



university of
 groningen

Group Discussion, Myside Bias and Truth-Tracking

Edoardo Baccini¹

joint work with Zoé Christoff¹, Stephan Hartmann² and Rineke Verbrugge¹

¹Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen

²Munich Center for Mathematical Philosophy, LMU Munich

Contact: e.baccini@rug.nl

<https://edoardobaccini.wordpress.com/>



Group Discussion

Opinion Formation

Decision-Making

Scientific Research



Individual Reasoning

Rationality

Cognitive Biases (e.g. **Myside Bias**)



Group Discussion

Opinion Formation

Decision-Making

Scientific Research



Individual Reasoning

Rationality

Cognitive Biases (e.g. **Myside Bias**)

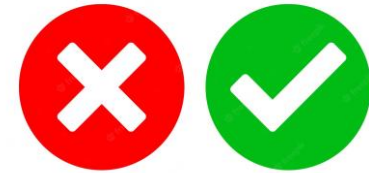


Group Discussion

Opinion Formation

Decision-Making

Scientific Research



Tracking the Truth

Correctly solving a binary
decision problem

Correctly answering a problem

Two (parallel) debates

Myside Bias and Truth-Tracking



What is the effect of myside bias on the group of agents' ability to track the truth?

Group Discussion and Truth-Tracking



What is the effect of group discussion on the group of agents' ability to collectively track the truth?

Two debates (I): Myside Bias and Truth-tracking

One's own prior beliefs influence one's **evaluation** (and production) of arguments.
(For an overview, Stanovich 2021)

Overestimation of confirming arguments

Underestimation of disconfirming arguments



Two debates (I): Myside Bias and Truth-tracking

One's own prior beliefs influence one's **evaluation** (and production) of arguments.
(For an overview, Stanovich 2021)

Overestimation of confirming arguments



Underestimation of disconfirming arguments

Overconfidence in previously held beliefs
regardless of their truth-value.



Two debates (I): Myside Bias and Truth-tracking

One's own prior beliefs influence one's **evaluation** (and production) of arguments.
(For an overview, Stanovich 2021)

Overestimation of confirming arguments

Underestimation of disconfirming arguments



Overconfidence in previously held beliefs
regardless of their truth-value.



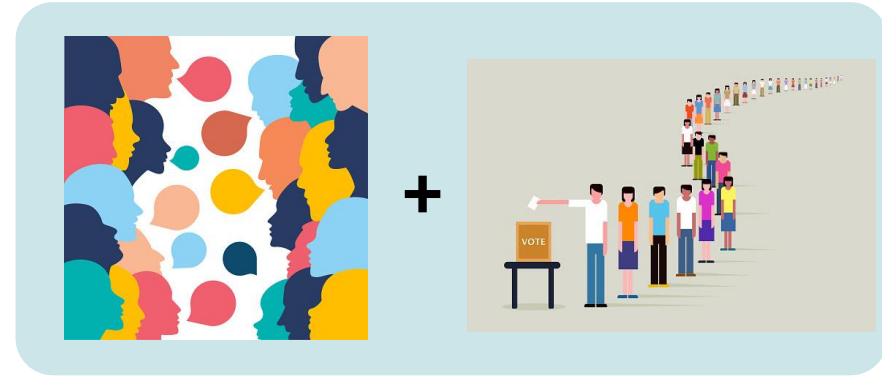
Detrimental effect on collective truth-
tracking

Enhancement of undesirable social
phenomena (polarization)



Two debates (II): Group Discussion and Truth-tracking

Does group discussion improve on the initial aggregate answers of a group of agents to a binary decision problem?



Two debates (II): Group Discussion and Truth-tracking

Group discussion is detrimental!

(e.g. Lorentz et al. 2011, Hahn et al. 2019)



Group discussion destroys the independence of the agents' opinions/votes.

Group discussion diminishes the diversity of the opinions of the agents.

Two debates (II): Group Discussion and Truth-tracking

Group discussion is beneficial!

(e.g. Hartmann and Rafiee Rad 2018, Mercier and Cladière 2022, Gabriel and O'Connor 2022)



Group discussion outperforms the aggregate answers of groups.

Why?

View 1: Myside Bias **Detrimental**

Prior-dependent argument evaluation



View 2: Myside Bias **Beneficial**

Coherence check, Stubbornness, Vigilance

View 1: Myside Bias Detrimental

Prior-dependent argument evaluation



Overconfidence in prior beliefs regardless of their truth-value.



View 2: Myside Bias Beneficial

Coherence check, Stubbornness, Vigilance

Avoiding falling prey of deceitful/false information.

View 1: Myside Bias Detrimental

Prior-dependent argument evaluation

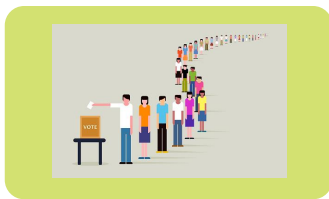


Overconfidence in prior beliefs regardless of their truth-value.



Detrimental effect on collective truth-tracking

Undesirable social phenomena (polarization)



Rabin and Schrag (1999), Stanovich (2021)

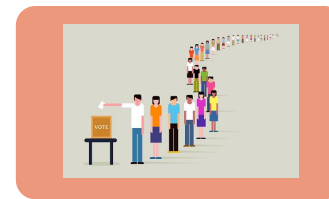
View 2: Myside Bias Beneficial

Coherence check, Stubbornness, Vigilance

Avoiding falling prey of deceitful/false information.

Cognitive division of labour between discussants at the opposite side of an issue.

Persistence of correct beliefs despite incorrect information.



Mercier and Sperber (2017), Gabriel and O'Connor (2022)

Overview

- A Bayesian Model of Myside-biased Argument Evaluation



- An Agent-Based Model (ABM) of Group Discussion with Myside-biased Agents



- Three Experiments



- Conclusion and Further Work

A Bayesian Model of Myside Bias Argument Evaluation



A Bayesian Model of Myside-biased Argument Evaluation: Bayesian Framework in a Nutshell

Belief: propositional random variables B

Argument: propositional random variable A

Prior belief: prior probability $P(B)$

Belief-update: Bayes' Update

$$\frac{P(B)}{P(B) + \frac{P(A|\neg B)}{P(A|B)} P(\neg B)}$$

Argument strength: likelihood ratio $\frac{P(A|\neg B)}{P(A|B)}$

Confirming argument (for B):

$$\frac{P(A|\neg B)}{P(A|B)} < 1$$

Disconfirming argument (for B):

$$\frac{P(A|\neg B)}{P(A|B)} > 1$$



A Bayesian Model of Myside-biased Argument Evaluation: The Myside Bias

Myside Bias: One's own prior beliefs influence one's evaluation of arguments.

(1) **Prior-dependent weighting:** over(under)-weighting confirming (disconfirming) arguments.

(2) **Neutrality:** do no overweighting or underweighting for neutral arguers.

(3) **Gradation:** stronger prior-dependent weighting, for arguer with stronger beliefs.



A Bayesian Model of Myside-biased Argument Evaluation: Modeling Myside Bias

Myside-biased update (*Bayes' rule on perceived argument strength*):

$$\frac{P(B)}{P(B) + x'(x, P(B)) P(\neg B)}$$

Perceived argument strength:

$$x'(x, P(B)) = \begin{cases} 2x \frac{P(\neg B)^\gamma}{P(B)^\gamma + P(\neg B)^\gamma}, & \text{if } P(B) \geq \frac{1}{2} \\ \frac{x P(B)^\gamma + P(\neg B)^\gamma}{2 P(B)^\gamma}, & \text{otherwise} \end{cases}$$

where $x = \frac{P(A|\neg B)}{P(A|B)}$ and the *radicality parameter* $\gamma \in (0,1)$.



