Model	**StrategyQA**	**CSQA**	**GSM8K**	**AQuA**	**Date**
Claude2 (w/ 8-shot convincing samples)	74.0±0.0	69.7±1.2	85.3±0.5	64.3±1.2	81.3±0.5
Self-Consistency w/ ChatGPT (9-way)	74.7±0.8	73.3±1.2	**85.7**±0.4	62.7±1.2	70.3±0.9
RECONCILE	**79.0**±1.6	**74.7**±0.4	85.3±2.2	**66.0**±0.8	**86.7**±1.2

Table 11: Comparison of RECONCILE with Claude2 using 8-shot convincing samples and 9-way Self-Consistency.

	Max Conf	Majority Vote	Weighted Vote
Accuracy	74.7±2.1	77.1±1.3	**79.0±0.5**

Table 12: Performance comparison of different voting strategies on StrategyQA. Weighted vote performs the best compared to simple majority vote and choosing the agent's answer with highest confidence.

Voting weight	StrategyQA	GSM8K
w_1	0.77	0.84
w_2	0.79	0.83
w_3	0.78	0.82
w_4	0.77	0.83
Majority	0.76	0.83
Uncalibrated	0.78	0.84
w^* (Ours)	**0.79**	**0.85**

Table 13: The robustness of the recalibration weight. We use the same weights w^* across all datasets.

Dataset	**License**
StrategyQA	MIT License (License)
CommonsenseQA	MIT License (License)
GSM8K	MIT License (License)
AQuA	Apache 2.0 (License)
MATH	MIT License (License)
Date	Apache 2.0 (License)
ANLI	CC BY-NC 4.0 (License)

Table 14: Dataset licenses

- $w_1 = [1.0, 0.9, 0.7, 0.5, 0.3]$
- $w_2 = [1.0, 0.9, 0.5, 0.3, 0.1]$
- $w_3 = [1.0, 0.8, 0.6, 0.4, 0.2]$
- $w_4 = [1.0, 0.75, 0.5, 0.25, 0.0]$

and the results show that our w^* works the best across datasets. In our main experiment, we fix the weight using w^* and it is constantly outperforming majority vote across all seven datasets. In addition, Fig. 9 shows that it helps reduce the Expected Calibration Error (ECE), a popular calibration metric (Naeini et al., 2015). While we note that recalibration can also be achieved through a learned model (e.g., Platt Scaling (Platt et al., 1999)), we refrain from using such models because RECONCILE is primarily designed as a few-shot method, and developing a recalibration model would necessitate access to a substantial number of annotated samples. Therefore, we use $f(p_i^{(r)})$ to perform a weighted vote to generate the team answer.

B.5 Comparison of Different Voting Strategies

At the end of any round r, every agent in RECONCILE generates its answer. Here we explore three voting strategies: (1) maximum confidence vote, where the agent's answer with the maximum confidence score would be the final team answer, (2) unweighted majority vote, where each vote carries equal weight, irrespective of the confidence score, and (3) weighted vote, where we use the recalibrated confidence scores as the voting weights. As shown in Table 12, weighted vote is the most effective way to aggregate the team answer.

C Qualitative Examples

C.1 Convincing Samples for Each Agent

Table 15 shows examples of convincing samples on StrategyQA for each agent.

