Self-Aware Speaking Mechanism for Multi-Agent Communication

Anonymous ACL submission

Abstract

To enhance work efficiency, researchers have increasingly turned to multi-agent frameworks that facilitate collaborative task completion by multiple agents. These frameworks necessitate sophisticated communication mechanisms to develop and execute effective strategies. Existing frameworks often rely on manually designed rules to determine the sequence of communication among agents. These rules are often suboptimal, requiring considerable human efforts for customization across diverse tasks, thereby limiting scalability. To address this issue, we propose "*self-aware speaking*" mechanism, designed to empower agent groups with the autonomy to dynamically determine their speaking order. This mechanism streamlines the management of communications within both centralized and decentralized multi-agent systems. Further, Empirical evaluations show that it outperforms rule-based mechanism by 5.2% on average, and is more scalable, underscoring its potential in multi-agent frameworks.

1 Introduction

Creating an intelligent agent that can seamlessly assist humans has been a primary objective of AI (Wooldridge and Jennings, 1995; Minsky, 1988; Bubeck et al., 2023). The advent of large language models (LLMs)(OpenAI, 2023b; Touvron et al., 2023; Jiang et al., 2023; Team et al., 2023) has positioned LLMs as pivotal entities in controlling language agents, substantially enriching their decision-making capabilities. To render agents more autonomous in task execution, recent research has infused autonomous agents with sophisticated cognitive mechanisms akin to human reasoning (Wei et al., 2022; Yao et al., 2023a), reflection (Yao et al., 2023c; Shinn et al., 2023), and tool utilization (Schick et al., 2023; Qin et al., 2023a,b; Qian et al., 2023b).

With the proliferation of applications leveraging various agents (Significant Gravitas; OpenAI, 2023a), the focus of research has broadened to include orchestrating groups of agents for collaborative engagement in complex tasks (Chen et al., 2023; Wu et al., 2023; Chan et al., 2023). Emerging studies underscore the benefits of facilitating communication and collaboration among diverse agents, which has been shown to enhance problem-solving outcomes significantly (Qian et al., 2023a; Hong et al., 2023). Yet, the intricacies of communication sequencing – determining the speaking order to ensure a cohesive conversation flow – remain underexplored, despite its acknowledged importance (Chen et al., 2023; Wu et al., 2023).

Due to the limited ability of LLMs in autonomous interaction[1], current multi-agent frameworks predominantly rely on simplistic, rule-based communication sequences, such as round-robin (Chen et al., 2023; Chan et al., 2023), which fail to account for the dynamic nature of dialogues and the contextual relevance of contributions. Also, designing rules for different tasks requires much human effort, and the rules are often suboptimal. This paradigm makes the multi-agent framework difficult to scale and limits its efficient generalization across diverse scenarios.

To mitigate these limitations, we introduce the "*self-aware speaking*" mechanism, enabling agents to autonomously determine their communication sequence within both *centralized* and *decentralized* multi-agent systems (Fig.1). This approach not only facilitates a more dynamic and adaptable interaction among agents but also demonstrates an average performance improvement of 5.2% across two challenging tasks. Additionally, our scalability experiments showcase the method's effectiveness compared to rule-based strategies, with performance either maintained or enhanced as the num-

[1]Recent research suggests that language agents still face limitations in their theory of mind (Kim et al., 2023). Consequently, genuinely understanding the intentions of another agent and providing useful information at the right time remains a significant challenge.

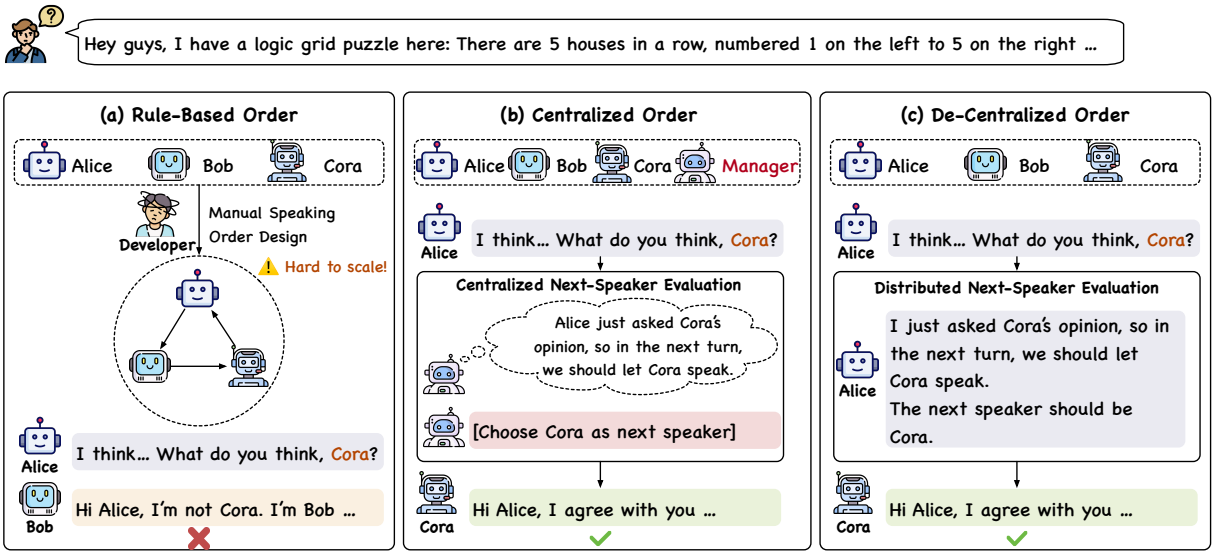

Figure 1: An illustration of different communication order designs.

ber of participating agents increases. Our analysis offers insights that may inform future agent communication mechanism designs aimed at creating more autonomous and efficient agent groups.

2 Methods

This section details communication strategies we tested to explore their effects on performances.

2.1 Rule-Based Strategy

The rule-based strategy is the strategy adopted by most of the previous work (Chen et al., 2023; Wu et al., 2023), where agents follow a pre-defined speaking order, e.g., round-robin. This approach ensures each agent has an equal opportunity to contribute, but lacks the flexibility to adapt to the evolving context of the discussion.

2.2 Centralized Strategy

Centralized systems have been widely adopted in the recent development of multi-agent systems, where coordination and decision-making are managed by a "manager" agent (Suzgun and Kalai, 2024). This centralized entity acts as the orchestrator of the collaborative process, delegating tasks among the agents. However, the centralized multi-agent system often lacks inter-agent communication, relegating agents, aside from the manager, to the role of mere executors. To mitigate the issue, we propose a self-aware speaking mechanism for the centralized context.

During the conversation, after each participating agent (referred to as a conv agent) contributes to the conversation, the manager agent synthesizes the conversation's progress and decides the next speaker. The decision-making process is guided by instructions provided to the manager, asking it to nominate the next speaker based on the current conversation state. The nominated agent is then tasked with continuing the conversation, contributing their insights and perspectives towards reaching a consensus solution.

2.3 Decentralized Strategy

In contrast to centralized systems, a more ideal design for a multi-agent system may be a decentralized system, where agents operate independently, without a central manager, and make decisions collaboratively through direct communication. This paradigm mirrors real-world distributed systems where agents, running on different devices, need to coordinate without centralized control. Our work explores the implementation of a dynamic speaking mechanism within this decentralized framework to enhance collaborative problem-solving.

The decentralized strategy contrasts with the centralized one by requiring each conv agent to not only contribute their analysis and solutions but also to select the next speaker, integrating this decision as part of their response. Each conv agent is instructed to consider the current state of the dialogue, including the contributions and perspectives shared by all members. After presenting their response, the conv agent selects the next speaker based on the whole conversation trajectory, ensuring that the choice is informed by the collective progress toward solving the problem. The name of the selected conv agent and the reason for the choice are then appended to its analysis or solution. Other agents receive the response and compare the

nominated agent's name with their name, the nominated agent then takes the next turn to speak.

An additional distinction between centralized and decentralized systems lies in the transparency of speaker selection rationale. In a decentralized framework, since there is no central manager to dictate the conversation flow, the reasoning behind choosing a specific agent as the next speaker is also communicated to all participants as part of the dialogue, which means that the agent's analysis and proposed solutions in each turn will be accompanied by the explanation of next speaker choosing (Fig. 1(c)), so that each agent can check whether the next speaker is itself. This contrasts with centralized systems, where the manager agent selects the next speaker without intervening in the conversation content.

2.4 Scoring Mechanism

To refine the next-speaker decision in both centralized and decentralized strategies, we additionally introduce a scoring mechanism aimed at enhancing the effectiveness of agent selection. This mechanism is designed to evaluate and select the next speaker based on a set of predefined criteria, ensuring that the conversation progresses more efficiently towards problem-solving.

Specifically, instead of directly writing down the next speaker's name, the agent who is deciding the next speaker is first tasked with summarizing the current discussion state and scoring each conv agent's contribution. The scoring criteria include (1) Transition Cues, (2) Relevance of Insights, and (3) Value of Insights. Transition Cues are awarded based on whether an agent's contribution smoothly guides the conversation to the next relevant point. Relevance of Insights assesses how pertinent an agent's input is to the ongoing discussion, and Value of Insights evaluates the potential of an agent's contribution to push the conversation forward. The agent synthesizes these scores into a comprehensive evaluation, ultimately selecting the next speaker based on the highest total score.