A Bayesian Model of Myside-biased Argument Evaluation: Modeling Myside Bias

Myside-biased update (*Bayes' rule on perceived argument strength*):

$$\frac{P(B)}{P(B) + x'(x, P(B)) P(\neg B)}$$

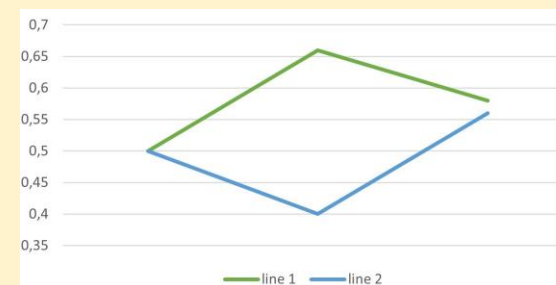
Perceived argument strength:

$$x'(x, P(B)) = \begin{cases} 2x \frac{P(\neg B)^\gamma}{P(B)^\gamma + P(\neg B)^\gamma}, & \text{if } P(B) \geq \frac{1}{2} \\ \frac{x P(B)^\gamma + P(\neg B)^\gamma}{2 P(B)^\gamma}, & \text{otherwise} \end{cases}$$

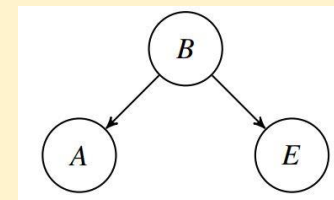
where $x = \frac{P(A|\neg B)}{P(A|B)}$ and the *radicality parameter* $\gamma \in (0,1)$.

What we get:

- **Prior-dependent weighting, neutrality and gradation.**
- **Bayes' Rule** for $\gamma = 0$.
- The order of arguments counts!



- **Justification** within a Bayesian framework.



An Agent-Based Model of Discussion with Mysided Agents

An ABM of Group Discussion with Mysided Agents (available [here](#)): **The Setup**



n agents

a unique issue (propositional variable)

prior degrees of belief drawn from a uniform distribution

average prior degree of belief of the group strictly above 0.5

radicality γ homogeneously/heterogeneously distributed

An ABM of Group Discussion with Mysided Agents (available [here](#)): **The Discussion**



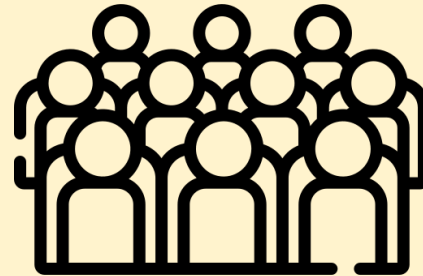
Discussion: sequence of argument exchanges

Argument Exchange

Randomly select one agent to be the **speaker**.



**Present an argument
confirming its own view**
Likelihood ratio drawn from a
distribution fixed at the start



Myside-biased Update
The other agents update
their degrees of belief

An ABM of Group Discussion with Mysided Agents (available [here](#)): **The Discussion**



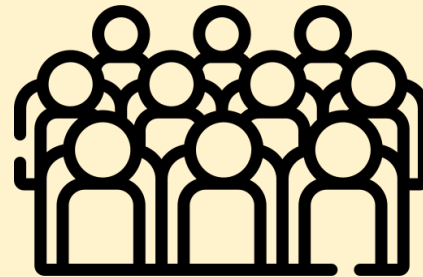
Discussion: sequence of argument exchanges

Argument Exchange

Randomly select one agent to be the **speaker**.



**Present an argument
confirming its own view**
Likelihood ratio drawn from a
distribution fixed at the start



Myside-biased Update
The other agents update
their degrees of belief



**Group consensus
or
Deep disagreement**

An ABM of Group Discussion with Mysided Agents (available [here](#)): **Monitoring Truth-Tracking**



1. What kind of discussion are the agents having? Effective/Ineffective exchange

An ABM of Group Discussion with Mysided Agents (available [here](#)): **Monitoring Truth-Tracking (Majority Rule)**



1. What kind of discussion are the agents having? Effective/Ineffective exchange

2. Is discussion between mysided agents beneficial or detrimental for a **majority** of agents to be correct?

Three Experiments

I

Experiment (I): Homogeneous Groups Uniform Radicality Distribution

- **Radicality γ** is the same for all agents (0 (no bias), 0.1, 0.2, 0.3, 0.5, 0.8)
- **Group sizes** (10,20,30,50,100,500)
- **10000** repetitions for each combinations of value of γ and group size

