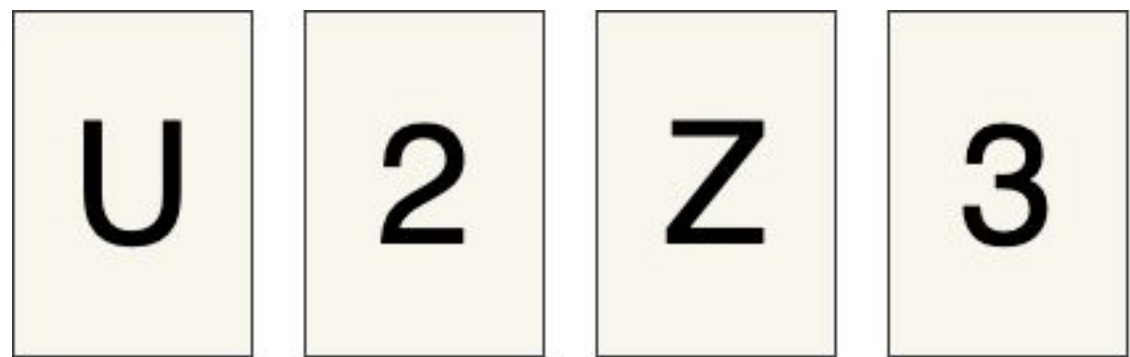


DeliData: A dataset for deliberation in multi-party problem solving

Georgi Karadzhov, Tom Stafford, Andreas Vlachos

1. Task - Wason Card Selection

- Extensively studied in psychology for group decision making
- Most people get it wrong at first (20% success rate)
- By engaging in conversation groups solve it better - 70-80% success rate



Each of the 4 cards below has a letter on one side and a number on the other. Which card(s) do you need to turn to test the rule: **All cards with vowels on one side have an even number on the other.**

Kiwi: What answers did everyone put?

Beaver: I put U and 2

Duck: I selected only U

Kiwi: Hmm... Do we actually need 2?

Duck: If we turn 2, there may be a consonant, but that's fine

Zebra: Surely we don't need 7 right

Kiwi: Why not 7 though?

Zebra: Actually, I am wrong. There may be a vowel behind 7

Kiwi: Do we all agree on U and 7?

ProbingModeration

NPDSolutioncomplete_solution

NPDSolutioncomplete_solution

ProbingReasoning

NPDReasoning

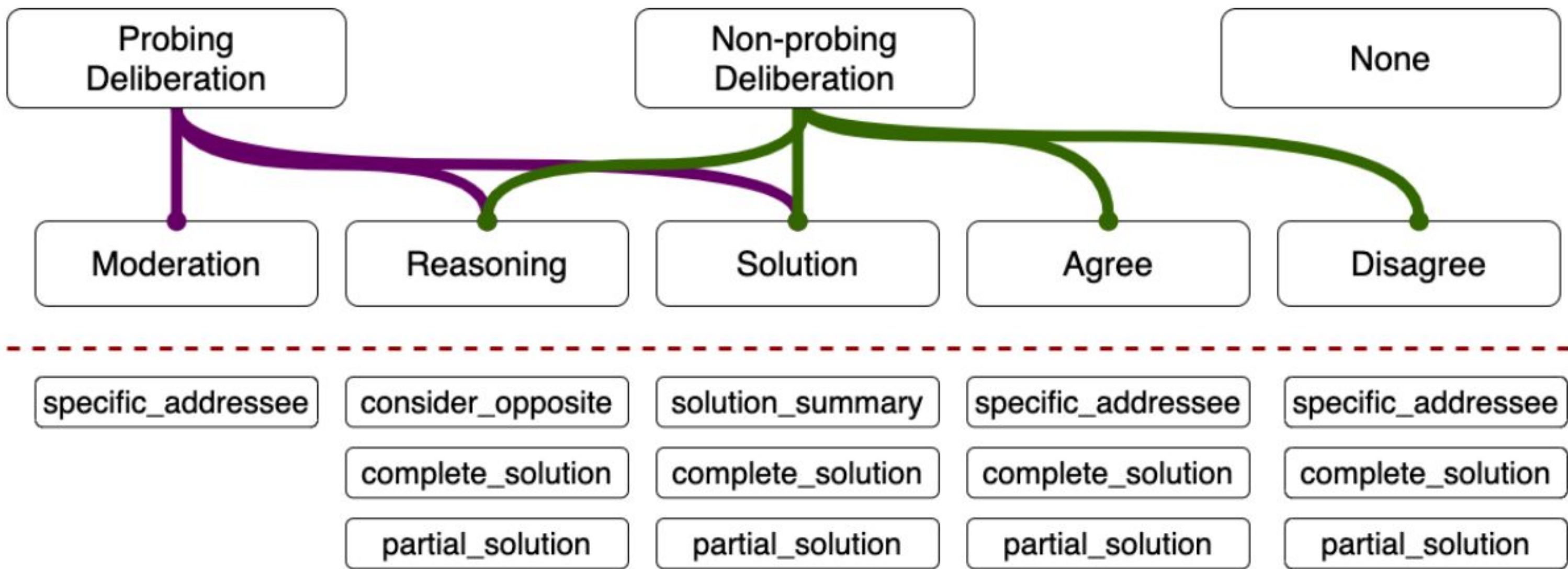
NPDSolutionpartial_solution

ProbingReasoningconsider_opposite

NPDReasoning

ProbingSolutionsolution_summary

4. Annotating Deliberation



	Probing	NPD	Total
Moderation	89	0	89
Reasoning	59	453	512
Solution	66	305	371
Agree	0	265	265
Disagree	0	9	9
Total	214	1032	1246

Additional Label	Count	%
specific_addressee	55	4.4
complete_solution	258	20.7
partial_solution	79	6.3
solution_summary	40	3.2
consider_opposite	11	0.9

2. Experimental Setup:

- We developed a platform for collecting group dialogues in-the-wild
- Each experiment run contains 3 phases: (i) participants venture a solo solution, (ii) then engage in a group discussion, finally (iii) submit a revised solution

3. In-lab and in-the-wild data collection:

DeliData will be the first publicly available dataset on group collaboration

	In-lab	In-the-wild	Total
Number of Dialogues	17	209	226
Total Participants	50	658	708
AVG utterances	41.05	28.7	29.6
AVG utterance length	8.3	7.95	7.98
Solo-performance	0.61	0.59	0.59
Group Performance	0.84	0.70	0.71
Agreement	0.91	0.82	0.83

Thanks to the Isaac Newton Trust and Cambridge University Press, we will be able to collect 360 more dialogues.

5. Annotation Findings

- Participants actively engage in Reasoning opposed to solution statements
- Inter-annotator agreement: 0.75 and 0.71 for the first 2 levels of annotation, and 0.53 for the additional labels
- Stronger correlation between Probing for reasoning and Success, compared to dialogue length and number of participants

Feature	Correlation
Probing-Reasoning	0.119
# of participants	0.056
Dialogue Length	0.077

6. The Future: Deliberation Enhancing Bots (DEliBots)

- Dialogue systems that participate in the group discussion in order to improve the group performance through deliberation
- The DEliBot doesn't know the answer to the task, but knows how to ask good probing questions

Let's talk about constructive discussions:

gmk34@cam.ac.uk