

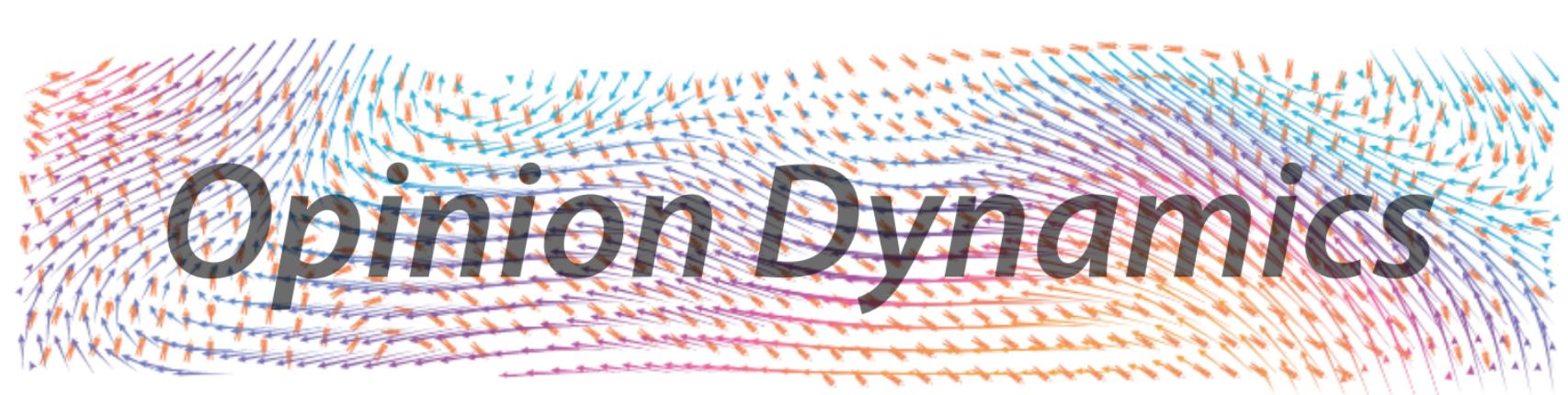
# *Validating argument-based opinion dynamics with survey experiments*

*Sven Banisch and Hawal Shamon*



# *Overview*

- ▶ *Opinion dynamics (a brief overview)*
- ▶ *Argument-based models (basic setting)*
- ▶ *Two model extensions (biased processing and external news)*
- ▶ *Collective dynamics (with and without external perturbation)*
- ▶ *Model validation (micro assumptions and macro outcomes)*

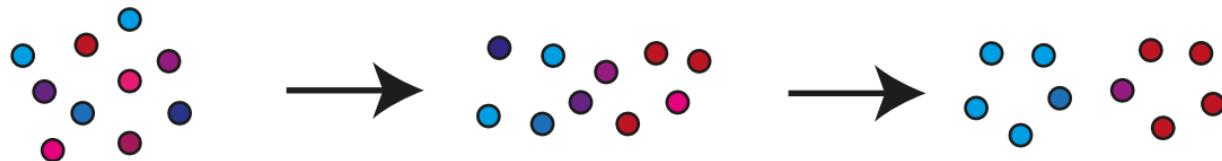


# Opinion Dynamics

*Opinion dynamics is a field that develops theoretical models of collective opinion processes to understand the mechanisms behind the emergence of consensus, polarization and conflict.*

# *Opinion Dynamics*

- ▶ Computational models for the evolution of opinions in a population of artificial agents
  - *N* artificial agents placed in a social information environment
  - These agents interact and exchange opinions
- ▶ Opinion dynamics studies the properties of these **complex dynamical systems** to understand basic mechanisms of consensus, polarization, media influence, etc.