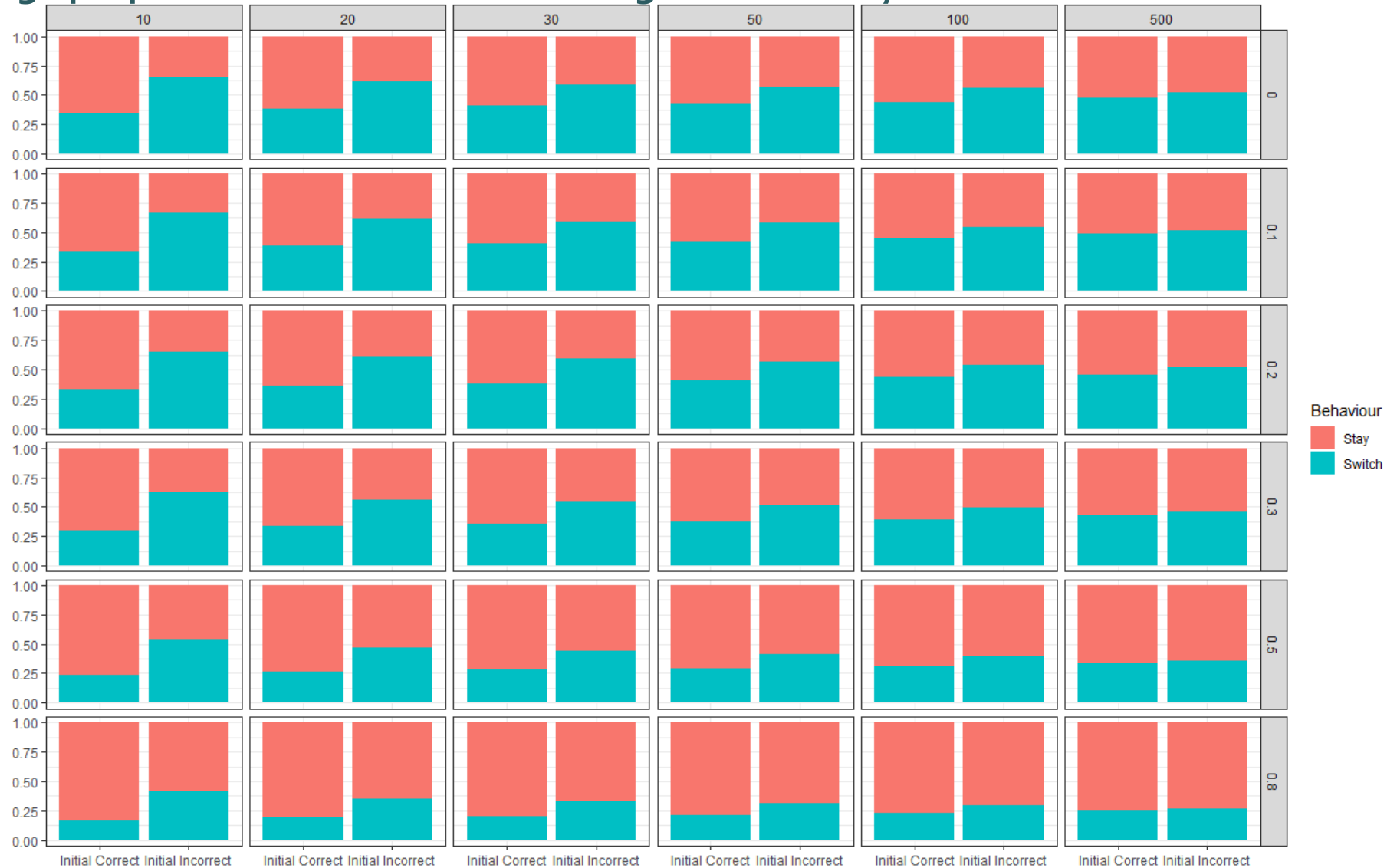
Experiment (I): Homogeneous Groups

Distribution of different consensus states



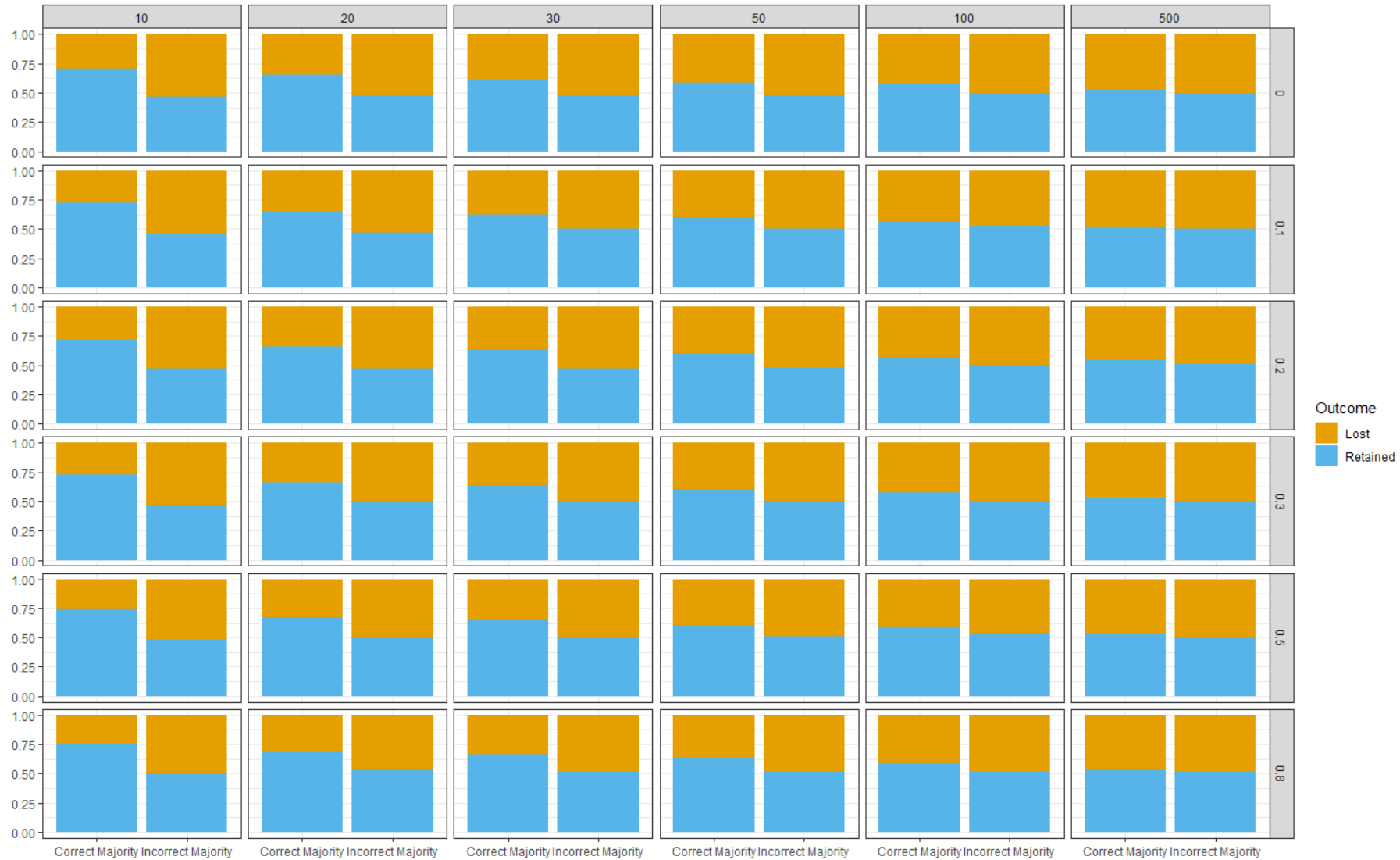
Experiment (I): Homogeneous Groups

Average proportion of correct/incorrect agents that stay in/switch from their initial belief