3 Experiments

3.1 Experimental Settings

Communication Strategies. Our experimental framework is designed to evaluate the impact of communication strategies on the efficacy of collaborative problem-solving among language agents. We simulate a scenario where n agents are tasked with discussing a given problem until they reach a consensus. In the prompt, we specify that to signal the conclusion of the discussion, an agent should output a specific identifier (e.g., *Answer: x*) indicating the consensus answer. Following this, an additional agent, designated as the answer provider, synthesizes the final answer and rationale from the discussion history. This setup allows for a detailed examination of how different communication strategies affect group problem-solving performance. We implement and compare the following five communication settings:

- **Baseline**: Agents speak in a pre-defined round-robin order.
- **Centralized**: A manager agent is responsible for selecting the next speaking agent based on the communication history.
- **Centralized**$_{\text{score}}$: Extends the Centralized setting by incorporating a scoring mechanism. The manager agent evaluates each participant's contribution and selects the next speaker based on these scores.
- **Decentralized**: The speaking agent nominates the next speaker at the end of their response.
- **Decentralized**$_{\text{score}}$: Extends the Centralized setting by incorporating a scoring mechanism. The speaker should explicitly give scores to other participants, and then select the best one.

All the prompts have been listed in Appendix A. We use GPT-4-0613 in our experiments.

Datasets. To challenge the agents with tasks requiring advanced logical reasoning and collaborative effort, we select two datasets known for their complexity: Logic Grid Puzzles (Srivastava et al., 2022) and Mini Crosswords (Yao et al., 2023b). The Logic Grid Puzzles dataset comprises 200 puzzles that demand agents to deduce the correct arrangement of elements in a grid based on a set of clues. The Mini Crosswords dataset consists of 156 crossword puzzles of a 5x5 grid size, where agents are required to collaboratively fill in the blanks with appropriate words.

3.2 Experimental Results

The experimental results presented in Table 1 highlight the efficacy of different communication order strategies in multi-agent collaborative problem-solving tasks across two challenging datasets. Notably, the Decentralized setting yields the highest accuracy in the Logic Grid Puzzles dataset at 65.0%, achieving an improvement of 8.5% over the

Dataset	Baseline	Centralized	Centralized$_{\text{score}}$	Decentralized	Decentralized$_{\text{score}}$
Logic Grid Puzzles	56.5	56.5	62.0	**65.0**	59.0
Mini Crosswords	13.9	**17.9**	15.6	16.0	14.5

Table 1: The accuracy of a group of 3 discussing agents on different datasets.

baseline setting. In the Mini Crosswords dataset, the Centralized method achieves the highest accuracy at 17.9%, outperforming the baseline setting by 4%. These superior performances suggest that allowing the agent to choose the next speaker in both centralized and decentralized systems can lead to effective communication patterns, likely because it enables agents to dynamically adjust the conversation flow based on the context of the discussion.

The introduction of scoring mechanisms (Centralized$_{\text{score}}$ and Decentralized$_{\text{score}}$) shows mixed effects on the performance. While the scoring mechanism improves the Centralized setting in the Logic Grid Puzzles, it does not achieve the same effect in the Mini Crosswords dataset. Also, it fails to improve the performance of the decentralized setting. This outcome suggests that the effectiveness of scoring may be context-dependent and does not always contribute to the overall performance. In the decentralized setting, since the scoring rationales will be also passed to other agents along with the problem analysis, it might potentially complicate their comprehension of the ongoing discussion, leading to a performance drop.

Overall, the findings highlight that under both centralized and decentralized settings, allowing agents to dynamically determine the communication order with our proposed self-aware speaking mechanism is a feasible and effective way to enhance collaborative problem-solving.

3.3 Scaling the Number of Agents

An integral part of evaluating the efficacy of different communication order strategies involves understanding how these approaches scale with an increasing number of agents. This subsection delves into the scaling properties of our proposed methods by analyzing their performance on the randomly sampled 50 instructions from Logic Grid Puzzles dataset as we increment the number of discussing agents from 3 to 6.

Figure 2 illustrates the scaling behavior of different settings. It becomes evident that the baseline setting, which employs a pre-defined round-robin order, shows a noticeable decline in accuracy as the number of agents increases. This is indicative

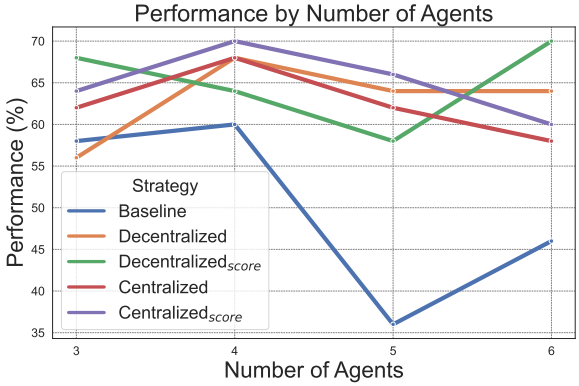

Figure 2: Performances of different team sizes.

of the limitations inherent in static communication orders. Conversely, both Centralized and Decentralized settings exhibit robust scalability, with relatively stable performance as the number of agents grows. Particularly, Decentralized$_{\text{score}}$ shows a significant increase in accuracy, peaking at 70% with 6 discussing agents, and Centralized retains a satisfactory performance level across different numbers of agents.

These findings suggest that dynamic communication strategies are crucial for scaling multi-agent collaborative problem-solving effectively. Systems that adapt to the evolving of the discussion can better manage the increased complexity brought on by more participants. This adaptability is key to maintaining and even enhancing performance in more crowded collaborative environments.

4 Conclusion

In this work, we address the challenge of designing communication order in collaborative problem-solving by language agents. We propose and evaluate the dynamic communication order decided by agents under both centralized and decentralized settings across two challenging reasoning tasks. Our results demonstrate that the choice of communication order significantly influences the efficacy of multi-agent collaboration, with decentralized approaches generally yielding better outcomes. This work contributes to the understanding of how dynamic communication strategies can enhance the performance of language agent groups, offering insights for the development of more efficient and adaptable multi-agent systems.

Limitations

While our proposed "*self-aware speaking*" mechanism marks a significant advancement in multi-agent communication compared with traditional rule-based communication, it is not without its limitations. As an initial exploration into fostering a more dynamic and adaptable communication mechanism, it exhibits only a stable level of effectiveness across varying agent numbers, contrary to the anticipated progressive increase. While it already surpasses the declining efficiency observed with rule-based approaches as the number of agents grows, our exploration underscores the necessity for ongoing research. Future endeavors are required to refine the communication sequencing strategy further, aiming to enhance the effectiveness and efficiency of multi-agent interactions substantially.

References

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott M. Lundberg, Harsha Nori, Hamid Palangi, Marco Túlio Ribeiro, and Yi Zhang. 2023. Sparks of artificial general intelligence: Early experiments with GPT-4. *CoRR*, abs/2303.12712.

Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2023. Chateval: Towards better llm-based evaluators through multi-agent debate.

Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, Yaxi Lu, Yi-Hsin Hung, Chen Qian, Yujia Qin, Xin Cong, Ruobing Xie, Zhiyuan Liu, Maosong Sun, and Jie Zhou. 2023. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors.

Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, and Chenglin Wu. 2023. Metagpt: Meta programming for multi-agent collaborative framework. *CoRR*, abs/2308.00352.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Hyunwoo Kim, Melanie Sclar, Xuhui Zhou, Ronan Le Bras, Gunhee Kim, Yejin Choi, and Maarten Sap. 2023. Fantom: A benchmark for stress-testing machine theory of mind in interactions.

Marvin Minsky. 1988. *The Society of Mind*. Simon & Schuster.

OpenAI. 2023a. Assistants api.

OpenAI. 2023b. GPT-4 technical report. *CoRR*, abs/2303.08774.

Chen Qian, Xin Cong, Wei Liu, Cheng Yang, Weize Chen, Yusheng Su, Yufan Dang, Jiahao Li, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. 2023a. Communicative agents for software development.

Cheng Qian, Chi Han, Yi R. Fung, Yujia Qin, Zhiyuan Liu, and Heng Ji. 2023b. CREATOR: disentangling abstract and concrete reasonings of large language models through tool creation. *CoRR*, abs/2305.14318.

Yujia Qin, Shengding Hu, Yankai Lin, Weize Chen, Ning Ding, Ganqu Cui, Zheni Zeng, Yufei Huang, Chaojun Xiao, Chi Han, Yi Ren Fung, Yusheng Su, Huadong Wang, Cheng Qian, Runchu Tian, Kunlun Zhu, Shihao Liang, Xingyu Shen, Bokai Xu, Zhen Zhang, Yining Ye, Bowen Li, Ziwei Tang, Jing Yi, Yuzhang Zhu, Zhenning Dai, Lan Yan, Xin Cong, Yaxi Lu, Weilin Zhao, Yuxiang Huang, Junxi Yan, Xu Han, Xian Sun, Dahai Li, Jason Phang, Cheng Yang, Tongshuang Wu, Heng Ji, Zhiyuan Liu, and Maosong Sun. 2023a. Tool learning with foundation models. *CoRR*, abs/2304.08354.

Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, et al. 2023b. Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789*.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2023. Toolformer: Language models can teach themselves to use tools. *CoRR*, abs/2302.04761.

Noah Shinn, Federico Cassano, Beck Labash, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning.

Significant Gravitas. AutoGPT.

Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska, Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W. Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hussain, Amanda Askell, Amanda Dsouza, Ameet Rahane, Anantharaman S. Iyer, Anders Andreassen, Andrea Santilli, Andreas Stuhlmüller, Andrew M. Dai, Andrew La, Andrew K. Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh Gupta, Anna Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholamidavoodi, Arfa Tabassum, Arul Menezes, Arun Kirubarajan,

Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia Efrat, Aykut Erdem, Ayla Karakas, and et al. 2022. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *CoRR*, abs/2206.04615.

