counterintuitive that social cognitive units improve the model's mathematical ability. However, our inspection of the outputs finds that CogFlow enforces a structured meta-cognitive scaffolding that standard CoT lacks. For example, `Observation` forces the model to explicitly parse and list all problem conditions before solving; `Attribution` helps identify the underlying mathematical principles or theorems required; `Regulation` acts as a self-correction mechanism, checking intermediate calculation steps to prevent the error propagation often seen in standard CoT. To further validate this transferability, we compare CogFlow-adapt against a Llama-3.1-8B model trained with standard RLVR on the same datasets, which achieves 74.7% accuracy on SocialIQA and 52.0% on MATH-500, both significantly lower than CogFlow-adapt ($p < 0.05$). These results collectively confirm that CogFlow functions as a domain-agnostic reasoning scaffold, generalizing effectively well beyond its original social reasoning domain.

Information Flow within Cognitive Flow To dissect the internal mechanism of cognitive reasoning, we visualize the information flow within our *CogFlow* during inference. We analyze attention

```
## **[Task]**
You are a thoughtful and persuasive mentor. Your friend encountered a task: [Question]
He has been provided with several responses, the best one is [best_response], and the rest
are [other_responses].
But he did not choose the best one.
Now, you plan to persuade his friend to reconsider. But you should be gentle, so you should
take a reasoning procedure [Best Reasoning] leading to the best response, and try to (1)
figure out based on the chosen response, what may not he considered in each step of the
reasoning procedure, and (2) try to teach him to think in better ways. You can guide him to
build up the reasoning procedure(You should assume that he is a beginner, and he may not know
the reasoning procedure. ), and make every node of the reasoning procedure better.

## **[Question]**
{question}

## **[other_responses]**
{other_responses}

## **[best_response]**
{best_response}

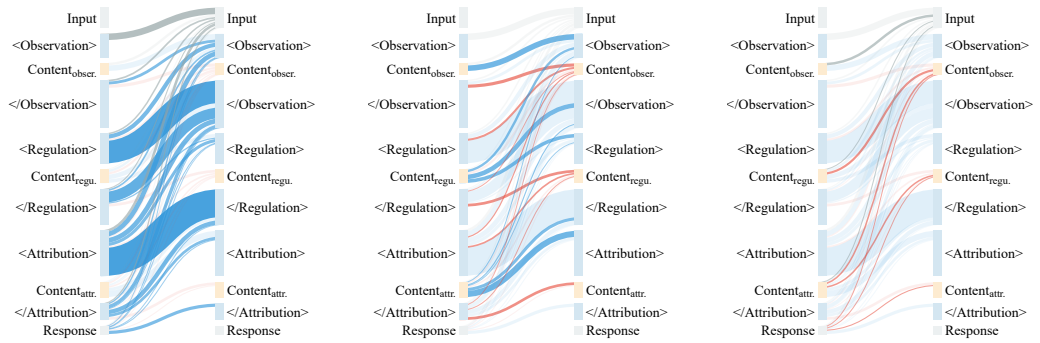
## **[Best Reasoning]**
{best_reasoning}

## **[Requirement]**
- YOU SHOULD NEVER MENTION OR HINT AT THE EXISTENCE OF THE [best_response].

## **[Output format]**

You should persuade him to reconsider. The persuasion should be concise. Please provide the
persuasion in JSON format like this:
```json
{{
    "words": "Your persuasion"
}}
```
```

Prompt 3: Prompt converting reasoning to intervention. {question}, {other_responses}, {best_responses}, {best_reasoning} are placeholders. {other_responses} is a list of all four candidate responses from CogFlow, Simulated-CogFlow, DeepSeek-R1, and Distilled-R1-GRPO$_{\mathcal{R}_{\text{Len}}}$. {question} is a social situation and question. {best_responses} and {best_reasoning} are the response content and reasoning of the ground truth model in the situation, respectively.



(a) *Unit-to-Unit* Flow (b) *Unit-to-Content* Flow (c) *Content-to-Content* Flow

Figure 6: Visualization of three information flow patterns (a-c) in performing cognitive reasoning. The patterns are composed of cognitive units (blue blocks) and their unit content (yellow blocks). The connecting line width reflects the degree of influence from the right block to the left.

patterns between four logical blocks: initial **input**, **cognitive unit tokens** (e.g., ⟨Observation⟩), **cognitive unit content** (the text generated for each unit), the final **response**. The attention weight between blocks is calculated by averaging the summed weights from all layers. For multi-token blocks (*input, content, response*), we mitigate dilution from non-essential tokens by averaging the top-10 attention weights within a block. The resulting weights (w) are then normalized ($w' = w^{0.2}$) to en-

Table 7: Example of guidance for intervention for two styles of reasoning, i.e., cognitive reasoning vs. R1 reasoning.