Experiment (I): Homogeneous Groups

Average proportion of correct/incorrect majority lost/retained after discussion

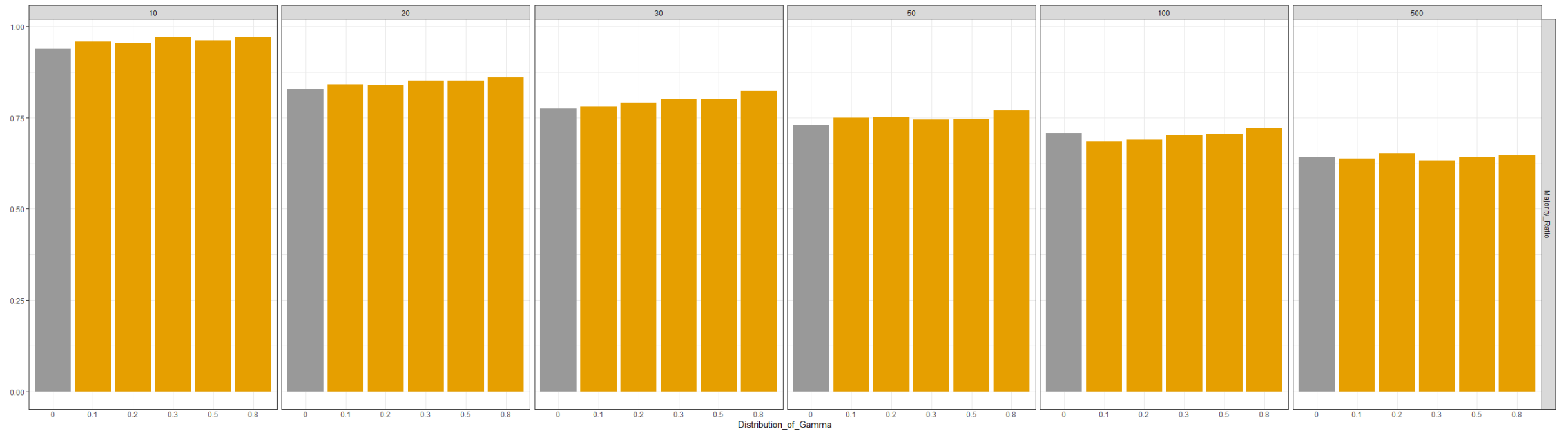


Experiment (I): Homogeneous Groups

Majority Ratio

Majority rate

$$\frac{n_{correct_maj_before_discussion}}{n_{correct_maj_after_discussion}}$$

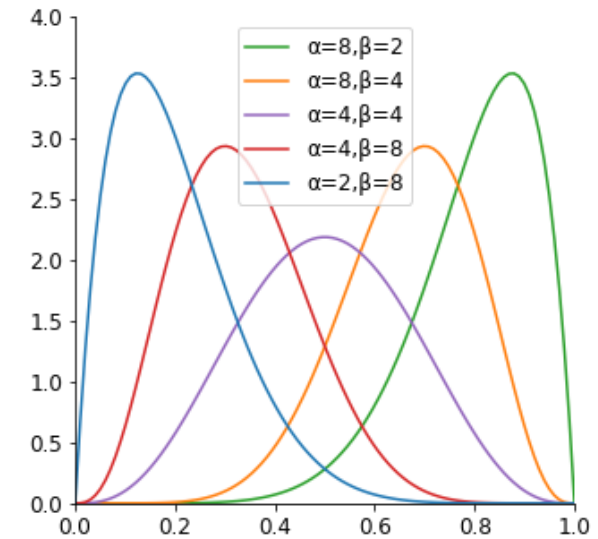


Three Experiments

II

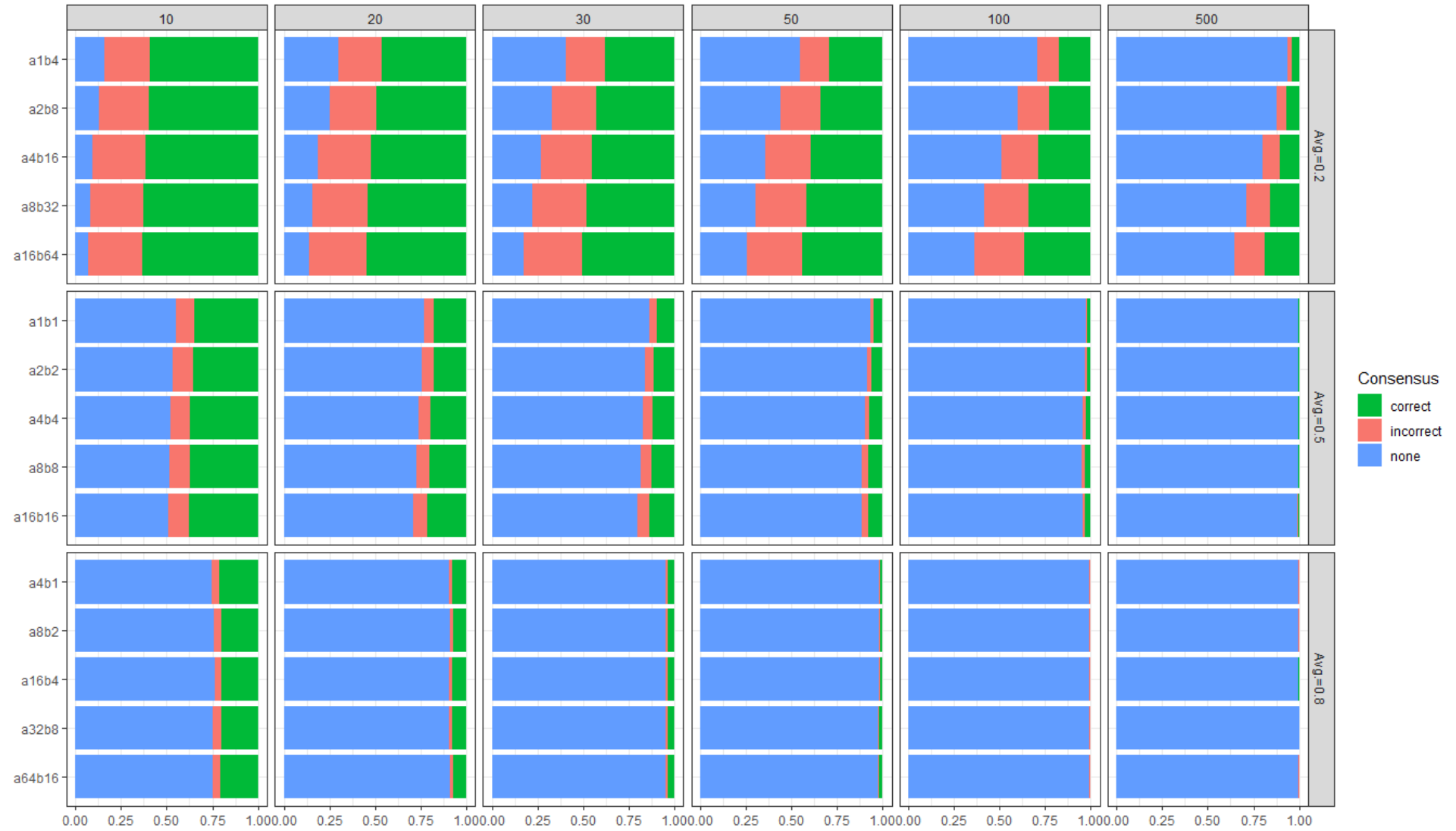
Experiment (II): Heterogeneous Groups Common Radicality Distribution

- **Radicality γ** is drawn from the same β –distribution for all agents
- **Group sizes** (10,20,30,50,100,500)
- **10000** repetitions for each combinations of value of γ and group size



Experiment (II): Heterogeneous Groups (Common radicality distribution)

Distribution of different consensus states (means of β -distributions (0.2, 0.5, 0.8))

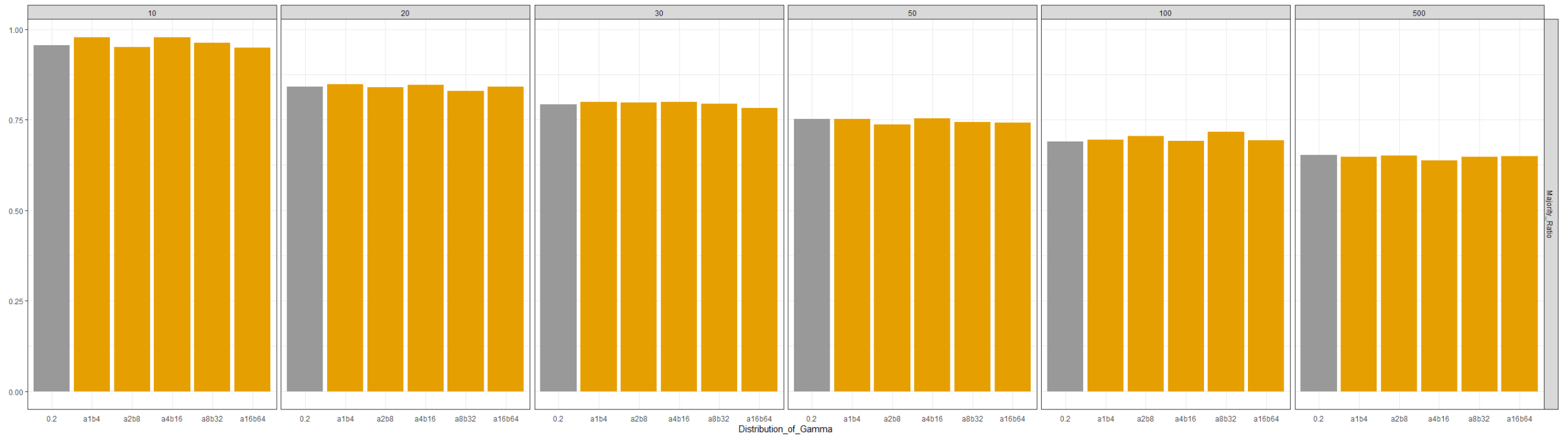


Experiment (II): Heterogeneous Groups (Common radicality distribution)