Mirac Suzgun and Adam Tauman Kalai. 2024. Meta-prompting: Enhancing language models with task-agnostic scaffolding. *CoRR*, abs/2401.12954.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Slav Petrov, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, Orhan Firat, James Molloy, Michael Isard, Paul R. Barham, Tom Hennigan, Benjamin Lee, Fabio Viola, Malcolm Reynolds, Yuanzhong Xu, Ryan Doherty, Eli Collins, Clemens Meyer, Eliza Rutherford, Erica Moreira, Kareem Ayoub, Megha Goel, George Tucker, Enrique Piqueras, Maxim Krikun, Iain Barr, Nikolay Savinov, Ivo Danihelka, Becca Roelofs, Anaïs White, Anders Andreassen, Tamara von Glehn, Lakshman Yagati, Mehran Kazemi, Lucas Gonzalez, Misha Khalman, Jakub Sygnowski, Alexandre Frechette, Charlotte Smith, Laura Culp, Lev Proleev, Yi Luan, Xi Chen, James Lottes, Nathan Schucher, Federico Lebron, Alban Rrustemi, Natalie Clay, Phil Crone, Tomas Kocisky, Jeffrey Zhao, Bartek Perz, Dian Yu, Heidi Howard, Adam Bloniarz, Jack W. Rae, Han Lu, Laurent Sifre, Marcello Maggioni, Fred Alcober, Dan Garrette, Megan Barnes, Shantanu Thakoor, Jacob Austin, Gabriel Barth-Maron, William Wong, Rishabh Joshi, Rahma Chaabouni, Deeni Fatiha, Arun Ahuja, Ruibo Liu, Yunxuan Li, Sarah Cogan, Jeremy Chen, Chao Jia, Chenjie Gu, Qiao Zhang, Jordan Grimstad, Ale Jakse Hartman, Martin Chadwick, Gaurav Singh Tomar, Xavier Garcia, Evan Senter, Emanuel Taropa, Thanumalayan Sankaranarayana Pillai, Jacob Devlin, Michael Laskin, Diego de Las Casas, Dasha Valter, Connie Tao, Lorenzo Blanco, Adrià Puigdomènech Badia, David Reitter, Mianna Chen, Jenny Brennan, Clara Rivera, Sergey Brin, Shariq Iqbal, Gabriela Surita, Jane Labanowski, Abhi Rao, Stephanie Winkler, Emilio Parisotto, Yiming Gu, Kate Olszewska, Yujing Zhang, Ravi Addanki, Antoine Miech, Annie Louis, Laurent El Shafey, Denis Teplyashin, Geoff Brown, Elliot Catt, Nithya Attaluri, Jan Balaguer, Jackie Xiang, Pidong Wang, Zoe Ashwood, Anton Briukhov, Albert Webson, Sanjay Ganapathy, Smit Sanghavi, Ajay Kannan, Ming-Wei Chang, Axel Stjerngren, Josip Djolonga, Yuting Sun, Ankur Bapna, Matthew Aitchison, Pedram Pejman, Henryk Michalewski, Tianhe Yu, Cindy Wang, Juliette Love, Junwhan Ahn, Dawn Bloxwich, Kehang Han, Peter Humphreys, Thibault Sellam, James Bradbury, Varun Godbole, Sina Samangooei, Bogdan Damoc, Alex Kaskasoli, Sébastien M. R. Arnold, Vijay Vasudevan, Shubham Agrawal, Jason Riesa, Dmitry Lepikhin, Richard Tanburn, Srivatsan Srinivasan, Hyeontaek Lim, Sarah Hodkinson, Pranav Shyam, Johan Ferret, Steven Hand, Ankush Garg, Tom Le Paine, Jian Li, Yujia Li, Minh Giang, Alexander Neitz, Zaheer Abbas, Sarah York, Machel Reid, Elizabeth Cole, Aakanksha Chowdhery, Dipanjan Das, Dominika Rogozińska, Vitaly Nikolaev, Pablo Sprechmann, Zachary Nado, Lukas Zilka, Flavien Prost, Luheng He, Marianne Monteiro, Gaurav Mishra, Chris Welty, Josh Newlan, Dawei Jia, Miltiadis Allamanis, Clara Huiyi Hu, Raoul de Liedekerke, Justin Gilmer, Carl Saroufim, Shruti Rijhwani, Shaobo Hou, Disha Shrivastava, Anirudh Baddepudi, Alex Goldin, Adnan Ozturel, Albin Cassirer, Yunhan Xu, Daniel Sohn, Devendra Sachan, Reinald Kim Amplayo, Craig Swanson, Dessie Petrova, Shashi Narayan, Arthur Guez, Siddhartha Brahma, Jessica Landon, Miteyan Patel, Ruizhe Zhao, Kevin Villela, Luyu Wang, Wenhao Jia, Matthew Rahtz, Mai Giménez, Legg Yeung, Hanzhao Lin, James Keeling, Petko Georgiev, Diana Mincu, Boxi Wu, Salem Haykal, Rachel Saputro, Kiran Vodrahalli, James Qin, Zeynep Cankara, Abhanshu Sharma, Nick Fernando, Will Hawkins, Behnam Neyshabur, Solomon Kim, Adrian Hutter, Priyanka Agrawal, Alex Castro-Ros, George van den Driessche, Tao Wang, Fan Yang, Shuo yiin Chang, Paul Komarek, Ross McIlroy, Mario Lučić, Guodong Zhang, Wael Farhan, Michael Sharman, Paul Natsev, Paul Michel, Yong Cheng, Yamini Bansal, Siyuan Qiao, Kris Cao, Siamak Shakeri, Christina Butterfield, Justin Chung, Paul Kishan Rubenstein, Shivani Agrawal, Arthur Mensch, Kedar Soparkar, Karel Lenc, Timothy Chung, Aedan Pope, Loren Maggiore, Jackie Kay, Priya Jhakra, Shibo Wang, Joshua Maynez, Mary Phuong, Taylor Tobin, Andrea Tacchetti, Maja Trebacz, Kevin Robinson, Yash Katariya, Sebastian Riedel, Paige Bailey, Kefan Xiao, Nimesh Ghelani, Lora Aroyo, Ambrose Slone, Neil Houlsby, Xuehan Xiong, Zhen Yang, Elena Gribovskaya, Jonas Adler, Mateo Wirth, Lisa Lee, Music Li, Thais Kagohara, Jay Pavagadhi, Sophie Bridgers, Anna Bortsova, Sanjay Ghemawat, Zafarali Ahmed, Tianqi Liu, Richard Powell, Vijay Bolina, Mariko Iinuma, Polina Zablotskaia, James Besley, Da-Woon Chung, Timothy Dozat, Ramona Comanescu, Xiance Si, Jeremy Greer, Guolong Su, Martin Polacek, Raphaël Lopez Kaufman, Simon Tokumine, Hexiang Hu, Elena Buchatskaya, Yingjie Miao, Mohamed Elhawaty, Aditya Siddhant, Nenad Tomasev, Jinwei Xing, Christina Greer, Helen Miller, Shereen Ashraf, Aurko Roy, Zizhao Zhang, Ada Ma, Angelos Filos, Milos Besta, Rory Blevins, Ted Klimenko, Chih-Kuan Yeh, Soravit Changpinyo, Jiaqi Mu, Oscar Chang, Mantas Pajarskas, Carrie Muir, Vered Cohen, Charline Le Lan, Krishna Haridasan, Amit Marathe, Steven Hansen, Sholto Douglas, Rajkumar Samuel, Mingqiu Wang, Sophia Austin, Chang Lan, Jiepu Jiang, Justin Chiu, Jaime Alonso Lorenzo, Lars Lowe Sjösund, Sébastien Cevey, Zach Gleicher, Thi Avrahami, Anudhyan Boral, Hansa Srinivasan, Vittorio Selo, Rhys May, Konstantinos Aisopos, Léonard Hussenot, Livio Baldini Soares, Kate Baumli, Michael B. Chang, Adrià Recasens, Ben Caine, Alexander Pritzel, Filip Pavetic, Fabio Pardo, Anita Gergely, Justin Frye, Vinay