*population at  $t-1$*

*population at  $t$*

*population at  $t+1$*

# *The Puzzle of Polarization*

»what on earth one must assume in order to generate the bimodal outcome of community cleavage studies?«

(Abelson, 1964)

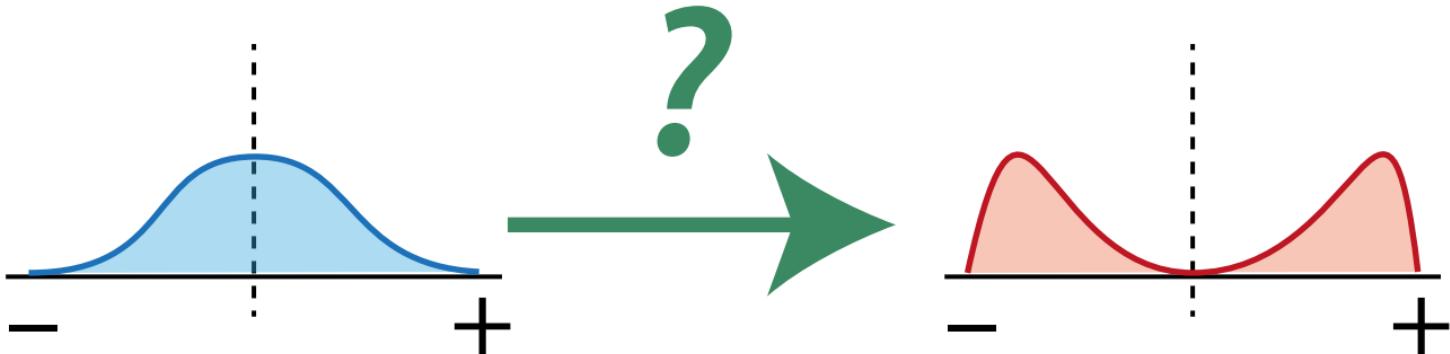
»If people tend to become alike in their beliefs, attitudes, and behavior when they interact, why do not all such differences eventually disappear?«

(Axelrod, 1997)

- These questions have inspired a lot of modeling work throughout the last two decades
  - This year is the 20th birthday of two prominent seminal contributions (Deffuant et al, 2002; Hegselmann & Krause, 2002)

# *The Puzzle of Polarization*

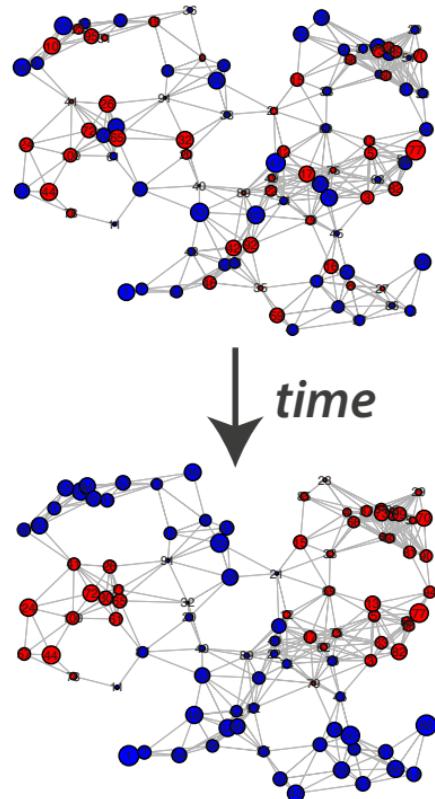
- ▶ How does a population with moderate initial opinions diverge into groups of agents that strongly support opposing views?