Majority rate (β –distributions with mean 0.2)

Majority rate

$$\frac{n_{correct_maj_before_discussion}}{n_{correct_maj_after_discussion}}$$

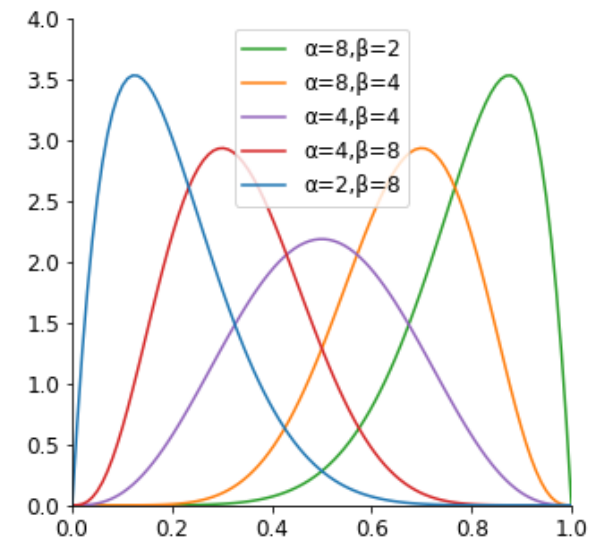


Three Experiments

III

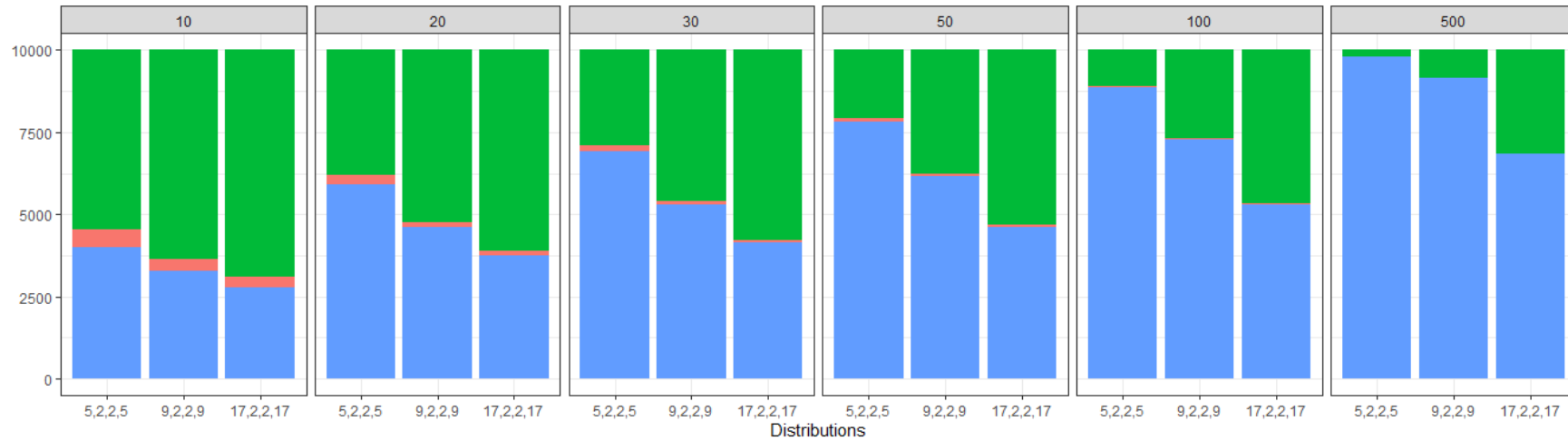
Experiment (III): Heterogeneous Groups (Common radicality distribution) Group-specific radicality distribution

- **Radicality γ** is drawn from two distinct β –distribution for **initially correct agents** and **initially incorrect agents**
- **Group sizes** (10,20,30,50,100,500)
- **10000** repetitions for each combinations of value of γ and group size



Experiment (III): Heterogeneous Groups (Common radicality distribution)

Distribution of different consensus states



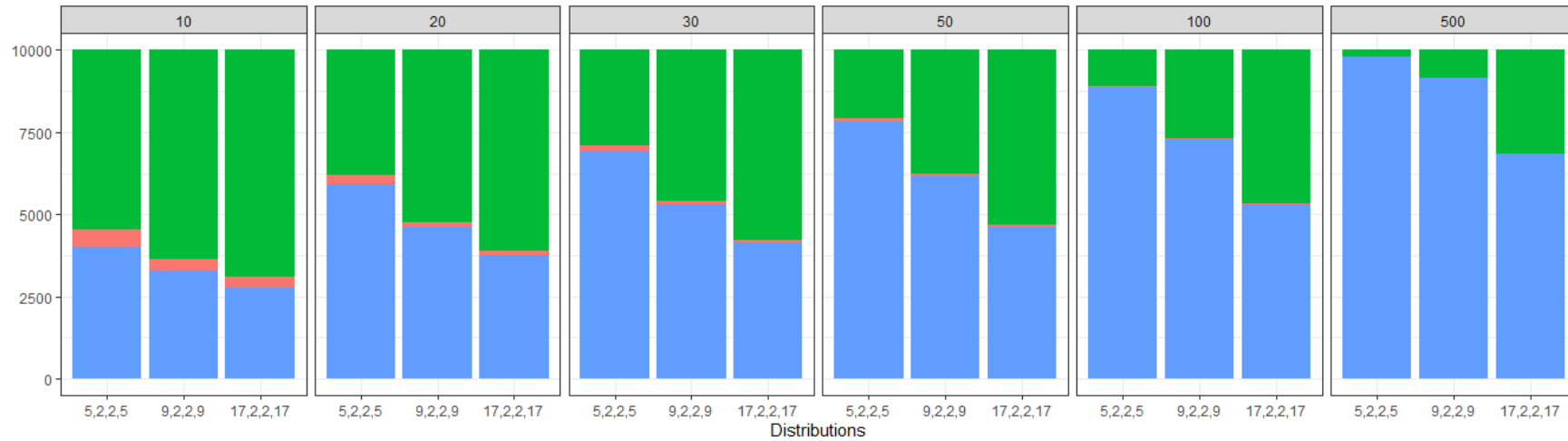
Initially correct agents are more radical

Consensus
 correct
 incorrect
 none

Pairs of distributions are denoted as:
 $\alpha_{correct_at_start}, \beta_{correct_at_start}$
 $\alpha_{incorrect_at_start}, \beta_{incorrect_at_start}$

Experiment (III): Heterogeneous Groups (Common radicality distribution)