Type	Guidance
Cognitive Flow Style	Hi there! I noticed you analyzing that school scenario - great focus on the visible emotions. Let's build your analysis step by step together. First, we should collect every observable fact neutrally, like Leo would: Maya's bracelet-twisting, the folded note exchange, Chris's smirk-to-whisper sequence, her abrupt step back and near-trip, the library retreat, Chris's chuckle, and the bell ringing. Now, when interpreting, we might consider adding layers: Could the note's secrecy create uncertainty about its purpose? Might Chris's immediate chuckle after her exit amplify concerns about intent? And importantly, does Maya's established anxiety before the interaction suggest potential patterns? The time pressure from the bell could also heighten Leo's urgency. This method helps ensure we connect all details to concerns without speculation. Want to walk through how this changes your interpretation?
R1 Reaoning Chain Style	Hey, let's walk through this step-by-step together. First, consider Maya's nervousness before the interaction - her bracelet-twisting, darting eyes, and untied shoelace show she's already anxious. Many beginners might overlook how these details establish vulnerability. Next, Chris's smirk while reading the note and his whispered comment are crucial - they suggest negative intent rather than kindness. Maya's physical reaction of stepping back abruptly and nearly tripping reveals shock or fear, directly linking to what Chris said. Then note her hasty retreat to the library (a safe space) instead of staying social, showing deep distress. Chris's solitary chuckle afterward confirms he wasn't being supportive. Finally, Leo's stopped eating and wrinkled forehead prove he's connecting these dots in real-time. By building the analysis this way - starting with initial cues, then reactions, then consequences - we see how every detail points to Leo worrying about emotional harm to Maya. What stands out most to you in this sequence?

hance the visibility of all connections. For clarity, we separate this information flow into 3 patterns: *unit-to-unit*, *unit-to-content*, and *content-to-content*. A flow, shown in Figure 6, reveals: **(1) The structural unit-to-unit flow dominates the reasoning process**, with the highest attention weights among all patterns. **(2) Cognitive unit tokens actively steer content generation**. The unit-to-content flow shows a strong link between a unit token and its unit content (e.g., $\texttt{Content}_{\text{attribution}}$ to $\langle\texttt{Attribution}\rangle$), confirming that the unit tokens are not just placeholders but actively guide the generation of relevant thoughts. **(3) Reasoning exhibits a hierarchical "structure-first" attention flow.** When generating new content, the model consistently attends more to the structural tokens of previous steps than to the textual content of those steps, e.g., $\texttt{Content}_{\text{regulation}}$ attends more strongly to the $\langle/\texttt{Observation}\rangle$ token than to the $\texttt{Content}_{\text{observation}}$. This shows that the structured scaffold built by unit tokens is the primary driver of the model's cognitive flow generation.

G Prompts for Methodology

G.1 Prompts for Seed Data Collection

We used R1 (Guo et al., 2025) to generate the seed data. Each data instance consists of three components: a situation (describing the background and story), a question (based on the situation), and format constraints (specifying the required output format). First, we used the Prompt 4 to generate the situation and question. Next, we used the Prompt 5 to generate the format constraint. Finally, we used the Prompt 6 to validate the quality of the resulting data instance. If it met our

```
## **[Task]**

Given a scenario description and suggestions related to the scenario, you are required to
generate a scenario and a question for the COGNITION TEST. You should just use the
description and suggestions as triggers; you can convert the scenario arbitrarily by yourself:

1. **Summarize the Scenario**:
    - Objective: Craft a scene of dynamic social interaction focusing on several with
motion-driven engagement. Describe the scenario using plain words.
    - It should focus on social interactions, with enough details, for example:
        - Environmental Context: Describe a specific time/place.
        - Specific Task: Clearly state efficient information. For example, the problem they
are facing or the activity they are doing.
        - Character Relationship: Clearly state the relationship between the roles.
        - Character Dynamics: Establish clear profiles of the characters.
    - IMPORTANT: The scenario should be concise with enough details (not necessarily related
to the original scenario or the question). It's better to include either relevant or
irrelevant details to the question stated in the next step.

2. **State the question**:
    - Concisely state the question based on the scenario in one sentence using the third
person perspective. But you should only state the question simply, using words of mouth.
    - You should double-check that the answer is NOT stated in the scenario. There should
also be no direct/indirect hints in the scenario.
    - The question should be suitable for a cognitive test. You should ignore the original
question stated in [Scenario Description].

3. **Output Format**:
    Present your result in JSON format using the following structure and respond in English.
You should only use simpler vocabulary at a high school level to form your answers. Make sure
that quotes inside all strings are escaped with backslashes:
'''json
{{
    "scenario": "Scenario Summary",
    "question": "Question in one sentence"
}}
'''

## **#Possible Tests#**
These examples are way too brief and easy; your output should be more detailed and harder.
For example, you should not give any hints, and it had better be open-ended.
'''json
{examples}
'''

## **[Scenario Description]**
{scenario_description}

## **[Suggestion]**
{suggestion}
```