# *Binary Opinion Dynamics*

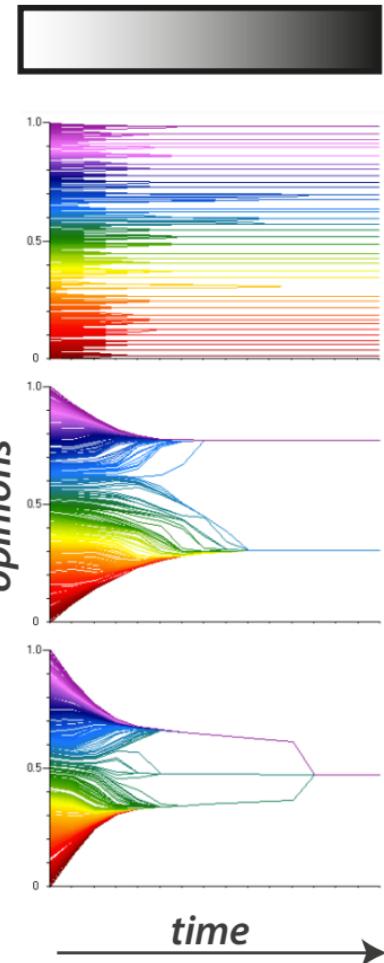


- Voter model, Sznajd model, Galam models
  - Opinions as Ising spins (yes/no)
  - Opinion change by imitation of a neighbor's state or adoption of majority opinion in the neighborhood
- Many theoretical results using tools from stochastic processes and statistical physics
  - Order-disorder transitions, convergence times, etc.
  - Key question: Influence of social network on spreading dynamics



# Continuous Opinion Dynamics

- ▶ Early consensus models (French 1956), social influence network theory (Friedkin & Johnsen, 2011)
  - Opinions as points on a continuous scale
  - Opinion change by weighted average of current opinion and the opinions in the neighborhood
- ▶ Bounded confidence models (Deffuant et al, 2002; Hegselmann & Krause, 2002; Flache et al, 2017)
  - **Opinion homophily:** »similarity leads to interaction and interaction leads to still more similarity«
  - Key question: Influence of extreme and stubborn agents, emergence of opinion landscapes

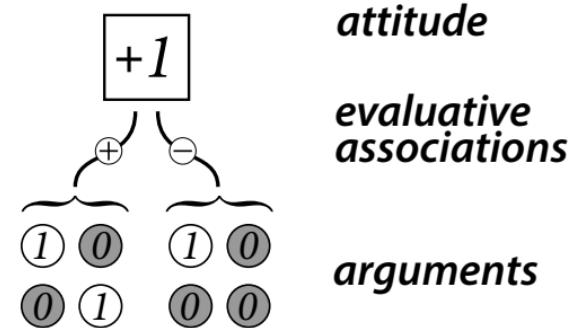


# *argument-based opinion dynamics*

*more complex & realistic  
representations of opinions*

# Argument Communication Theory

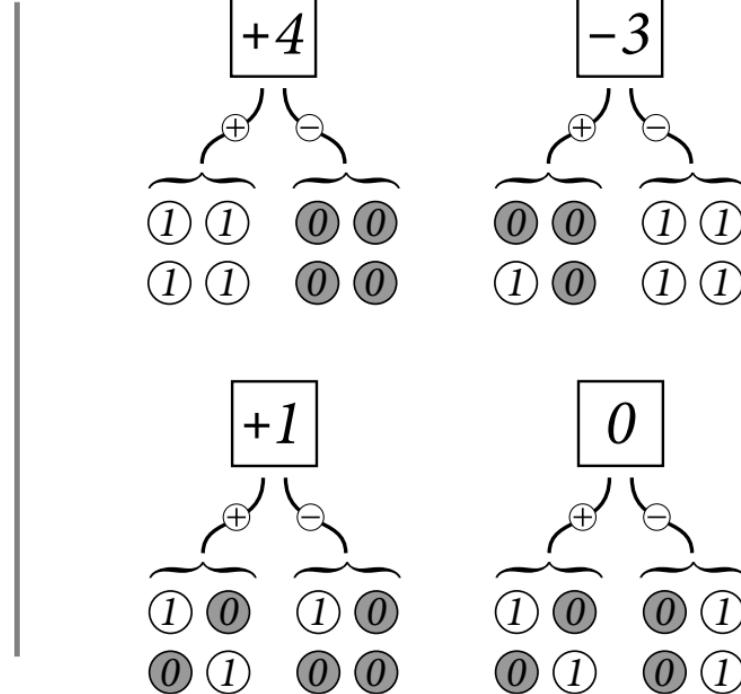
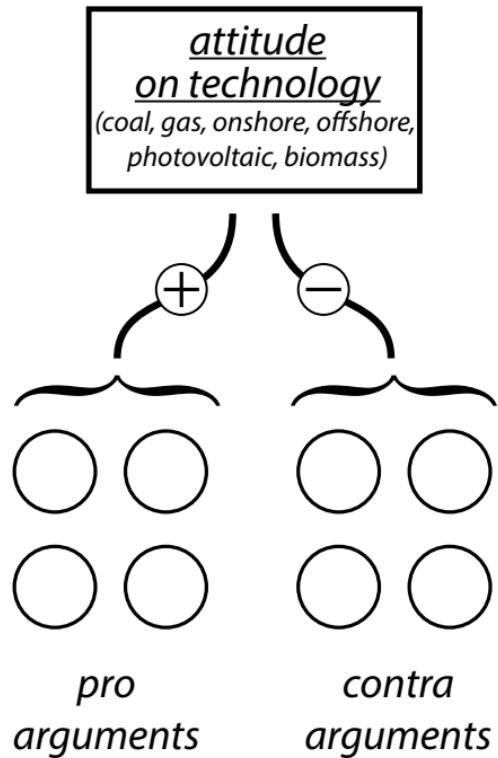
- Models of collective deliberation by social exchange of pro and con arguments regarding a certain issue (attitude object)
- Multidimensional & multilayered representation of opinions
- Agents communicate arguments and update them in memory
- Attitudes/opinions as number of pro versus con arguments are updated after peer exposure to an argument
- Basic ideas from attitude theories (cf. Banisch & Olbrich, 2021)