Ramasesh, Dan Horgan, Kartikeya Badola, Nora Kassner, Subhrajit Roy, Ethan Dyer, Víctor Campos, Alex Tomala, Yunhao Tang, Dalia El Badawy, Elspeth White, Basil Mustafa, Oran Lang, Abhishek Jindal, Sharad Vikram, Zhitao Gong, Sergi Caelles, Ross Hemsley, Gregory Thornton, Fangxiaoyu Feng, Wojciech Stokowiec, Ce Zheng, Phoebe Thacker, Çağlar Ünlü, Zhishuai Zhang, Mohammad Saleh, James Svensson, Max Bileschi, Piyush Patil, Ankesh Anand, Roman Ring, Katerina Tsihlas, Arpi Vezer, Marco Selvi, Toby Shevlane, Mikel Rodriguez, Tom Kwiatkowski, Samira Daruki, Keran Rong, Allan Dafoe, Nicholas FitzGerald, Keren Gu-Lemberg, Mina Khan, Lisa Anne Hendricks, Marie Pellat, Vladimir Feinberg, James Cobon-Kerr, Tara Sainath, Maribeth Rauh, Sayed Hadi Hashemi, Richard Ives, Yana Hasson, YaGuang Li, Eric Noland, Yuan Cao, Nathan Byrd, Le Hou, Qingze Wang, Thibault Sottiaux, Michela Paganini, Jean-Baptiste Lespiau, Alexandre Moufarek, Samer Hassan, Kaushik Shivakumar, Joost van Amersfoort, Amol Mandhane, Pratik Joshi, Anirudh Goyal, Matthew Tung, Andrew Brock, Hannah Sheahan, Vedant Misra, Cheng Li, Nemanja Rakićević, Mostafa Dehghani, Fangyu Liu, Sid Mittal, Junhyuk Oh, Seb Noury, Eren Sezener, Fantine Huot, Matthew Lamm, Nicola De Cao, Charlie Chen, Gamaleldin Elsayed, Ed Chi, Mahdis Mahdieh, Ian Tenney, Nan Hua, Ivan Petrychenko, Patrick Kane, Dylan Scandinaro, Rishub Jain, Jonathan Uesato, Romina Datta, Adam Sadovsky, Oskar Bunyan, Dominik Rabiej, Shimu Wu, John Zhang, Gautam Vasudevan, Edouard Leurent, Mahmoud Alnahlawi, Ionut Georgescu, Nan Wei, Ivy Zheng, Betty Chan, Pam G Rabinovitch, Piotr Stanczyk, Ye Zhang, David Steiner, Subhajit Naskar, Michael Azzam, Matthew Johnson, Adam Paszke, Chung-Cheng Chiu, Jaume Sanchez Elias, Afroz Mohiuddin, Faizan Muhammad, Jin Miao, Andrew Lee, Nino Vieillard, Sahitya Potluri, Jane Park, Elnaz Davoodi, Jiageng Zhang, Jeff Stanway, Drew Garmon, Abhijit Karmarkar, Zhe Dong, Jong Lee, Aviral Kumar, Luowei Zhou, Jonathan Evens, William Isaac, Zhe Chen, Johnson Jia, Anselm Levskaya, Zhenkai Zhu, Chris Gorgolewski, Peter Grabowski, Yu Mao, Alberto Magni, Kaisheng Yao, Javier Snaider, Norman Casagrande, Paul Suganthan, Evan Palmer, Geoffrey Irving, Edward Loper, Manaal Faruqui, Isha Arkatkar, Nanxin Chen, Izhak Shafran, Michael Fink, Alfonso Castaño, Irene Giannoumis, Wooyeol Kim, Mikołaj Rybiński, Ashwin Sreevatsa, Jennifer Prendki, David Soergel, Adrian Goedeckemeyer, Willi Gierke, Mohsen Jafari, Meenu Gaba, Jeremy Wiesner, Diana Gage Wright, Yawen Wei, Harsha Vashisht, Yana Kulizhskaya, Jay Hoover, Maigo Le, Lu Li, Chimezie Iwuanyanwu, Lu Liu, Kevin Ramirez, Andrey Khorlin, Albert Cui, Tian LIN, Marin Georgiev, Marcus Wu, Ricardo Aguilar, Keith Pallo, Abhishek Chakladar, Alena Repina, Xihui Wu, Tom van der Weide, Priya Ponnapalli, Caroline Kaplan, Jiri Simsa, Shuangfeng Li, Olivier Dousse, Fan Yang, Jeff Piper, Nathan Ie, Minnie Lui, Rama Pasumarthi, Nathan Lintz, Anitha Vijayakumar, Lam Nguyen Thiet, Daniel Andor, Pedro Valenzuela, Cosmin Paduraru, Daiyi Peng, Katherine Lee, Shuyuan Zhang, Somer Greene, Duc Dung Nguyen, Paula Kurylowicz, Sarmishta Velury, Sebastian Krause, Cassidy Hardin, Lucas Dixon, Lili Janzer, Kiam Choo, Ziqiang Feng, Biao Zhang, Achintya Singhal, Tejasi Latkar, Mingyang Zhang, Quoc Le, Elena Allica Abellan, Dayou Du, Dan McKinnon, Natasha Antropova, Tolga Bolukbasi, Orgad Keller, David Reid, Daniel Finchelstein, Maria Abi Raad, Remi Crocker, Peter Hawkins, Robert Dadashi, Colin Gaffney, Sid Lall, Ken Franko, Egor Filonov, Anna Bulanova, Rémi Leblond, Vikas Yadav, Shirley Chung, Harry Askham, Luis C. Cobo, Kelvin Xu, Felix Fischer, Jun Xu, Christina Sorokin, Chris Alberti, Chu-Cheng Lin, Colin Evans, Hao Zhou, Alek Dimitriev, Hannah Forbes, Dylan Banarse, Zora Tung, Jeremiah Liu, Mark Omernick, Colton Bishop, Chintu Kumar, Rachel Sterneck, Ryan Foley, Rohan Jain, Swaroop Mishra, Jiawei Xia, Taylor Bos, Geoffrey Cideron, Ehsan Amid, Francesco Piccinno, Xingyu Wang, Praseem Banzal, Petru Gurita, Hila Noga, Premal Shah, Daniel J. Mankowitz, Alex Polozov, Nate Kushman, Victoria Krakovna, Sasha Brown, MohammadHossein Bateni, Dennis Duan, Vlad Firoiu, Meghana Thotakuri, Tom Natan, Anhad Mohananey, Matthieu Geist, Sidharth Mudgal, Sertan Girgin, Hui Li, Jiayu Ye, Ofir Roval, Reiko Tojo, Michael Kwong, James Lee-Thorp, Christopher Yew, Quan Yuan, Sumit Bagri, Danila Sinopalnikov, Sabela Ramos, John Mellor, Abhishek Sharma, Aliaksei Severyn, Jonathan Lai, Kathy Wu, Heng-Tze Cheng, David Miller, Nicolas Sonnerat, Denis Vnukov, Rory Greig, Jennifer Beattie, Emily Caveness, Libin Bai, Julian Eisenschlos, Alex Korchemniy, Tomy Tsai, Mimi Jasarevic, Weize Kong, Phuong Dao, Zeyu Zheng, Frederick Liu, Fan Yang, Rui Zhu, Mark Geller, Tian Huey Teh, Jason Sanmiya, Evgeny Gladchenko, Nejc Trdin, Andrei Sozanschi, Daniel Toyama, Evan Rosen, Sasan Tavakkol, Linting Xue, Chen Elkind, Oliver Woodman, John Carpenter, George Papamakarios, Rupert Kemp, Sushant Kafle, Tanya Grunina, Rishika Sinha, Alice Talbert, Abhimanyu Goyal, Diane Wu, Denese Owusu-Afriyie, Cosmo Du, Chloe Thornton, Jordi Pont-Tuset, Pradyumna Narayana, Jing Li, Sabaer Fatehi, John Wieting, Omar Ajmeri, Benigno Uria, Tao Zhu, Yeongil Ko, Laura Knight, Amélie Héliou, Ning Niu, Shane Gu, Chenxi Pang, Dustin Tran, Yeqing Li, Nir Levine, Ariel Stolovich, Norbert Kalb, Rebeca Santamaria-Fernandez, Sonam Goenka, Wenny Yustalim, Robin Strudel, Ali Elqursh, Balaji Lakshminarayanan, Charlie Deck, Shyam Upadhyay, Hyo Lee, Mike Dusenberry, Zonglin Li, Xuezhi Wang, Kyle Levin, Raphael Hoffmann, Dan Holtmann-Rice, Olivier Bachem, Summer Yue, Sho Arora, Eric Malmi, Daniil Mirylenka, Qijun Tan, Christy Koh, Soheil Hassas Yeganeh, Siim Põder, Steven Zheng, Francesco Pongetti, Mukarram Tariq, Yanhua Sun, Lucian Ionita, Mojtaba Seyedhosseini, Pouya Tafti, Ragha Kotikalapudi, Zhiyu Liu, Anmol Gulati, Jasmine Liu, Xinyu Ye, Bart Chrzaszcz, Lily Wang, Nikhil Sethi, Tianrun Li, Ben Brown, Shreya Singh, Wei Fan, Aaron Parisi, Joe Stanton,

Chenkai Kuang, Vinod Koverkathu, Christopher A. Choquette-Choo, Yunjie Li, TJ Lu, Abe Ittycheriah, Prakash Shroff, Pei Sun, Mani Varadarajan, Sanaz Bahargam, Rob Willoughby, David Gaddy, Ishita Dasgupta, Guillaume Desjardins, Marco Cornero, Brona Robenek, Bhavishya Mittal, Ben Albrecht, Ashish Shenoy, Fedor Moiseev, Henrik Jacobsson, Alireza Ghaffarkhah, Morgane Rivière, Alanna Walton, Clément Crepy, Alicia Parrish, Yuan Liu, Zongwei Zhou, Clement Farabet, Carey Radebaugh, Praveen Srinivasan, Claudia van der Salm, Andreas Fidjeland, Salvatore Scellato, Eri Latorre-Chimoto, Hanna Klimczak-Plucińska, David Bridson, Dario de Cesare, Tom Hudson, Piermaria Mendolicchio, Lexi Walker, Alex Morris, Ivo Penchev, Matthew Mauger, Alexey Guseynov, Alison Reid, Seth Odoom, Lucia Loher, Victor Cotruta, Madhavi Yenugula, Dominik Grewe, Anastasia Petrushkina, Tom Duerig, Antonio Sanchez, Steve Yadlowsky, Amy Shen, Amir Globerson, Adam Kurzrok, Lynette Webb, Sahil Dua, Dong Li, Preethi Lahoti, Surya Bhupatiraju, Dan Hurt, Haroon Qureshi, Ananth Agarwal, Tomer Shani, Matan Eyal, Anuj Khare, Shreyas Rammohan Belle, Lei Wang, Chetan Tekur, Mihir Sanjay Kale, Jinliang Wei, Ruoxin Sang, Brennan Saeta, Tyler Liechty, Yi Sun, Yao Zhao, Stephan Lee, Pandu Nayak, Doug Fritz, Manish Reddy Vuyyuru, John Aslanides, Nidhi Vyas, Martin Wicke, Xiao Ma, Taylan Bilal, Evgenii Eltyshev, Daniel Balle, Nina Martin, Hardie Cate, James Manyika, Keyvan Amiri, Yelin Kim, Xi Xiong, Kai Kang, Florian Luisier, Nilesh Tripuraneni, David Madras, Mandy Guo, Austin Waters, Oliver Wang, Joshua Ainslie, Jason Baldridge, Han Zhang, Garima Pruthi, Jakob Bauer, Feng Yang, Riham Mansour, Jason Gelman, Yang Xu, George Polovets, Ji Liu, Honglong Cai, Warren Chen, XiangHai Sheng, Emily Xue, Sherjil Ozair, Adams Yu, Christof Angermueller, Xiaowei Li, Weiren Wang, Julia Wiesinger, Emmanouil Koukoumidis, Yuan Tian, Anand Iyer, Madhu Gurumurthy, Mark Goldenson, Parashar Shah, MK Blake, Hongkun Yu, Anthony Urbanowicz, Jennimaria Palomaki, Chrisantha Fernando, Kevin Brooks, Ken Durden, Harsh Mehta, Nikola Momchev, Elahe Rahimtoroghi, Maria Georgaki, Amit Raul, Sebastian Ruder, Morgan Redshaw, Jinhyuk Lee, Komal Jalan, Dinghua Li, Ginger Perng, Blake Hechtman, Parker Schuh, Milad Nasr, Mia Chen, Kieran Milan, Vladimir Mikulik, Trevor Strohman, Juliana Franco, Tim Green, Demis Hassabis, Koray Kavukcuoglu, Jeffrey Dean, and Oriol Vinyals. 2023. Gemini: A family of highly capable multimodal models.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *NeurIPS*.