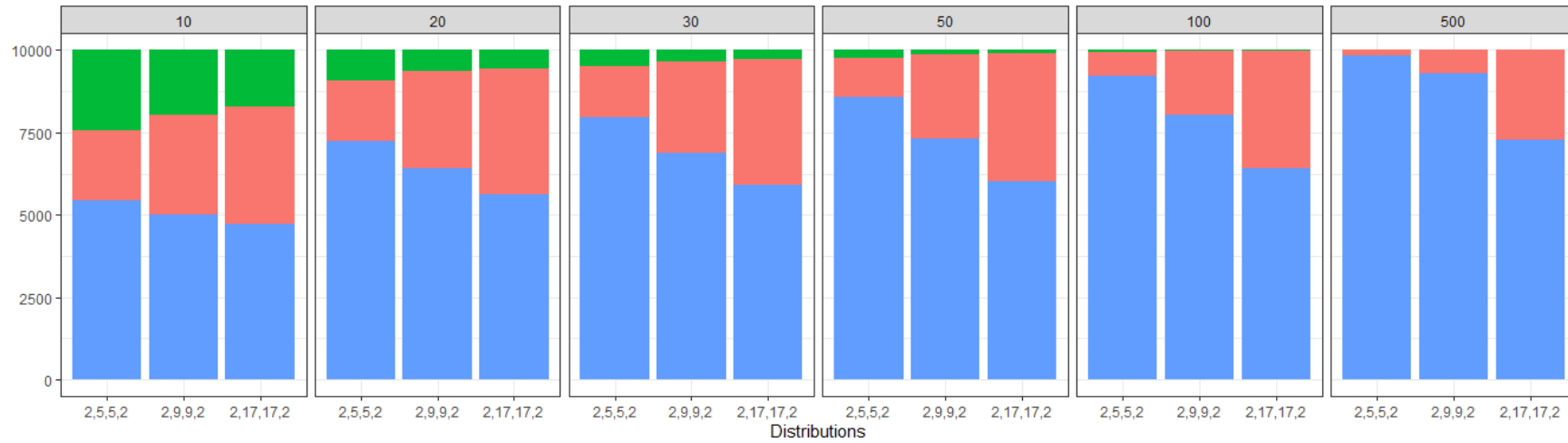
Distribution of different consensus states



Initially correct agents are more radical

Consensus
■ correct
■ incorrect
■ none

Pairs of distributions are denoted as:
 $\alpha_{correct_at_start}, \beta_{correct_at_start}$
 $\alpha_{incorrect_at_start}, \beta_{incorrect_at_start}$



Initially incorrect agents are more radical

Consensus
■ correct
■ incorrect
■ none

Experiment (III): Heterogeneous Groups (Common radicality distribution)

Average proportion of correct/incorrect agents that stay in/switch from their initial belief



Initially correct agents are more radical

Behaviour
■ Stay
■ Switch

Experiment (III): Heterogeneous Groups (Common radicality distribution)

Average proportion of correct/incorrect agents that stay in/switch from their initial belief



Initially correct agents are more radical

Behaviour
Stay
Switch

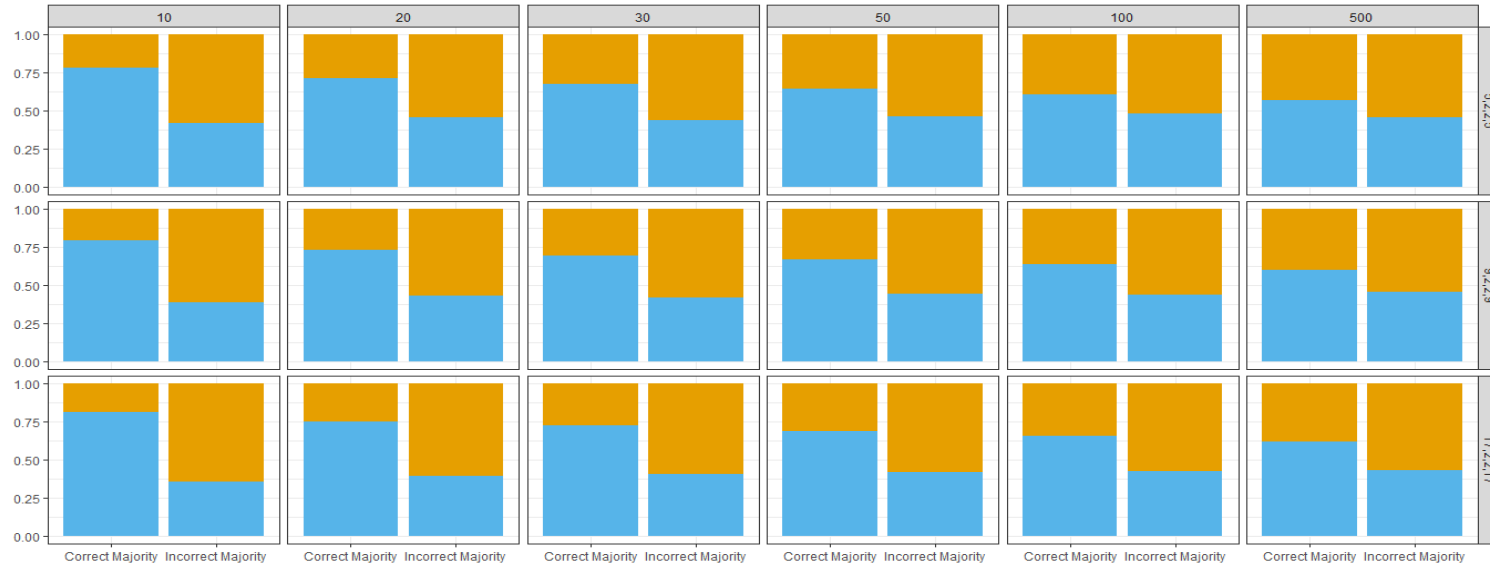


Initially incorrect agents are more radical

Behaviour
Stay
Switch

Experiment (III): Heterogeneous Groups (Common radicality distribution)

Average proportion of correct/incorrect majorities retained/lost after discussion

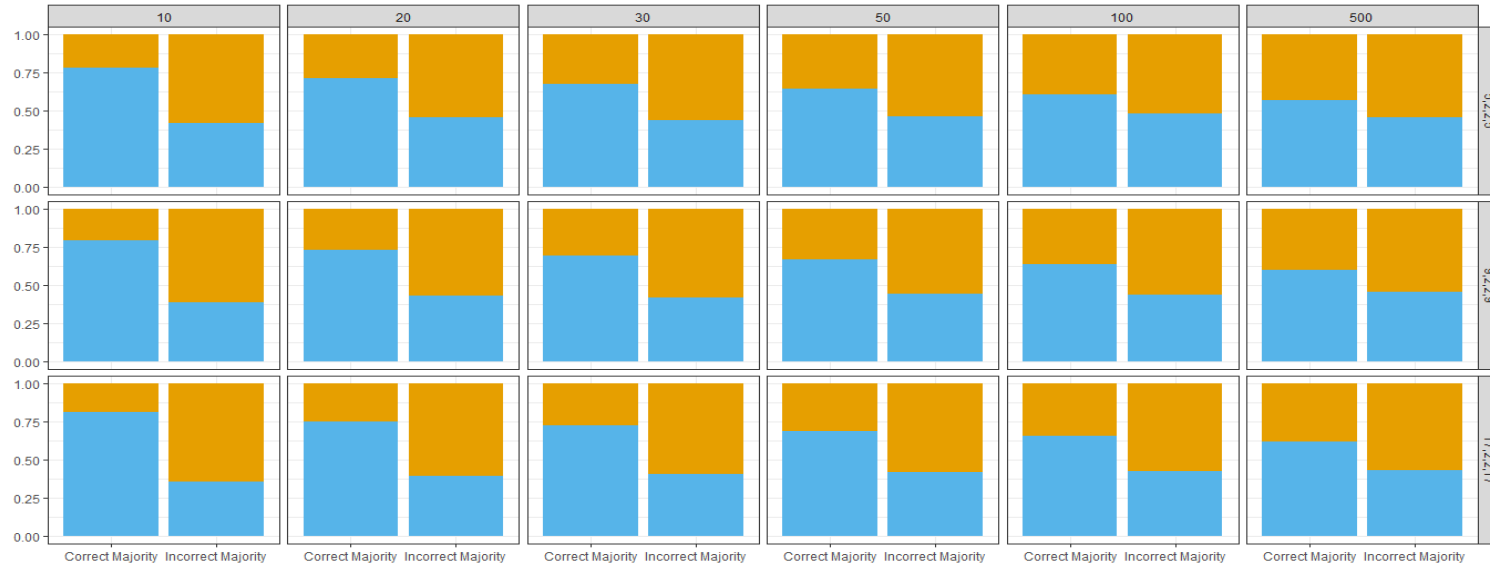


Initially correct agents are more radical

Outcome
Lost
Retained

Experiment (III): Heterogeneous Groups (Common radicality distribution)