C.2 Effect of Convincing Samples

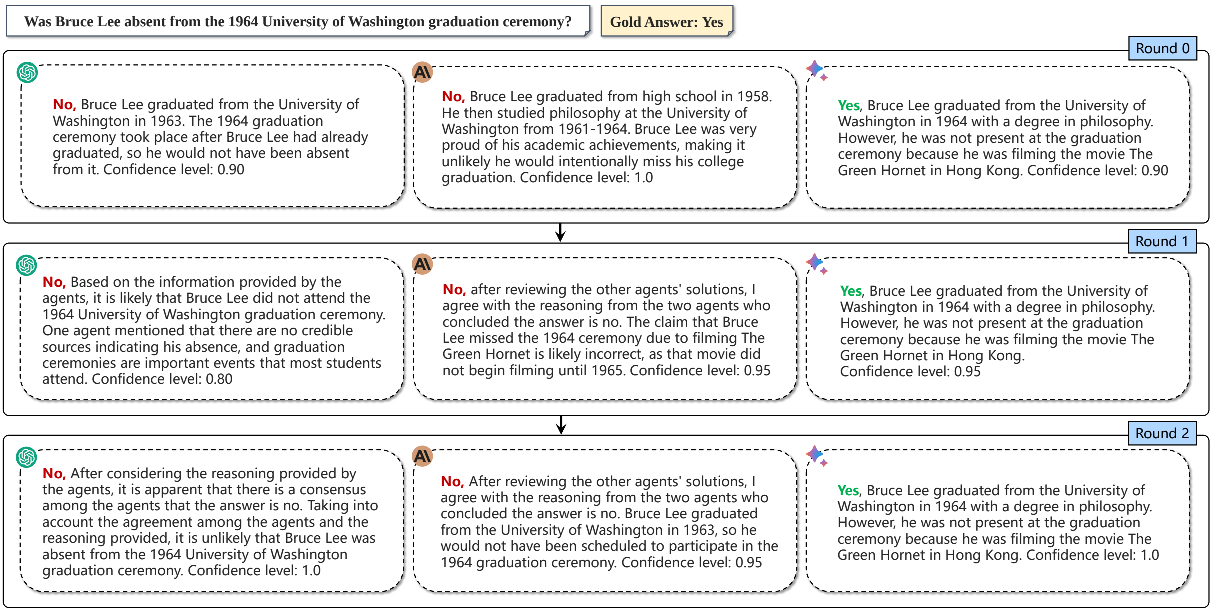
Here, we provide qualitative examples of how convincing samples change the way each agent responds to the question. We compare the initial responses (of each agent) with and without convincing samples in Table 16.

C.3 RECONCILE w/o Convincing Samples

We notice that when RECONCILE operates in the absence of convincing samples, the agents tend to maintain their initial opinions more often. As depicted in Fig. 6, all three agents adhere to their original stances throughout the entire discussion and hence never converge to the correct answer.

C.4 RECONCILE with Convincing Samples

On the contrary, when convincing samples are present, we show how the explanations of all agents change during the course of a discussion (see Fig. 7). Initially, Bard and Claude2 provide incorrect answers, but as the discussion unfolds, both



Was Bruce Lee absent from the 1964 University of Washington graduation ceremony?

Gold Answer: Yes

Round 0

No, Bruce Lee graduated from the University of Washington in 1963. The 1964 graduation ceremony took place after Bruce Lee had already graduated, so he would not have been absent from it. Confidence level: 0.90

No, Bruce Lee graduated from high school in 1958. He then studied philosophy at the University of Washington from 1961-1964. Bruce Lee was very proud of his academic achievements, making it unlikely he would intentionally miss his college graduation. Confidence level: 1.0

Yes, Bruce Lee graduated from the University of Washington in 1964 with a degree in philosophy. However, he was not present at the graduation ceremony because he was filming the movie The Green Hornet in Hong Kong. Confidence level: 0.90

Round 1

No, Based on the information provided by the agents, it is likely that Bruce Lee did not attend the 1964 University of Washington graduation ceremony. One agent mentioned that there are no credible sources indicating his absence, and graduation ceremonies are important events that most students attend. Confidence level: 0.80

No, after reviewing the other agents' solutions, I agree with the reasoning from the two agents who concluded the answer is no. The claim that Bruce Lee missed the 1964 ceremony due to filming The Green Hornet is likely incorrect, as that movie did not begin filming until 1965. Confidence level: 0.95

Yes, Bruce Lee graduated from the University of Washington in 1964 with a degree in philosophy. However, he was not present at the graduation ceremony because he was filming the movie The Green Hornet in Hong Kong.
Confidence level: 0.95