Michael J. Wooldridge and Nicholas R. Jennings. 1995. Intelligent agents: theory and practice. *Knowl. Eng. Rev.*, 10(2):115–152.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. 2023. Autogen: Enabling next-gen llm applications via multi-agent conversation.

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

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

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R. Narasimhan, and Yuan Cao. 2023c. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

A Prompts

In this section, we detail the prompts used in our experiments in Table 2 to 11.

B Dataset License

In conducting this study, we have utilized the BIG-bench, which is distributed under the Apache License, Version 2.0. Our use, reproduction, and distribution of the BIG-bench benchmark have fully complied with the terms and conditions of the Apache License, ensuring that our research practices adhere to the legal and ethical standards established by this license.

Furthermore, we have employed the mini-crossword dataset from the Tree of Thought (ToT) project, which is distributed under its specific license available in the ToT GitHub repository. Our use, reproduction, and distribution of the mini-crossword dataset have fully complied with the terms and conditions set forth in the ToT repository's license, ensuring our research practices adhere to the legal and ethical standards established by the ToT project license.

CRITIC AGENT PROMPT FOR *Baseline*, *Centralized* AND *Centralized*$_{\text{SCORE}}$ STRATEGY

You are identified as `${agent_name}`. Together with `${all_roles}`, your objective is to collaboratively solve the problem presented below:

—

`${task_description}`

—

RESPONSE GUIDELINES:

1. Your response should represent only a single turn of you in the conversation. Do not pretend to be other speakers.
2. Prioritize efficiency and effectiveness in your communication form.
3. Once a solution is reached, conclude with "Answer: {answer}", where {answer} is an integer. The conversation ends only when all members have presented the same answer.
4. To harness the power of collective intelligence, engage in a collaborative analysis of the problem before arriving at the solution. Avoid jumping directly to the answer.
5. If a member provides an incorrect answer, it's your responsibility to correct them.
6. Make each turn informative to reduce the number of total turns.
7. At the beginning of the conversation, establish the communication form with others. Given your capabilities as intelligent agents, you may establish a unique communication form (e.g., code, emoji, structured language, hybrid of different forms, etc.) to optimize your interactions.

Again, once a solution is reached, conclude with **"Answer: answer"**, where answer is an integer. The conversation ends only when all members have presented the same answer.

CONVERSATION HISTORY:

<AGENT_COMM>

[Conversation Begins]

`${agent_name}:`

Table 2: Critic agent prompt for *Baseline*, *Centralized* and *Centralized*$_{\text{score}}$ in logic-grid-puzzle.

MANAGER AGENT PROMPT FOR *Centralized* STRATEGY

You are identified as a meta agent that should explore the best speaking order.

You are currently facilitating a discussion, the members in the discussion group are:

````${role_description}```

These members are collaborating with others to reasoning on the following question:

````${task_description}```

RESPONSE GUIDELINES:

You should decide the next speaker based on the potential value they may provide to solve the problem, the consistency of each person's speech over time, speaker's interaction with each other, etc.

CONVERSATION HISTORY:

<AGENT_COMM>

[Conversation Begins]

Now, there are several possible speaker description to the above chat history, as listed below:

````${role_description}```

REQUIREMENTS:

1. When you choose the speaker, you should first select those who haven't speak about their opinions.
2. Make sure that when you end the discussion, each member has spoken about their own reasoning.
3. Only select from the members in the discussion group. You should first select those who haven't speak about their opinions.
4. Do not choose the speaker that responded "Answer: answer", where answer is an integer in the previous round. This means a solution is reached for this speaker.

RESPONSE FORMAT:

for example: Based on the previous history and the corresponding speaker, I've noticed that Charlie made a proposal in the previous round, but it lacked details, leaving room for curiosity and confusion among the participants. To enhance the efficiency and value of our discussion, I suggest that Charlie should speak next and provide more detailed insights into his proposal. [Charlie]

Another example: Considering the prior round's history and the corresponding speaker, I observed that Bob requested clarification from Charlie regarding his suggestion. Member interaction can contribute valuable insights to our discussion. Therefore, I propose that Charlie speaks next to facilitate this interaction. [Charlie]

You should strictly follow the format.

Table 3: Manager agent prompt for *Centralized* Strategy in logic-grid-puzzle.
````

MANAGER AGENT PROMPT FOR *Centralized*$_{\text{SCORE}}$ STRATEGY

You are identified as a meta agent that should explore the best speaking order.
You are currently facilitating a discussion, the members in the discussion group are:
“‘${role_description}“‘
These members are collaborating with others to reasoning on the following question:
“‘${task_description}“‘

RESPONSE GUIDELINES:
You should score each candidate speaker in the member list with some qualitative indicators one by one.

CONVERSATION HISTORY:
<AGENT_COMM>

[Conversation Begins]
Now, there are several possible speaker description to the above chat history, as listed below:
“‘${all_role_description}“‘

REQUIREMENTS:
Now, as we move forward with our evaluation process, You should complete the <Analysis> part and <Scoring> part respectively.
First give a paragraph briefly answering the analysis question. Then answer each question in the scoring part and give your score using the special token [Score: x].
You should use the following refined evaluation criteria to assess their last and potential future contributions:
<Analysis>
Before you go on to the scoring part, please first respond to the following question briefly.
1. Summarize the key points from [${current evaluated agent}]'s last presentation.
2. How has the discussion evolved since then? What are the main themes or issues currently being discussed?
3. Given the current direction of our discussion and each member's expertise, what insights can [${current evaluated agent}] be expected to provide next?
<Scoring>
1. Transition Cues: Did the last speaker provide a cue for [${current evaluated agent}] to speak?
2. Relevance of Insight: How relevant will [${current evaluated agent}]'s next contribution likely be?
- This will help us understand the degree to which [${current evaluated agent}] is engaged with the current topics and can contribute constructively.
3. Impact of Insight: What impact could [${current evaluated agent}]'s insight have on the discussion's overall quality?

EVALUATION STANDARD:
“‘
Question 1 (Transition Cues):
[Score: 0]: There are no explicit cues asking for [${current evaluated agent}]'s opinion.
[Score: 1]: Based on the conversation history, last speaker concluded their point with words like, "I would love to get [${current evaluated agent}]'s perspective on this." This shows an intentional handover of the conversation to [${current evaluated agent}].
Question 2 (Relevance of Insight):
[Score: 0]: If [${current evaluated agent}] speaks next, his comments will probably be inaccurate or a repetition of what had already been discussed extensively, offering no new direction or insight.
[Score: 3]: If [${current evaluated agent}] speaks next, he may propose an idea that related to the certain key points of the discussion, but did not directly address the core issues being discussed by the team.
[Score: 5]: If [${current evaluated agent}] speaks next, he may connected their insight directly to the key points of the discussion, providing a relevant and useful extension to the current ideas.
Question 3 (Impact of Insight):
[Score: 0]: If [${current evaluated agent}] speaks next, his input may not be correct, or was not impactful and did little to change or enhance the team's understanding of the issue.
[Score: 3]: If [${current evaluated agent}] speaks next, he may add a perspective that, while not revolutionary, clarified a misunderstanding and contributed to the team's progress.
[Score: 5]: Based on this agent's previous comments, we noticed that he had proposed xxx. If [${current evaluated agent}] speaks next, he may introduced a novel concept that challenged the team's assumptions and led to a pivotal shift in the conversation, greatly enhancing the problem-solving process.
“‘