Average proportion of correct/incorrect majorities retained/lost after discussion

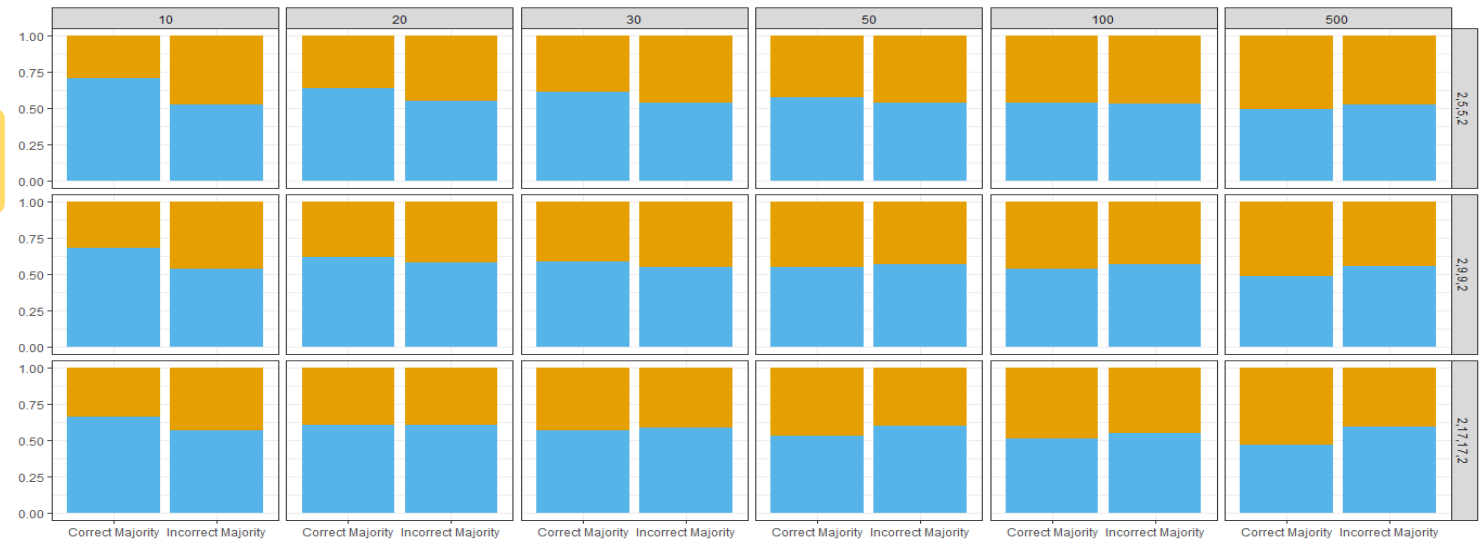


Initially correct agents are more radical

Initially incorrect agents are more radical

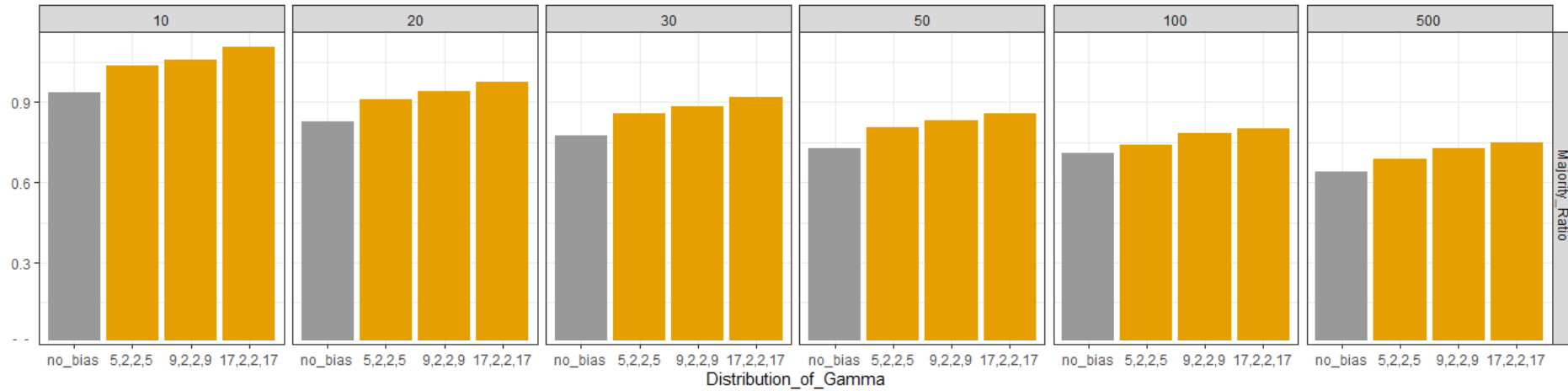
Outcome

- Lost
- Retained



Experiment (III): Heterogeneous Groups (Common radicality distribution)