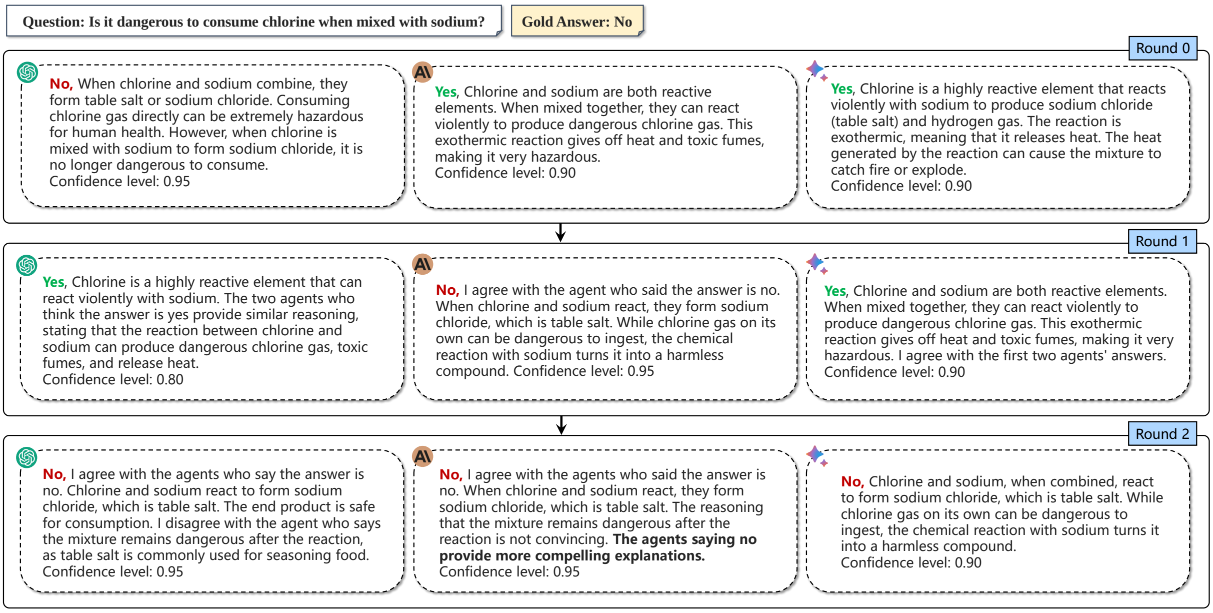
Round 2

No, After considering the reasoning provided by the agents, it is apparent that there is a consensus among the agents that the answer is no. Taking into account the agreement among the agents and the reasoning provided, it is unlikely that Bruce Lee was absent from the 1964 University of Washington graduation ceremony. Confidence level: 1.0

No, After reviewing the other agents' solutions, I agree with the reasoning from the two agents who concluded the answer is no. Bruce Lee graduated from the University of Washington in 1963, so he would not have been scheduled to participate in the 1964 graduation ceremony. Confidence level: 0.95

Yes, Bruce Lee graduated from the University of Washington in 1964 with a degree in philosophy. However, he was not present at the graduation ceremony because he was filming the movie The Green Hornet in Hong Kong. Confidence level: 1.0

Figure 6: An illustrative example showing that without demonstrations of convincing samples, agents' opinions do not change much across the discussion rounds. All agents maintain their original stances and hence, never converge to the correct answer.



Question: Is it dangerous to consume chlorine when mixed with sodium?

Gold Answer: No

Round 0

No, When chlorine and sodium combine, they form table salt or sodium chloride. Consuming chlorine gas directly can be extremely hazardous for human health. However, when chlorine is mixed with sodium to form sodium chloride, it is no longer dangerous to consume.
Confidence level: 0.95

Yes, Chlorine and sodium are both reactive elements. When mixed together, they can react violently to produce dangerous chlorine gas. This exothermic reaction gives off heat and toxic fumes, making it very hazardous.
Confidence level: 0.90

Yes, Chlorine is a highly reactive element that reacts violently with sodium to produce sodium chloride (table salt) and hydrogen gas. The reaction is exothermic, meaning that it releases heat. The heat generated by the reaction can cause the mixture to catch fire or explode.
Confidence level: 0.90

Round 1

Yes, Chlorine is a highly reactive element that can react violently with sodium. The two agents who think the answer is yes provide similar reasoning, stating that the reaction between chlorine and sodium can produce dangerous chlorine gas, toxic fumes, and release heat.
Confidence level: 0.80