Prompt 4: The prompt template for social situation and question generation. {examples}, {suggestion}, {scenario_description} are placeholders. {examples} is the examples randomly sampled from a manually crafted set, {suggestion} are the comments of the Reddit post, {scenario_description} is the original Reddit post.

quality standards, it was stored; otherwise, it was discarded, and the process was repeated from the beginning.

G.2 Prompts for Cognitive Flow Simulation

We used R1 to perform the following tasks:

- **Planning**: LLMs are used in the following two operations: **Generation**: To generate a cognitive unit's thought content, we use the Prompt 7. **Prediction**: The Prompt 8 to identify the most relevant subsequent units. The choice `Terminate` stands for the reasoning has reached a terminal state.
- **Completion**: We prompted the LLMs using the Prompt 9 to get the final response.

```
## **[Task]**

Given a [User Input] containing a Story and an open-ended Question, please propose an
appropriate constraint on the output format for the answer to the Question containing all
constraints in [Required Constraints]. The proposed constraint should be concise.

Note:
- The generated instructions cannot contain any content related to or hinting at the answer.
- The output format should follow [Output Format].

## **[User Input]**
{user_input}

## **[Required Constraints]**
{required_constraints}

## **[Output Format]**
You should directly output the instruction as natural sentences without any additional words
or explanations (especially explain how you generate the output). In one word, your whole
output can be directly used as a constraint.
```

Prompt 5: Prompt template for constraint generation. {user_input} is the generated situation and question. {required_constraints} is randomly sampled from the options in Table 8.

Table 8: Prompts for generating output constraints.

Constraint Type	Prompt Template Options
Format	The output should be formatted in JSON / YAML / Markdown/ Bullet or any other suitable format. To state this constraint, you should choose one specific format and give a brief demonstration of the format of the output to make it clear. Note that your instructions should be concise. **IMPORTANT: You should make sure that your demonstration is just formal, without any hint of the real answer. **
Verbosity	• **High:** The output should be of high verbosity, which means detailed (but still needs to be concise). • **Medium:** The output should be of medium verbosity, which means balancing between brief and detailed. • **Low:** The output should be of low verbosity, which means brief and concise.

G.3 Prompts for Dual-Validation based Filtering

We used R1 to perform the following tasks:

- **Comparison Pool Construction**: We randomly gathered snippets from the generated cognitive flows, and reused the Prompt 9 to get the final response under the guidance of the reconstructed fake cognitive flow.
- **Two-stage Comparative Preference Ranking**: For both two stages in preference ranking, we use a unified Prompt 10, which can score the first response listed to be 5 as an anchor.

G.4 Prompts for Cognitive Flow Pruning

LLMs were used to evaluate the quality of cognitive flows with Prompt 11. We screened out those who scored below 4 in at least one category.

```
## **[Task]**

Please check the scenario, question, and constraint given in the [User Input], determine in
order whether the following conditions are met, and provide your response following the
output requirements in [Check Output Format].

1. Please check if the question is relevant to the scenario. For example, the content
involved must be mentioned in the scenario.
2. Please check if the constraint does not imply the answer to the question, but only
provides formatting content or restates the content of the question.
3. Please check if the constraint does not contain confusing content. The constraint must be
a reasonable format restriction for someone answering the question. For example, the act of
"requiring in the constraint not to imply the answer" does not meet the requirement.

## **[User Input]**

{user_input}

## **[Check Output Format]**

Please use JSON format for the output, with only one key named 'result', and the value being
a boolean type. true indicates that all requirements are met, and false indicates that at
least one requirement is not met. Please follow the structure below:

'''json
{{
        "result": true / false
}}
'''
```

Prompt 6: The prompt template for seed data quality validation. {user_input} is a placeholder for the social situation, question, and format constraints.

Table 9: Prompt templates for cognitive units.