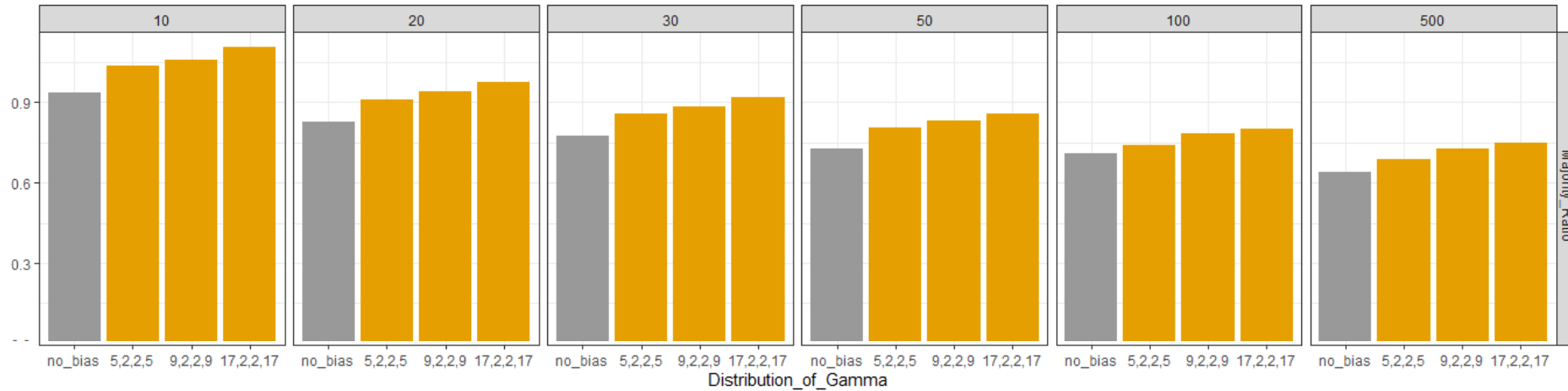
Majority Rate



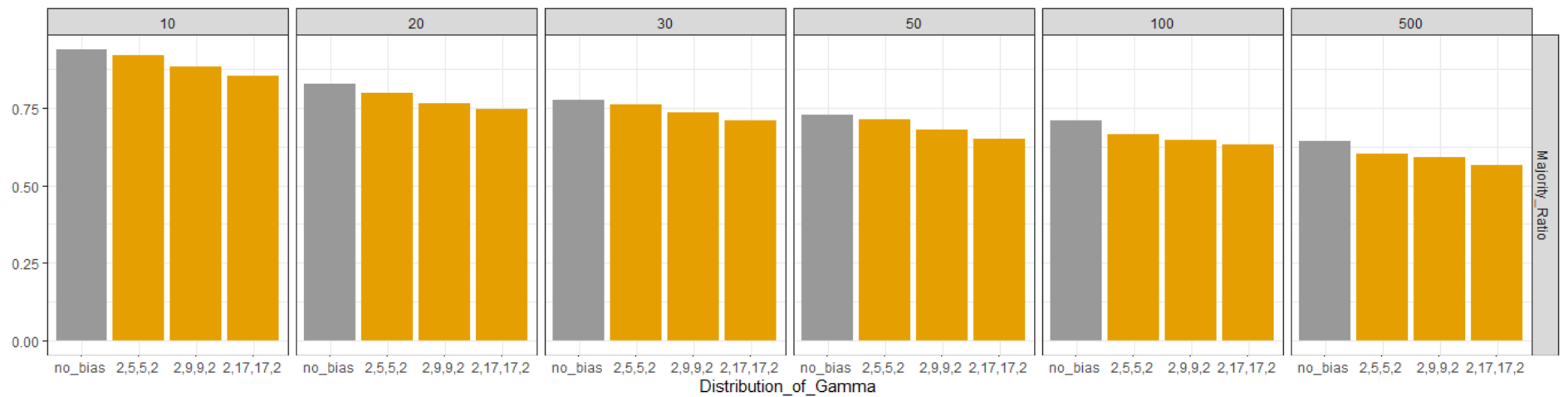
Initially correct agents are more radical

Experiment (III): Heterogeneous Groups (Common radicality distribution)

Majority Rate



Initially correct agents are more radical



Initially incorrect agents are more radical

Conclusions and Further Work

Summing up: The two views

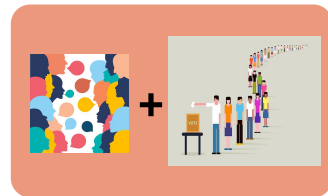
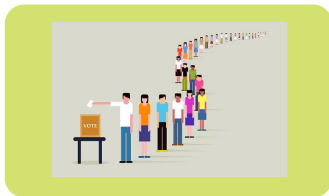
View 1

Prior-dependent argument evaluation

Overconfidence in previously held beliefs regardless of their truth-value.

Detrimental effect on collective truth-tracking

Enhancement of undesirable social phenomena (polarization)



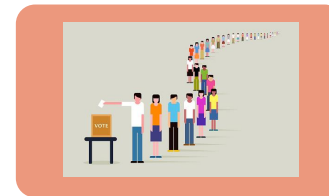
View 2

Coherence check, Stubbornness, Vigilance

Avoiding falling prey of deceitful/false information.

Cognitive division of labour between discussants at the opposite side of an issue.

Persistence of correct beliefs despite the diffusion of incorrect information.



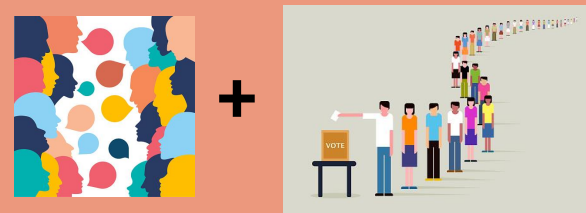
Summing up: Experiments (I),(II)

Increasing Radicality

- Inhibition of effective opinion-changing communication
- Neither harm, nor gain on collective wisdom (compared to no bias)

Increasing Group Size

- Inhibition of truth-conducive communication
- Inhibition of consensus-conducive communication
- Anchoring on wrong opinions

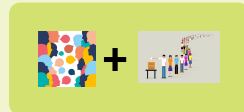


Summing up: Experiments (III)

(Differentially) Increasing Radicality

Among initially correct agents

Epistemic Benefits



Among initially incorrect agents

Epistemic Harm



Increasing Group Size

- Inhibition of truth-conducive communication
- Inhibition of consensus-conducive communication
- Anchoring on wrong opinions

Conclusions and Further Work:

Conclusions

- Increasing group size is detrimental to majority truth-tracking via discussion.
- Majority truth-tracking via discussion can be sensitive to myside bias, if subgroups of agents holding different beliefs differ in radicality.
- Overall, group discussion does not often improve aggregate collective answers.



Conclusions and Further Work:

Further Work

- Testing further on **group size effects** on discussion (network structures)
- Comparing truth-tracking abilities for different **communication protocols** (formal properties)
- Investigating groups where agents have **different competences** in producing arguments
- Implementing in the model aspects of **strategic communication**



university of
groningen

Group Discussion, Myside Bias and Truth-Tracking

Edoardo Baccini¹

joint work with Zoé Christoff¹, Stephan Hartmann² and Rineke Verbrugge¹

¹Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen

²Munich Center for Mathematical Philosophy, LMU Munich

Contact: e.baccini@rug.nl

References

- Gabriel, N. & O'Connor, C. (2022). Can confirmation bias improve group learning? *PhilSci-Archive*
- Hahn, U., von Sydow, M. & Merdes, C. (2019). How communication can make voters choose less well. *Topics in Cognitive Science*, 11(1), 194–206. doi:10.1111/tops.12401
- Hartmann, S. & Rafiee Rad, S. (2018). Voting, deliberation and truth. *Synthese*, 195(3), 1273–1293. doi:10.1007/s11229-016-1268-9
- Lorenz, J., Rauhut, H., Schweitzer, F. & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025. doi:10.1073/pnas.1008636108
- Mercier, H. & Claidière, N. (2022). Does discussion make crowds any wiser? *Cognition*, 222, 104912. doi:https://doi.org/10.1016/j.cognition.2021.104912
- Mercier, H. & Sperber, D. (2017). *The Enigma of Reason: A New Theory of Human Understanding*. Cambridge, MA: Harvard University Press
- Rabin, M. & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1), 37–82. doi:10.1162/003355399555945
- Stanovich, K. E. (2021). *The Bias That Divides Us: The Science and Politics of Myside Thinking*. Cambridge: MIT Press.

Images

- <https://www.istockphoto.com/nl/vector/mensen-profileren-hoofden-in-dialoog-vector-achtergrond-gm1313518767-401998773>
- <https://vectorified.com/download-image#thought-leadership-icon-17.png>
- <https://www.freepik.com/vectors/correct-incorrect>
- https://stock.adobe.com/nl/search?k=%22casting+vote%22&as_campaign=ftmigration2&as_channel=dpcft&as_campclass=brand&as_source=ft_web&as_camptype=acquisition&as_audience=users&as_content=closure_tag-page&asset_id=393346066
- <https://thenounproject.com/icon/speak-1064313/>
- https://www.flaticon.com/fr/icone-gratuite/public_5059287
- https://en.wikipedia.org/wiki/File:Nederlands_verkeersbord_B7.svg