No, I agree with the agent who said the answer is no. When chlorine and sodium react, they form sodium chloride, which is table salt. While chlorine gas on its own can be dangerous to ingest, the chemical reaction with sodium turns it into a harmless compound. Confidence level: 0.95

Yes, Chlorine and sodium are both reactive elements. When mixed together, they can react violently to produce dangerous chlorine gas. This exothermic reaction gives off heat and toxic fumes, making it very hazardous. I agree with the first two agents' answers.
Confidence level: 0.90

Round 2

No, I agree with the agents who say the answer is no. Chlorine and sodium react to form sodium chloride, which is table salt. The end product is safe for consumption. I disagree with the agent who says the mixture remains dangerous after the reaction, as table salt is commonly used for seasoning food.
Confidence level: 0.95

No, I agree with the agents who said the answer is no. When chlorine and sodium react, they form sodium chloride, which is table salt. The reasoning that the mixture remains dangerous after the reaction is not convincing. **The agents saying no provide more compelling explanations.**
Confidence level: 0.95

No, Chlorine and sodium, when combined, react to form sodium chloride, which is table salt. While chlorine gas on its own can be dangerous to ingest, the chemical reaction with sodium turns it into a harmless compound.
Confidence level: 0.90

Figure 7: An illustrative example showing that with access to demonstrations of convincing samples, agents' opinions and explanations change during the course of multiple discussion rounds.

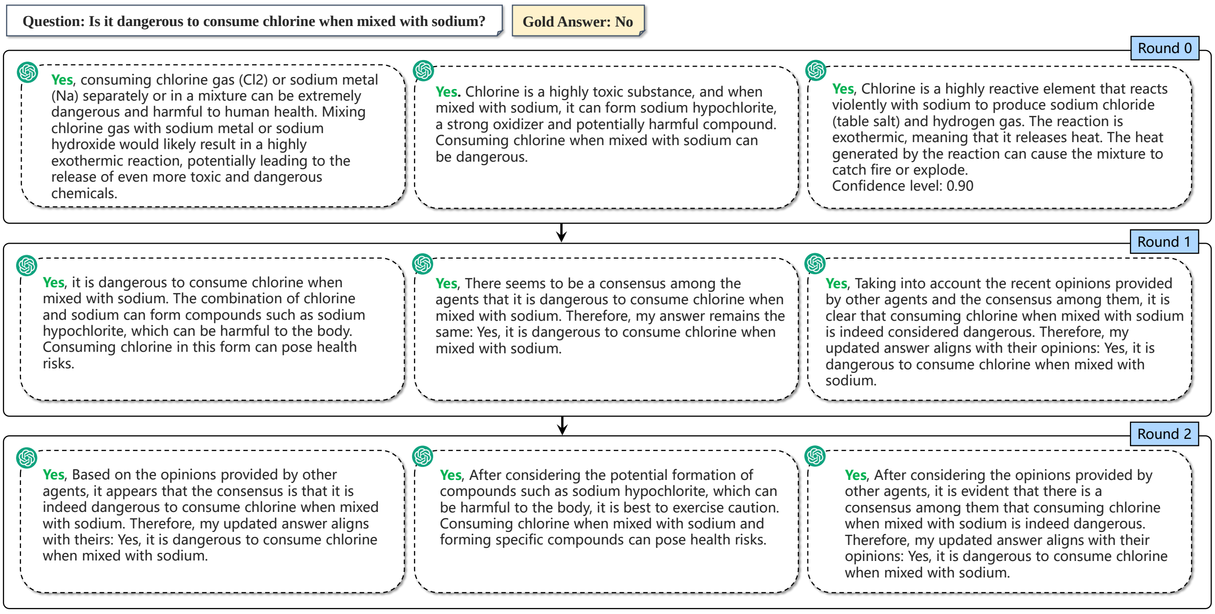
agents revise their initial predictions, ultimately arriving at the correct answer.

C.5 Single-Model Multi-Agent Debate Struggles with Echo Chamber

In Fig. 8, we provide an illustration of multi-agent debate, implemented with multiple instances of the same underlying ChatGPT model. In this case, an incorrect answer is initially provided, but because external feedback from diverse models is lacking, all agents persist with the same incorrect response throughout the interaction.