Prompt Type	Prompt Template
Observation	- **Task**: Observe and interpret the specific behaviors, attitudes, or other information from the current context. The extracted facts must be precise and detailed without vague information. **NOTE: ALL THE INFORMATION MUST BE ALIGNED WITH THE CONTEXT. DO NOT MAKE UP FAKE INFORMATION.** - **Output**: State observation comprehensively.
Attribution	- **Task**: Attribute and evaluate the events or behaviors. It might include: Causal reasoning for others' actions / Impact assessment on current context / Further analysis and explanation. - **Output**: State the specific reason comprehensively.
Motivation	- **Task**: Generate motivation and goals, addressing the main problem discovered in other steps. - **Output**: State the goal and motivation.
Regulation	- **Task**: Check and adjust the previous thought to form a revised motivation or perception, or action plan. You should check (1) whether it lacks consideration, (2) whether other requirements need to be noticed. Think of the effect of the current plan or behavior, and check if there exists any risk. You should also check if there are any misunderstandings and be suspicious of the information in the analysis. - **Output**: Accurately and comprehensively state the problem and how to solve it.
Efficacy	- **Task**: Assess the internal perceptions, emotions, and beliefs of the actor of some behavior, and adjust the perception or action plan. - **Output**: State the efficacy and adjustment of action.
Behavior	- **Task**: Determine a more complete behavior based on the current environment and the analysis.

```
## **[Task]**
- Background: You are an assistant helping to answer problems. You need to carry out the next
step [{node_name}] of a reasoning chain to help respond to [User Input].
- Requirements:
   * follow the instructions in **Reasoning Step: [{node_name}]** which defines the reasoning
step.
   * It should be the next step of the half-finished reasoning chain in [Existing Analysis].
The result can only be not aligned with the analysis in [Existing Analysis] if you need to
fix mistakes or explore aspects not considered. You can refer to [Analysis Expectation] for
guidance, but you do not need to strictly follow it.
   * **Important**: Your output should be specific, without fake information.
   * **Important**: Your output should be comprehensive, including any possible aspects.
   * **Important**: Your output should only contain one step. If other things are in need,
state the need and reserve the reasoning for the next steps.
   * You should only use simpler vocabulary at a high school level to form your answers.

## **[User Input]**
{user_input}

## **[Existing Analysis]**
{previous_nodes}

## **[Analysis Expectation]**
{analyze_expect}

## **Reasoning Step: [{node_name}]**
{node_description}

## **[Output Format]**
Please output in English. The content should be a smooth and coherent paragraph, following
the format below:
'''json
{{
        "content": "the content of the required step"
}}
'''
```

Prompt 7: The prompt template for generating cognitive units in the cognitive flow simulation. {user_input}, {previous_nodes}, {analyze_expect}, {node_description}, {node_name} are placeholders. The {node_name} and {node_description} are prompts from Table 9. The {analyze_expect} is the justification for choosing this unit, which was generated during the unit selection.

```
## **[Task]**
- Background: You are an assistant helping with problems. You need to choose the next step of
a reasoning chain to help respond to the user's input in [User Input]. The chain should be
comprehensive.
- Requirements:
   * You should select **ALL** possible candidates from [Candidate Next Steps] that can be a
reasonable next ONE step of the half-finished reasoning chain provided in [Existing
Analysis]. You could visit the same step several times to get more information or analyze
further.
   * You should select **ALL** possible candidates from [Candidate Next Steps] that can be a
reasonable next ONE step of the half-finished reasoning chain provided in [Existing
Analysis]. You could visit the same step several times to get more information or analyze
further.
   * If analysis is sufficient for responding to the [User Input], and there are no concerns,
DIRECTLY select the [Terminate] step. (NOTE: If you are not certain or you think there might
be other potentials, you must choose other nodes along with Terminate. )
   * If you find some bad steps in [Existing Analysis] (for example: misinformation, unclear
statement, etc. ), redoing it again might refine it.
   * If more than one valid options exist, list the most applicable 2 or 3 steps, and put the
most applicable one in the first place.
   * The names of the next steps should be exactly the same as the name, e.g., Attribution
and Evaluation.
   * You should first review the prior steps in [Existing Analysis], and then determine the
candidates for the next step.


## **[User Input]**
{user_input}

## **[Existing Analysis]**
{previous_nodes}

## **[Candidate Next Steps]**

- **[Observation]**
   * Observe the specific behaviors or attitudes from the current context.

- **[Regulation]**
   * Validate and refine previous thoughts: (1) consider twice to polish the thought,
behavior, or motivation, (2) check if there exists more information in the scenario that
needs to be considered.

- **[Behavior]**
   * derive context-specific behaviors.

- **[Efficacy]**
   * analyze and adjust internal perceptions of the scene and action plan.

- **[Attribution]**
   * further interprets the result of previous steps, may include Causal reasoning for
others' actions, or Impact assessment on the current context.

- **[Terminate]**
   * Terminate analysis, synthesize final conclusion, and respond to the user.

- **[Motivation]**
   * formulate one's primary drivers of oneself, based on their needs/desires identified in
other steps.

## **[Output Format]**
'''json
{{
    "rationale": "Concise justification for selecting the next one step candidates, and
choose the most likely one",
    "next_step_candidates": ["step name", ...]
}}
'''
```