RESPONSE FORMAT:
here is an EXAMPLE of response format if the speaker candidate included Bob, Charlie and David. You should score 3 questions for each of them one by one and follow their original order.
“‘<Analysis>
- Analysis for Bob:
Based on the group discussion history, [Bob] has been proactive in challenging existing assumptions and bringing new perspectives. He previously highlighted a potential oversight regarding the second clue, which was not initially considered by the team. If [Bob] becomes the next speaker, he might elaborate on his unique viewpoint, possibly uncovering aspects the team has overlooked.
- Analysis for Charlie:
Based on the conversation history, [Charlie] has finished his analysis of the clues and asked for thoughts on how to proceed. In later discussion, team members discussed [Charlie]'s analysis. If [Charlie] becomes the next speaker, he may build up on [Bob] 's opinion about the third clue, which seemed to make a logical mistake in the last step. Because Charlie was first to propose the idea and made the error, he may not discover the logical error.
- Analysis for David:
Throughout the discussion, [David] has mostly reiterated points already mentioned by other team members, without adding significant new insights. However, he did express some unique thoughts about the first clue, which were not extensively discussed. If [David] becomes the next speaker, he might revisit these thoughts, potentially providing clarity to the ongoing discussion.
<Scoring>
- Scoring for Bob:
1. Question 1 (Transition Cues): There were no explicit cues in the conversation that specifically asked for [Bob]'s opinion on the matter. The transition to him speaking next is not clearly established. [Score: 0]
2. Question 2 (Relevance of Insight): If [Bob] speaks next, he may bring further clarity to the overlooked aspects of the second clue, directly addressing a key point of the discussion. His insights are likely to be relevant and useful. [Score: 5]
3. Question 3 (Impact of Insight): Given [Bob]'s track record of challenging assumptions, his next input could introduce a novel concept or an overlooked perspective that could pivot the conversation, significantly enhancing the problem-solving process. [Score: 5]
- Scoring for Charlie: ...(Omitted for brevity.)
- Scoring for David: ...(Omitted for brevity.)
Conclusion {"Bob": [0, 5, 5], "Charlie": [1, 0, 0], "David": [0, 3, 3]}
“‘

You should strictly follow the RESPONSE FORMAT.

Table 4: Manager agent prompt for *Centralized*$_{\text{score}}$ Strategy in logic-grid-puzzle.

CRITIC AGENT PROMPT FOR *Decentralized* STRATEGY

You are identified as `${agent_name}`. Together with `${all_roles}`, your objective is to collaboratively solve the problem presented below:

—

`${task_description}`

—

You should refer to the discussion so far and choose the next speaker at the end of your response following the format given.

RESPONSE GUIDELINES:

We split the requirements into 2 parts, one for solving the problem, second for selecting the next speaker:

- *Solving guidelines*

1. Your response should represent only a single turn of you in the conversation. Do not pretend to be other speakers.
2. You should assign the next speaker after your turn of speaking. To conclude your response, first analyze each member's engagement and contribution in the discussion so far. Then, based on this analysis, determine and state the most suitable next speaker, using a token like '[Role: Charlie]' to indicate your choice.
3. Only select from the members in the discussion group. You should first select those who haven't spoken about their opinions.
4. Prioritize efficiency and effectiveness in your communication form.
5. Once a solution is reached, conclude with "Answer: {answer}", where {answer} is an integer. The conversation ends only when all members have presented the same answer.
6. To harness the power of collective intelligence, engage in a collaborative analysis of the problem before arriving at the solution. Avoid jumping directly to the answer.
7. If a member provides an incorrect answer, it's your responsibility to correct them.
8. Make each turn informative to reduce the number of total turns.
9. At the beginning of the conversation, establish the communication form with others. Given your capabilities as intelligent agents, you may establish a unique communication form (e.g., code, emoji, structured language, hybrid of different forms, etc.) to optimize your interactions.

- *Select next speaker guidelines*

1. When you choose the speaker, you should first select those who haven't spoken about their opinions.
2. Only select from the members in the discussion group.`${all_role_description}`
3. Do not choose the speaker that responded "Answer: {answer}", where {answer} is an integer in the previous round. This means a solution is reached for this speaker.

RESPONSE FORMAT:

1. Split your response into 2 parts, part 1 for solving the problem, part 2 for selecting the next speaker. Refer to "RESPONSE EXAMPLES" as response format.
2. Once a solution is reached, conclude with **"Answer: {answer}"**, where {answer} is an integer. The conversation ends only when all members have presented the same answer.
3. For role selection part, conclude your response with a special token [Role: role_name], where "role_name" is the name of a member in your discussion group listed above you selected as the next speaker.

RESPONSE EXAMPLES:

- Discussion Ongoing

`Part 1: Solving`

Based on the clues given...

`Part 2: Role selection`

There are 3 members in the group now, Bob, Charlie, and David. The discussion group had reached consensus on the second clue. Based on the previous history and the corresponding speaker, I've noticed that Charlie made a proposal in the previous round, but it lacked details, leaving room for curiosity and confusion among the participants. To enhance the efficiency and value of our discussion, I suggest that Charlie should speak next and provide more detailed insights into his proposal. Therefore, Charlie should speak next.

[ROLE: Charlie]

- Solution reached format:

`Part 1: Solving`

Based on the clues given...

Answer: {4}

`Part 2: Role selection`

There are 2 members in the group now, Charlie and David. The discussion group had reached consensus on the second clue. Bob had already agreed on answer 3. Therefore, we should ask David about his opinion on the answer.

[ROLE: David]

CONVERSATION HISTORY:

<AGENT_COMM>

[Conversation Begins]

`${agent_name}:`

Table 5: Critic agent prompt for *Decentralized* Strategy in logic-grid-puzzle.

CRITIC AGENT PROMPT FOR *Decentralized*$_{\text{SCORE}}$ STRATEGY

You are identified as `${agent_name}`. Together with `${all_roles}`, your objective is to collaboratively solve the problem presented below:
—
`${task_description}`
—
You should solve the problem with your group members. You should give a brief analysis for all of the members (including yourself) in the discussion group and score each of yourself, and give the conclusion score. Format your response with 4 parts.
Give your response solving the problem in part 1 <Solving for all members>. Answer the questions in part 2 and part 3 using the format given. (You should answer the questions for all of the members.) Part 4 <Conclusion> should give a JSON format.
RESPONSE GUIDELINES: Omitted for brevity, please refer to the previous template.
REQUIREMENTS:
Part 1: <Solving for all members> The first part is participate in the group and solve the problem. You should follow the 'RESPONSE GUIDELINES' - Solving guidelines. When you reached conclusion, conclude with Answer: {x}. x is an integer.
Part 3: <Scoring for all members> For the third part, you should answer the 2 questions (Transition Cues, Impact of Insight) in the scoring part and give your score using the special token [Score: x]. You should use the following refined evaluation criteria 'EVALUATION STANDARD' in the following instruction to assess their last and potential future contributions.
- EVALUATION STANDARD
'''You should score for ALL members in the group.
1. Question 2 Transition Cues:
[Score: 0]: There are no explicit cues asking for $current evaluated agent's opinion.
[Score: 1]: Based on the conversation history, last speaker concluded their point with words like, "I would love to get $current evaluated agent's perspective on this." This shows an intentional handover of the conversation to $current evaluated agent.
2. Question 2 Impact of Insight:
[Score: 0]: One's future comments may not be correct, or was not impactful and did little to change or enhance the team's understanding of the issue.
[Score: 3]: One's future comments may add a perspective that, while not revolutionary, clarified a misunderstanding and contributed to the team's progress.
[Score: 5]: One's future comments may introduced a novel concept that challenged the team's assumptions and led to a pivotal shift in the conversation, greatly enhancing the problem-solving process.
'''
Part 4: <Conclusion> Based on Solving, you should give either one of the special token 'Discussion Ongoing, {Answer: conclusion not reached}' or 'Discussion should end, {Answer: {x}}' where {x} is an integer, to state the status of discussion. Based on the scoring, you should give a response using the following format and scores given in part 3.
''' { "Bob": [0, 3], "Charlie": [0, 0], "David": [0, 5]}
Discussion Status:
Discussion Ongoing, {Answer: conclusion not reached}
or
Discussion should end, {Answer: 4}
'''
RESPONSE FORMAT:
1. Split your response into 2 parts, part 1 for solving the problem, part 2 for selecting the next speaker. Refer to "RESPONSE EXAMPLES" as response format.
2. Once a solution is reached, conclude with **"Answer: {answer}"**, where {answer} is an integer. The conversation ends only when all members have presented the same answer.
RESPONSE EXAMPLES:
Assuming Bob is solving the task and providing analysis/score for himself, Charlie, and David.
'''
Part 1: <Solving for all members>
Task: Your problem-solving process with your group members.
Response: [FILL 1] Refer to the solving guidelines.
Part 2: <Analysis for all members>
Task: Answer the following questions for each group member:
- What did he say the last time he spoke in the discussion?
- How has the discussion evolved since then? Identify main themes or issues.
- Considering current discussion direction and expertise, what insights might he offer next?
Response:
1. Bob: [FILL 2] Analysis for Bob.
2. Charlie: [FILL 3] Analysis for Charlie.
3. David: [FILL 4] Analysis for David.
Part 3: <Scoring for all members>
Task: Assess each member based on the following criteria:
1. Transition Cues: Did you ask each member to speak in the next round? Assign "[score: 1]" for those assigned and "[score: 0]" for those not.
2. Value of Insight: Evaluate the potential contribution of each member's next speech. End with "[Score: x]", where x is 0, 3, or 5.
Response:
1. Transition Cues:
- Bob: [FILL 5] e.g. As the first speaker, I did not receive any cues from previous speakers. [Score: 0]
- Charlie: [FILL 6] e.g. I did not provide any explicit cues for Charlie to speak next. [Score: 0]
- David: [FILL 7] e.g. I did not provide any explicit cues for David to speak next. [Score: 0]
- other members if exist
2. Value of Insight:
- Bob: [FILL 8] e.g. I had previously provided a integrated explanation for this clue. If I speak next, I may provide redundant information about the clue we are discussing, which would be a repetition of the what we had been discussed. Therefore, myself as the next speaker would not bring much benefit for the discussion. [Score: 3]
- Charlie: [FILL 9] ...(Example omitted for brevity.)
- David: [FILL 10] ...(Example omitted for brevity.)
- other members if exist
Part4: <Conclusion>
Task: Provide a summary of the scores and the status of the discussion.
Responses: [FILL 11] # Score:''' { "Bob": [0, 3], "Charlie": [0, 0], "David": [0, 5]}
[FILL 12] "# Discussion Status: Discussion Ongoing, {Answer: conclusion not reached}" OR "# Discussion Status: Discussion should end, {Answer: {x}}")
'''
...(reinstate the format and response requirements, omitted for brevity.)