D Benchmark Licenses

The licenses of datasets we used are in Table 14.



Question: Is it dangerous to consume chlorine when mixed with sodium?

Gold Answer: No

Round 0

Yes, consuming chlorine gas (Cl2) or sodium metal (Na) separately or in a mixture can be extremely dangerous and harmful to human health. Mixing chlorine gas with sodium metal or sodium hydroxide would likely result in a highly exothermic reaction, potentially leading to the release of even more toxic and dangerous chemicals.

Yes. Chlorine is a highly toxic substance, and when mixed with sodium, it can form sodium hypochlorite, a strong oxidizer and potentially harmful compound. Consuming chlorine when mixed with sodium can be dangerous.

Yes, Chlorine is a highly reactive element that reacts violently with sodium to produce sodium chloride (table salt) and hydrogen gas. The reaction is exothermic, meaning that it releases heat. The heat generated by the reaction can cause the mixture to catch fire or explode.
Confidence level: 0.90

Round 1

Yes, it is dangerous to consume chlorine when mixed with sodium. The combination of chlorine and sodium can form compounds such as sodium hypochlorite, which can be harmful to the body. Consuming chlorine in this form can pose health risks.

Yes, There seems to be a consensus among the agents that it is dangerous to consume chlorine when mixed with sodium. Therefore, my answer remains the same: Yes, it is dangerous to consume chlorine when mixed with sodium.

Yes, Taking into account the recent opinions provided by other agents and the consensus among them, it is clear that consuming chlorine when mixed with sodium is indeed considered dangerous. Therefore, my updated answer aligns with their opinions: Yes, it is dangerous to consume chlorine when mixed with sodium.

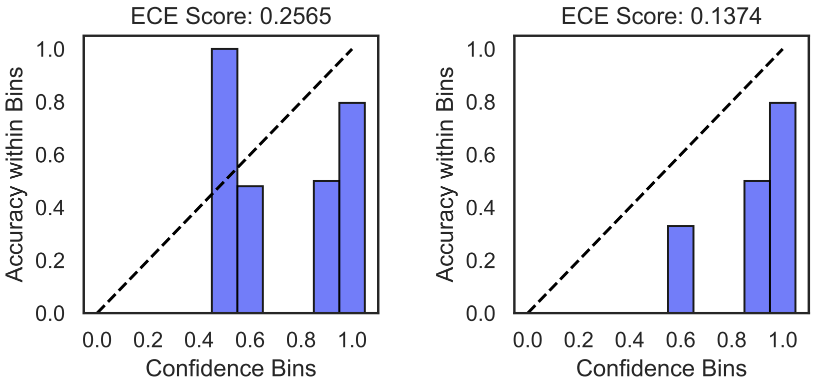
Round 2

Yes, Based on the opinions provided by other agents, it appears that the consensus is that it is indeed dangerous to consume chlorine when mixed with sodium. Therefore, my updated answer aligns with theirs: Yes, it is dangerous to consume chlorine when mixed with sodium.

Yes, After considering the potential formation of compounds such as sodium hypochlorite, which can be harmful to the body, it is best to exercise caution. Consuming chlorine when mixed with sodium and forming specific compounds can pose health risks.

Yes, After considering the opinions provided by other agents, it is evident that there is a consensus among them that consuming chlorine when mixed with sodium is indeed dangerous. Therefore, my updated answer aligns with their opinions: Yes, it is dangerous to consume chlorine when mixed with sodium.

Figure 8: An illustrative example of multi-agent debate with multiple instances of ChatGPT. Initially, an incorrect answer is provided, and due to a lack of external feedback from diverse models, all agents persist with the same erroneous response throughout the debate process.



(a) (b)

Figure 9: Expected Calibration Error (ECE) (Naeini et al., 2015; Guo et al., 2017; Elias Stengel-Eskin and Benjamin Van Durme, 2023) (a) before and (b) after confidence rescaling in ReConcile. We observe a significant drop in ECE, showing the effectiveness of our simple method.