»different beliefs about any given object, i.e. many different characteristics attributes, goals, and objects are positively or negatively associated with a given object«  
(Fishbein 1963)

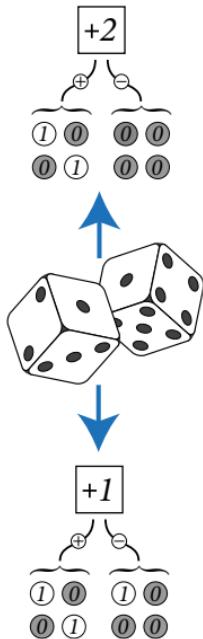
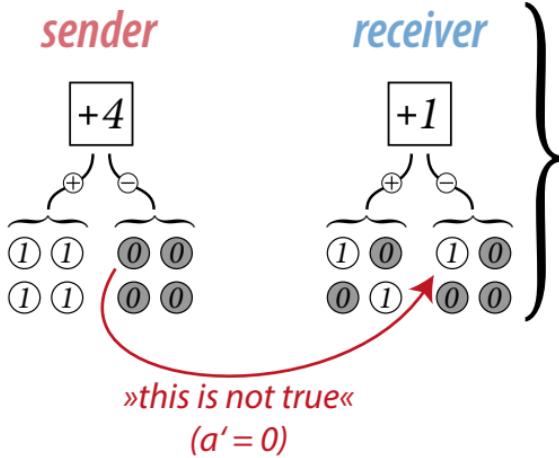
# Cognitive Model of Attitudes/Opinions