Prompt 8: The prompt template for choosing the next cognitive units in the cognitive flow simulation. {user_input}, {previous_nodes} are placeholders. {user_input} is the social situation, {previous_units} is a linear chain of existing cognitive units. LLMs are required to predict the units that directly follow the end of the sequence.

```
{user_input}
Please answer under the guidance of the following thought:
<|begin think|>
{previous_units}
<|end think|>
```

Prompt 9: The prompt template for generating a response under the guidance of cognitive flow. {user_input}, {previous_units} are placeholders. {user_input} is the social situation, {previous_units} is the simulated cognitive flow.

```
**[Task]**

Given the [User Input] and the corresponding multiple answers in [Answers] (which are in
random order), please score these answers on a scale of 1-10 (the higher the score, the
better) according to the following principles:

(1) Based on the [User Input] and all the answers in [Answers], propose evaluation criteria
that can assess the quality of the given answers. Ensure that under these criteria, the first
answer listed receives a score of 5.
(2) Explain the scoring principle for each score value sequentially.
(3) Score all the answers in [Answers] based on the established evaluation criteria. You must
use the first listed answer as a 5-point reference sample and provide a reason for each score.
(4) Refer to the [Output Format] for the output structure.

Note: You must ensure that the scores for the [Answers] are well-differentiated.

Special Attention: You must ensure that the first answer listed receives a score of 5 under
your scoring standard. Use this first answer as a baseline (referred to as "Baseline"). For
subsequent answers, a score greater than 5 must mean it is better than the first answer, and
a score less than 5 must mean it is worse than the first answer.

**[User Input]**
{user_input}

**[Answers]**
{answers}

**[Output Format]**
Output in JSON format, with every answer giving one score in 1-10. The answer is identified
by 'id'.
```json
{{
    "think": "analyze the user's input, come up with some criterion",
    "standard": [
        {{
            "score": 10,
            "standard": "standard of score 10"
        }},
        ...
        {{
            "score": 5,
            "standard": "standard of score 5"
        }},
        ...
        {{
            "score": 1,
            "standard": "standard of score 1"
        }},
    ],
    "result": [
        {{
            "id": ...,
            "reason": "compare with the first answer (Baseline), and then judge the quality
of this answer",
            "score": evaluated socre
        }},
        ...
        {{
            "id": ...,
            "reason": "compare with the first answer (Baseline), and then judge the quality
of this answer",
            "score": evaluated score
        }}
    ]
}}
```
```

Prompt 10: The prompt template for Two-Stage Comparative Preference Ranking. {user_input}, {answers} are placeholders. {answers} is a list of answers to be evaluated, each with a unique integer id. {user_input} is the social situation with question and format constraint.

```
**[Task]**

Please evaluate the cognitive flow provided in the [Reasoning Flow] based on the three core
criteria listed below. You need to score each criterion independently on a scale of 1-10 (the
higher the score, the better) and provide a reason for each score.

Evaluation Criteria:
- **Coherence**: Is it logically sound and free of internal contradictions?
- **Interpretability**: Does it clearly explain the social dynamics or core mechanisms
involved?
- **Predictability**: Does it offer reasonable insight into the future evolution of the
social dynamics?

Please strictly follow the JSON format required in the [Output Format].

**[Reasoning Flow]**
{reasoning_flow}

**[Output Format]**
Please output in JSON format. The JSON structure should include your thought process, the
independent scores, and reasons for each criterion.
'''json
{{
    "think": "evaluation process",
    "evaluation_result": {{
        "coherence": {{
            "reason": "Explain your reasoning",
            "score": evaluated_score
        }},
        "interpretability": {{
            "reason": "Explain your reasoning",
            "score": evaluated_score
        }},
        "predictability": {{
            "reason": "Explain your reasoning",
            "score": evaluated_score
        }}
    }}
}}
'''
```

Prompt 11: The prompt template for cognitive flow evaluation. {reasoning_flow} is a placeholder for the cognitive flow to be evaluated.