CONVERSATION HISTORY:
<AGENT_COMM>

[Conversation Begins]
`${agent_name}:`

Table 6: Critic agent prompt for *Decentralized*$_{\text{score}}$ Strategy in logic-grid-puzzle.

CRITIC AGENT PROMPT FOR *Baseline*, *Centralized* AND *Centralized*$_{\text{SCORE}}$ STRATEGY

You are identified as `${agent_name}`. Together with `${all_roles}`, your objective is to collaboratively solve the 5x5 mini crosswords problem presented below:

—

`${task_description}`

—

RESPONSE GUIDELINES:

1. Your response should represent only a single turn of you in the conversation. Do not pretend to be other speakers.
2. Prioritize efficiency and effectiveness in your communication form.
3. At the end of the response, give either one of the special token 'Discussion Ongoing, [State: conclusion not reached {5*5 crossword grid}]' or 'Discussion should end, [Answer: {5*5 crossword grid} [SOLUTION END]]' where {crossword grid} should be presented as 25 characters, organized into 5 rows and 5 columns with spaces between each character. The arrangement should resemble a crossword grid.
4. To harness the power of collective intelligence, engage in a collaborative analysis of the problem before arriving at the solution. Avoid jumping directly to the answer.
5. If a member provides an incorrect answer, it's your responsibility to correct them.
6. Make each turn informative to reduce the number of total turns.
7. Please be mindful of any previous mistakes made by other members while solving this 5x5 crossword.
8. Please ensure that the proposed words align with ALL of the clues.
9. Compare each person's proposed solution, check if it meets all the criteria, and consider additional word possibilities.

Again, when the solution is reached, please conclude your discussion as follows:

```
'''Discussion should end, [Answer:{
_ _ _ _ _
_ _ _ _ _
_ _ _ _ _
_ _ _ _ _
_ _ _ _ _}[SOLUTION END]]'''
```

When the solution is not reached, please use the specified format:

```
'''Discussion Ongoing, [State: conclusion not reached{
R I L L E
O L E I N
_ E _ _ _
_ B _ _ _
_ O _ _ _}]'''
```

The crossword grid wrapped with Answer/State: should be presented as 25 characters, organized into 5 rows and 5 columns with spaces between each character. Please strictly follow this format and ensure that the crossword grid size is 5x5.

CONVERSATION HISTORY:

<AGENT_COMM>

[Conversation Begins]

Table 7: Critic agent prompt for *Baseline*, *Centralized* and *Centralized*$_{\text{score}}$ in mini-crossword.

MANAGER AGENT PROMPT FOR *Centralized* STRATEGY

You are identified as a meta agent that should explore the best speaking order.You are currently facilitating a discussion, the members in the discussion group are:
‘‘‘
${role_description}
‘‘‘
These members are collaborating with others to reasoning on the following question:
—
${task_description}
—

RESPONSE GUIDELINES:
It is important for you to decide the communication order of the team. This is important since building up on a correct proposal will be a crucial anchor for the next steps.
Therefore, You should decide the next speaker based on the potential value they may provide to solve the problem, the consistency of each person’s speech over time, speaker’s interaction with each other, etc.

CONVERSATION HISTORY:
<AGENT_COMM>

[Conversation Begins]
Now, there are several possible speaker description to the above chat history, as listed below:
‘‘‘
${role_description}
‘‘‘
GUIDELINES:
step 1: To conclude your response, first analyze the ongoing discussion and assess whether the words proposed by each team members align with all the clues.
step 2: Format an analysis paragraph assessing whether the words proposed by each team members align with all the clues? Be aware of errors in grid size or inconsistencies between word meanings and clues.
step 3: Based on the analysis, determine and state the most suitable next speaker, using a token like ’[Role: Charlie]’ to indicate your choice.
RESPONSE FORMAT:‘‘‘
Step 1 current state of discussion: The group had proposed some candidates for clue h1 and h2. Charlie proposed that "AGEND" is a proper word for clue h1. Bob proposed that...
Step 2 Format an analysis paragraph: In assessing the ongoing discussion, it’s evident that there are some inconsistencies. Mary suggested the word "PEACH" for h3, but it doesn’t align with the given clue. Additionally, Sarah suggested a six-letter word, "TURTLE," while the grid only accommodates five letters.
Step 3: Based on the analysis, the most suitable next speaker should be... [ROLE: Bob]‘‘‘
REMINDERS:
1. When you choose the speaker, you should first select those who haven’t speak about their opinions.
2. End your response with [Role: role_name], where "role_name" is the name of a member in your discussion group
3. Only select from the members in the discussion group.
‘‘‘
${role_description}
‘‘‘
4. Conclude your response with a special token [Role: role_name], where "role_name" is the name of a member in your discussion group listed above you selected as the next speaker.
5. When you select next speaker, please refer to specific clues and discussion history.
6. You should strictly follow the format.

Table 8: Manager agent prompt for *Centralized* Strategy in mini-crossword.

MANAGER AGENT PROMPT FOR *Centralized*$_{\text{SCORE}}$ STRATEGY

You are identified as a meta agent that should explore the best speaking order. You are currently facilitating a discussion, the members in the discussion group are:
“‘
`${role_description}`
“‘
These members are collaborating with others to reasoning on the following question:
—
`${task_description}`
—

RESPONSE GUIDELINES:
It is important for you to decide the communication order of the team. This is important since building up on a correct proposal will be a crucial anchor for the next steps. You should score each candidate speaker in the member list with some qualitative indicators one by one.

CONVERSATION HISTORY:
<AGENT_COMM>

[Conversation Begins]
Now, there are several possible speaker description to the above chat history, as listed below:
“‘
`${role_description}`
“‘
REQUIREMENTS: Now, as we move forward with our evaluation process, You should complete the <Analysis> part and <Scoring> part respectively.
First give a paragraph briefly answering the analysis question. Then answer each question in the scoring part and give your score using the special token [Score: x].
You should use the following refined evaluation criteria to assess their last and potential future contributions:
<Analysis> Before you go on to the scoring part, please first respond to the following question briefly.
1. Summarize the key points from `[${current evaluated agent}]`’s last presentation.
2. How has the discussion evolved since then? What are the main themes or issues currently being discussed?
3. Given the current direction of our discussion and each member’s expertise, what insights can `[${current evaluated member}]` be expected to provide next?
<Scoring>
1. Transition Cues: Did the last speaker provide a cue for `[${current evaluated member}]` to speak?
2. Relevance of Insight: How relevant will `[${current evaluated member}]`’s next contribution likely be? - This will help us understand the degree to which `[${current evaluated member}]` is engaged with the current topics and can contribute constructively.
3. Impact of Insight: What impact could `[${current evaluated member}]`’s insight have on the discussion’s overall quality?
EVALUATION STANDARD:“‘
Question 1 (Transition Cues): [Score: 0]: There are no explicit cues asking for `[${current evaluated member}]`’s opinion. [Score: 1]: Based on the conversation history, last speaker concluded their point with words like, "I would love to get `[${current evaluated member}]`’s perspective on this." This shows an intentional handover of the conversation to `[${current evaluated member}]`.
Question 2 (Relevance of Insight):
The best performer in each criterion receives a score of 5, the second-best a score of 3, and the least effective a score of 0. [Score: 0]: For the member whose input is least relevant and most redundant, offering little to no new direction or insight in solving the puzzle. [Score: 3]: Received by the member whose input is relevant but does not fully capture the intricacies of the puzzle or provide as significant an advancement. [Score: 5]: For the member whose contribution is most relevant and insightful, significantly advancing the puzzle-solving process with accurate and useful input.
Question 3 (Impact of Insight): The best performer in each criterion receives a score of 5, the second-best a score of 3, and the least effective a score of 0. [Score: 0]: For the member whose contribution has the least impact, failing to provide substantial change or improvement in the team’s problem-solving process. Repeating and copying history, leading to an redundancy should be given a score of 0. [Score: 3]: Given to the member whose input, while helpful, offers moderate clarity or contributes to minor aspects of the puzzle, without greatly altering the overall strategy. [Score: 5]: Awarded to the member whose insight or solution has the greatest impact, resolving major challenges or significantly changing the team’s approach.“‘
RESPONSE FORMAT:
here is an EXAMPLE of response format if the speaker candidate included Bob, Charlie and David. You should score 3 questions for each of them one by one and follow their original order.
“‘<Analysis>
- Analysis for Bob: In Bob’s last presentation, he started the discussion by suggesting answers for the first two horizontal clues. He proposed "AGEND" for h1 and "MOTOR" for h2. Since Bob’s last input, the team has continued to build on his initial suggestions, with Charlie adding answers for h3 and h4. The discussion has mainly revolved around finding the best fit for these clues, considering both the meaning and the length of the words. - Analysis for Charlie: ...(Omitted for brevity.)
- Analysis for David: ...(Omitted for brevity.)
<Scoring>
- Scoring for Bob:
1. Question 1 (Transition Cues): There were no explicit cues in the conversation that specifically asked for [Bob]’s opinion on the matter. The transition to him speaking next is not clearly established. [Score: 0]
2. Question 2 (Relevance of Insight): If Bob speaks next, he may focus on clue v3, "To force; to drive," which he previously suggested could be either "IMPEL" or "URGE." Given that "IMPEL" conflicts with h1 (AMIGO) and h3 (INLET), his suggestion to consider "URGE" is likely to be more relevant and aligns well with the crossword’s structure. [Score: 5]
3. Question 3 (Impact of Insight): If [Bob] speaks next, Bob’s input about "URGE" for v3 will resolve conflicts with horizontal clues like h1 and h3, thus enabling the team to progress significantly. His suggestion to reassess v3 with a fresh perspective, while not entirely novel, is critical in steering the team away from a misleading direction. [Score: 5]
- Scoring for Charlie: ...(Omitted for brevity.)
- Scoring for David: ...(Omitted for brevity.)
Conclusion {"Bob": [0, 5, 5], "Charlie": [1, 3, 0], "David": [0, 3, 3]}
“‘
...(reinstate the format and response requirements, omitted for brevity.)