- Attitude defined as the number of pro versus con arguments



# Update Process

*social interaction*



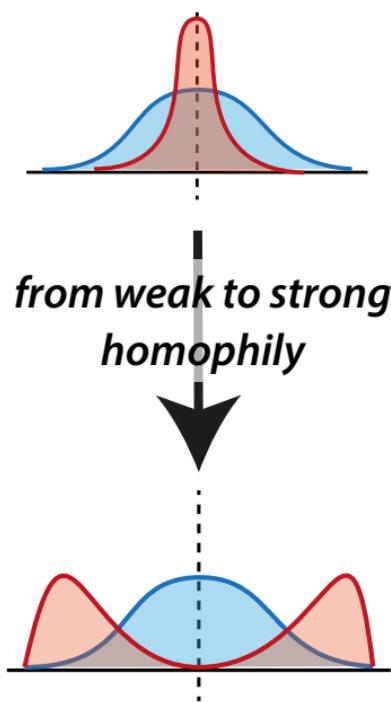
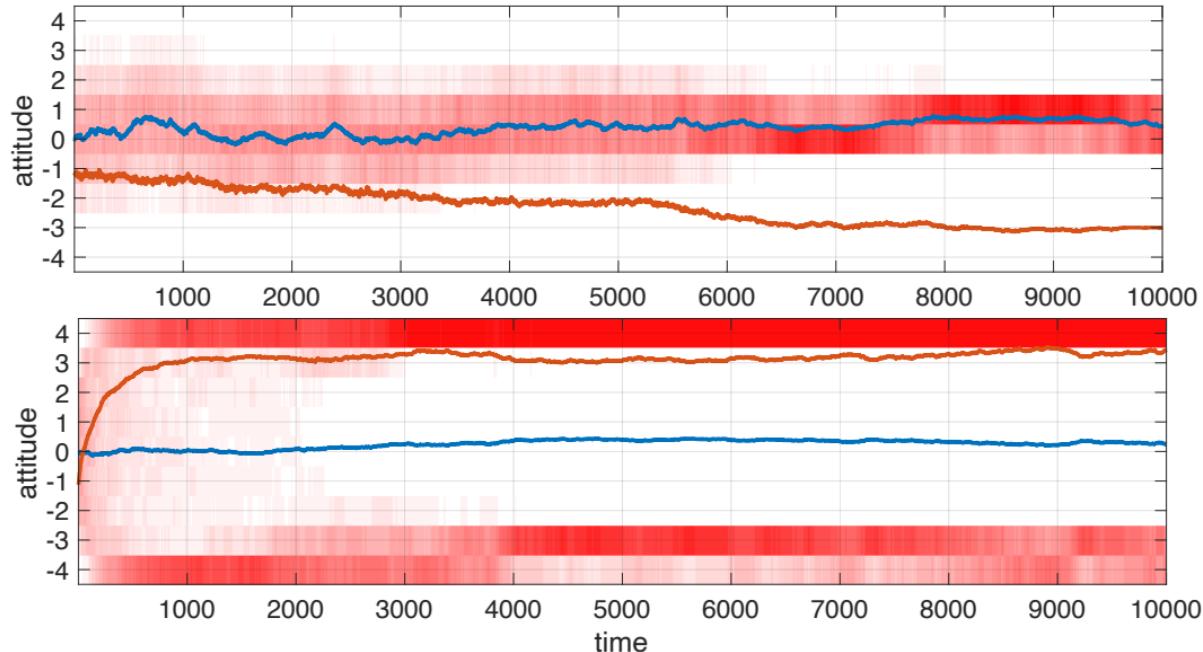
*cognitive processing*

1. A number ( $N$ ) of cognitive agents is paired at random.
2. The sender articulates a random argument to a receiver.
3. The receiver adopts this argument with probability  $1/2$  and updates its attitude accordingly.

► Paired with homophily at the attitude level: Attitude similarity drives interaction (Mäs/Flache, 2013)

# *Model Behavior: original model*

- ▶ Paired with homophily at the attitude level: Attitude similarity drives interaction (Mäs/Flache, 2013)



# *Argument Communication Theory*

## Online Demos

### UniVerseCity

#### Research – Curriculum – Papers – Demos

#### Complexity Science

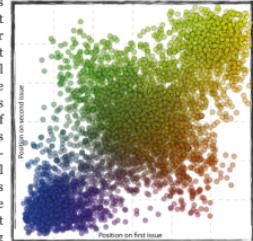
#### Interactive Exploration of Opinion Models

**What is the purpose of models?** Opinion dynamics is an exciting field which has given rise to the development of a considerable number of models by which agents (ex)change opinions in interaction. The number of different proposals is in fact hard to oversee and, that said, the question of how these models should be used is particularly relevant for the field. One can, on the one hand, develop models that contain as much as possible the different approaches in one encompassing tool in order to facilitate model comparison. On the other hand, models are mostly developed and articulated as addressing how particular assumptions on inter-individual influence processes play out at the macroscopic scales of groups or societies. In order to understand this connection from mechanisms to macroscopic outcomes, it is often more convenient to concentrate on a particular mechanism and provide (interactive) control on the main parameters of opinion

**www.universecity.de**  
**/index.php?site=demos**

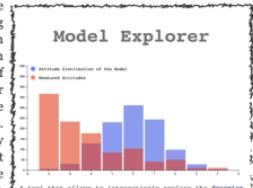


**From Attitude Structure to Political Spaces.** Arguments in a discussion often address different aspects of the issue at stake. But, some of these aspects are also relevant for other issues, which induces correlations between opinions on different issues. Those correlations could originate from factual interdependencies between the considered processes in the world, but they give also rise to ideologies and group identities which can induce further dependencies on their part. Many of the classical models of opinion dynamics studied in sociophysics are not able to address these issues. Drawing upon expectancy-value models in attitude research and the theory of conceptual spaces we developed a multi-level representation of opinions which allows to study of opinion dynamics on multiple interrelated issues. The model is based on three different ingredients: (1) interacting agents align their views regarding the significance of different argumentative domains; (2) different (partially overlapping) sets of these domains are associated with different political issues and an agent's attitude is a function of the importance assigned to the argument domains and their evaluative relevance for the issues; and (3) agents preferentially interact with other agents that hold similar attitudes. Under some conditions these combined processes give rise to polarization and reinforce correlations between attitudes towards multiple political issues. (See [arxiv](#))



1. **Argument Communication Model with 1000 agents** ([run it](#)) (This version accompanies a paper that will appear in JASSS. See the [brief guide to the model](#) for how to use the demo.)
2. **Argument Communication Model with 5000 agents** ([run it](#) - time is ok)

**Argument Communication with Biased Processing.** We look at data from an experiment on biased argument processing from the perspective of the cognitive architecture employed in argument communication models of collective opinion formation. The empirical experiment realized in the context of attitudes toward energy reveals a strong tendency to consider arguments aligned with the current attitude more persuasive and to downgrade those speaking against the current attitude. This is integrated into a theoretical model of cognitive agents by assuming that the coherence of an argument with the current attitude controls the probability to adopt it and to change the attitude accordingly. The strength of this bias is included as a free parameter which can be estimated from experimental data. We find a clear signature of moderate biased processing. This tool allows to explore simulations with interacting cognitive agents that exchange arguments. It relates the opinion distributions emerging in the model to surveyed attitude distributions (coal, gas, wind, photovoltaic, biomass). A good match is found transitory periods of the model and with unbiased external information.



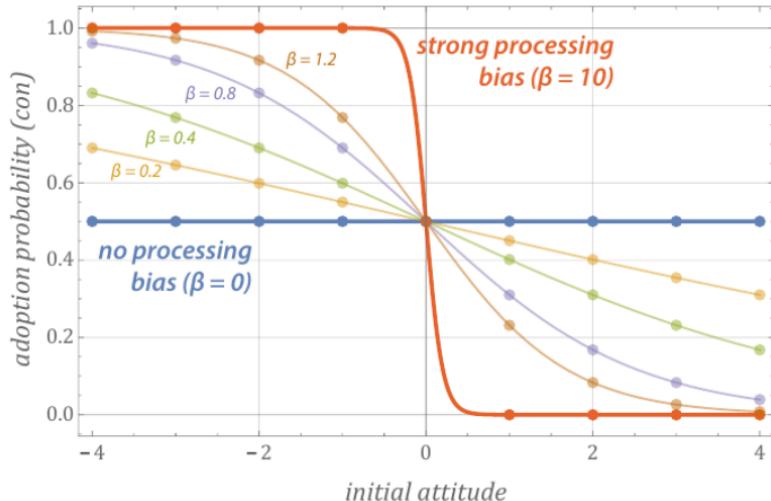
1. **Argument Communication Model with 1078 agents** ([run it](#))

*two model extensions*

*biased processing*  
&  
*unbiased news*

# First Extension: Biased Argument Processing

- Humans process information in a biased way: **attitude-congruent arguments are favoured over challenging ones!**
  - Argument adoption is assumed to depend on the evaluation  $V(a)$  and the strength of biased processing  $\beta$  by