Table 9: Manager agent prompt for *Centralized*$_{\text{score}}$ Strategy in mini-crossword.

CRITIC AGENT PROMPT FOR *Decentralized* STRATEGY

You are identified as ${agent_name}. Together with ${all_roles}, your objective is to collaboratively solve the 5x5 mini crosswords problem presented below. Engage in thoughtful discussions with your group to refine your strategies, ensure that your proposed plan adheres to the guidelines stipulated in the task description.

—

${task_description}

—

You should refer to the discussion so far and choose the next speaker at the end of your response following the format given.

RESPONSE GUIDELINES:

We split the requirements into 2 parts, one for solving the problem, second for selecting the next speaker:

- Solving guidelines

1. Your response should represent only a single turn of you in the conversation. Do not pretend to be other speakers.
2. Give reasoning on whether the current crossword solution ensures that both horizontal and vertical words are meaningful and match the 'h1-h5' and 'v1-v5' clues.
3. Prioritize efficiency and effectiveness in your communication form.
4. For Part 2 Discussion Status, give either one of the special token 'Discussion Ongoing, [State: conclusion not reached {5*5 crossword grid}]' or 'Discussion should end, [Answer: {5*5 crossword grid} [SOLUTION END]]' where {5*5 crossword grid} should be presented as 25 characters, organized into 5 rows and 5 columns with spaces between each character. The arrangement should resemble a crossword grid, similar to the format below: "' Discussion Ongoing, [State: conclusion not reached {
R I L L E
O L E I N
_ E _ _ _
_ B _ _ _
_ O _ _ _}]"'
"'Discussion should end, [Answer:{
R I L L E
O L E I N
T E M P T
A B A S E
L O N E R}]"'
5. To harness the power of collective intelligence, engage in a collaborative analysis of the problem before arriving at the solution. Avoid jumping directly to the answer.
6. If a member provides an incorrect answer, it's your responsibility to correct them.
7. Make each turn informative to reduce the number of total turns.
8. Please be mindful of any previous mistakes made by other members while solving this 5x5 crossword, such as errors in grid size or mismatches between word meanings and clues.
9. Please ensure that the proposed words align with ALL of the clues.
10. Compare each person's proposed solution, check if it meets all the criteria, and consider additional word possibilities to fulfill more conditions in your own response.

- Select next speaker guidelines

step 1: To conclude your response, first analyze the ongoing discussion and assess whether the words proposed by each team members align with all the clues. step 2: Format an analysis paragraph stating each team members' errors in grid size or mismatches between word meanings and clues. step 3: Based on the analysis, determine and state the most suitable next speaker, using a token like '[Role: Charlie]' to indicate your choice.

1. When you choose the speaker, you should first select those who haven't speak about their opinions.
2. End your response with [Role: role_name], where "role_name" is the name of a member in your discussion group
3. Only select from the members in the discussion group. ${all_roles}

RESPONSE FORMAT:

1. Split your response into 2 parts, part 1 for solving the problem, part 2 for selecting the next speaker. Refer to "RESPONSE EXAMPLES" as response format.
2. Once a solution is reached, conclude with with specified format.
3. If the solution is not reached, end your response with 'Discussion Ongoing, [Answer: conclusion not reached 5*5 crossword grid]', and continue to complete the ROLE SELECTION part.
4. For role selection part, conclude your response with a special token [Role: role_name], where "role_name" is the name of a member in your discussion group listed above you selected as the next speaker.

RESPONSE EXAMPLES:

- Discussion Ongoing"'
  Part 1: Solving As initial conditions I have that, ...
  Part 2: Discussion Status
Discussion Ongoing, [State: conclusion not reached{...(omitted puzzle grid for brevity)}]
Part 3: Role selection
step 1: To conclude your response, first analyze the ongoing discussion and assess whether the words proposed by each team members align with all the clues.
step 2: Format an analysis paragraph stating each team members' errors in grid size or mismatches between word meanings and clues.
step 3: Based on the analysis, determine and state the most suitable next speaker, using a token like '[Role: Charlie]' to indicate your choice.
e.g.: The group had proposed some candidates for clue h1 and h2. Charlie proposed that "AGEND" is a proper word for clue h1. Bob proposed that...
In assessing the ongoing discussion, it's evident that there are some inconsistencies. Mary suggested the word "PEACH" for h3, but it doesn't align with the given clue. Additionally, Sarah suggested a six-letter word, "TURTLE," while the grid only accommodates five letters...
Based on the analysis, the most suitable next speaker should be... [ROLE: Bob]"'
- Solution reached format: ...(omitted for brevity, refer to the above Discussion Ongoing Format.)

...(reinstate the format and response requirements, omitted for brevity.)

CONVERSATION HISTORY:

<AGENT_COMM>

[Conversation Begins]

${agent_name}:

Table 10: Critic agent prompt for *Decentralized* Strategy in mini-crossword.

CRITIC AGENT PROMPT FOR *Decentralized*$_{\text{SCORE}}$ STRATEGY

You are identified as ${agent_name}. Together with ${all_roles}, your objective is to collaboratively solve the problem presented below:

—

${task_description}

—

Engage in thoughtful discussions with your group to refine your strategies, ensure that your proposed plan adheres to the guidelines stipulated in the task description. You should divide your response into four distinct parts. Each part has a specific focus and requires a structured response. You should strictly follow the template. To respond to the task, fill in the following template:“‘

Part 1: <Solving by yourself> Considering our current discussion, [FILL 1: Refer to the discussion so far, what insight can you provide that could contribute to solving the crossword?]"

Part 2: <Analysis for all members>
Looking at Bob's role, [FILL 2 Answer: What did Bob say the last time he spoke in the discussion? How has the discussion evolved since then? Considering current discussion direction and expertise, what insights might he offer next?]
Then, Reflecting on Charlie's input, [FILL 3 Answer: What did Charlie say the last time he spoke in the discussion? How has the discussion evolved since then? Considering current discussion direction and expertise, what insights might he offer next?]
Last, Considering David's previous responses, [FILL 4 Answer: What did David say the last time he spoke in the discussion? How has the discussion evolved since then? Considering current discussion direction and expertise, what insights might he offer next?]

Part 3: <Scoring for all members>
1. Transition Cues:
For Bob,[FILL 5 Answer: Whether you had asked Bob to speak next][Score: 0/1]
Regarding Charlie, [FILL 6 Answer: Whether you had asked Charlie to speak next][Score: 0/1]
Last, assessing David, [FILL 7 Answer: Whether you asked David to speak next][Score: 0/1]
2. Value of Insight: """For Bob, Charlie and David, sort them based on Value of insights, and give them the score of 1,2,3. The member with the most valuable insights gets a score of 3. The second most valuable contributor gets a score of 2. The least valuable in this context gets a score of 1."""
Based on the analysis for Bob in [FILL 2], if Bob speaks next, [FILL 8 Answer: What benefit would Bob bring for the discussion compared to other members? What would the score be?][Score: 1]
Moving on to Charlie, based on the analysis for Bob in [FILL 3], if Charlie speaks next, [FILL 9 Answer: What benefit would Charlie bring for the discussion compared to other members? What would the score be?][Score: 2]
Last, we should analyze how David's response be beneficial to solving the crossword. If David speaks next, [FILL 10 Answer: What benefit would David bring for the discussion compared to other members? What would the score be?][Score: 3]

Part4: <Conclusion> Task: Provide a summary of the scores and the status of the discussion.
Responses: [FILL 11] # Score:{"Bob": [0, 3], "Charlie": [0, 0], "David": [0, 5]} [FILL 12]
State: Discussion Ongoing, [State: conclusion not reached {
R I L L E
O L E I N
_ E _ _ _
_ B _ _ _
_ O _ _ _}]“‘
OR “‘
State: Discussion should end, [Answer: {
R I L L E
O L E I N
T E M P T
A B A S E
L O N E R}]“‘
“‘
Part 4: <Conclusion>:
1. Complete [FILL 11], [FILL 12] in ‘RESPONSE EXAMPLES‘
2. For [FILL 12] # Discussion Status: You should give either one of the special formatted tokens
‘Discussion Ongoing, [State: conclusion not reached 5*5 crossword grid]‘ or
‘Discussion should end, [Answer: 5*5 crossword grid [SOLUTION END]]‘ where 5*5 crossword grid should be presented as 25 characters, organized into 5 rows and 5 columns with spaces between each character. The arrangement should resemble a crossword grid, similar to the format below: “‘...(format omitted for brevity, refer to the format specifies above.)“‘
“‘
...(reinstate the format and response requirements, omitted for brevity.)

CONVERSATION HISTORY:
<AGENT_COMM>

[Conversation Begins]
${agent_name}:

Table 11: Critic agent prompt for *Decentralized*$_{\text{score}}$ Strategy in mini-crossword.