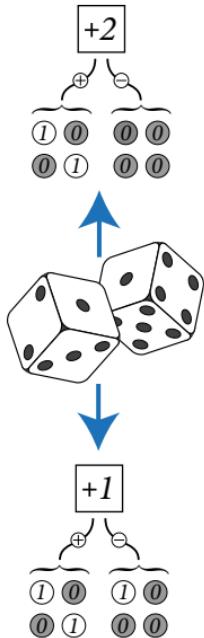
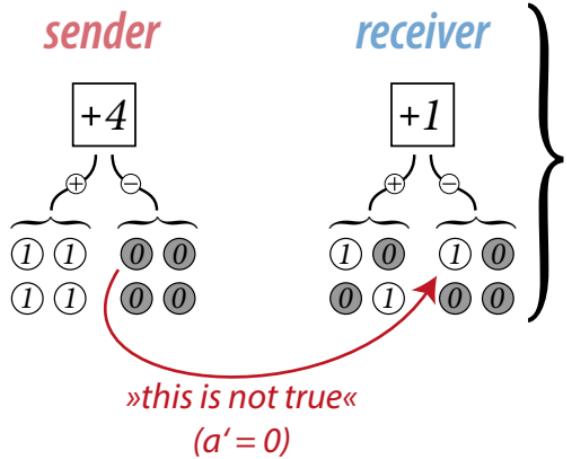


$$p_\beta(V(a'_i)) = \frac{1}{1 + e^{-\beta V(a'_i)}}$$

- The free parameter  $\beta$  accounts for the strength of biased processing
- $\beta = 0$  means unbiased adoption and if  $\beta$  is large only coherent arguments are adopted

# Update Process

*social interaction*



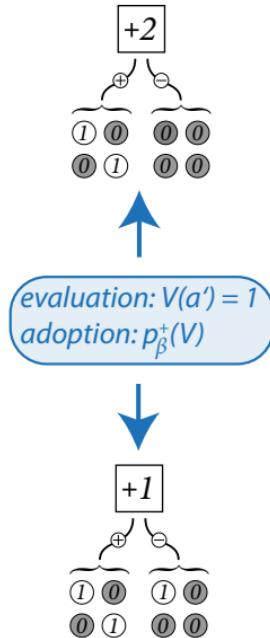
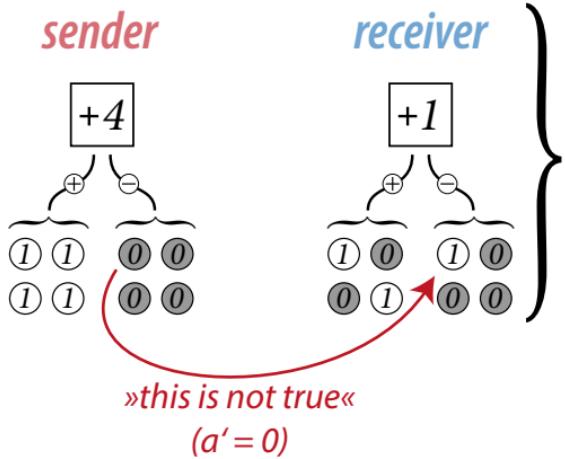
*cognitive processing*

1. A number ( $N$ ) of cognitive agents is paired at random.
2. The sender articulates a random argument to a receiver.
3. The receiver adopts this argument with probability  $p_\beta$  and updates its attitude accordingly.

► Consistent information/congruent arguments are adopted with higher probability (Banisch/Shamon, 2022)

# Update Process

*social interaction*



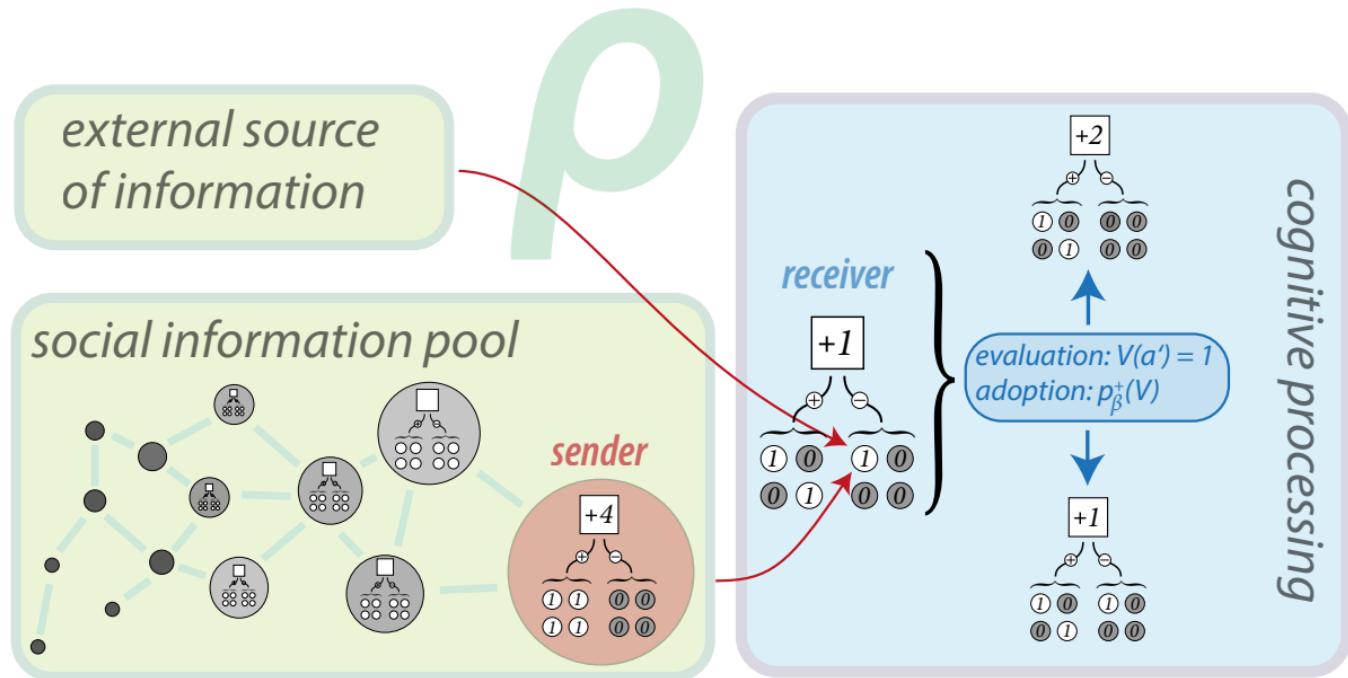
*cognitive processing*

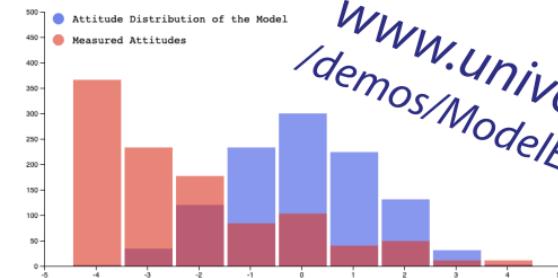
1. A number ( $N$ ) of cognitive agents is paired at random.
2. The sender articulates a random argument to a receiver.
3. The receiver adopts this argument with probability  $p_\beta$  and updates its attitude accordingly.

► Consistent information/congruent arguments are adopted with higher probability (Banisch/Shamon, 2022)

# Second Extension: Unbiased External News

- With a certain probability  $p$  agents receive a random argument from an unbiased external source



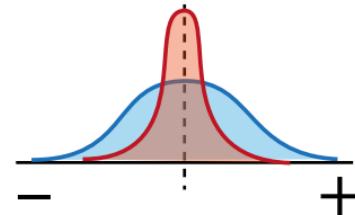


*www.universecity.de  
/demos/ModelExplorer.html*

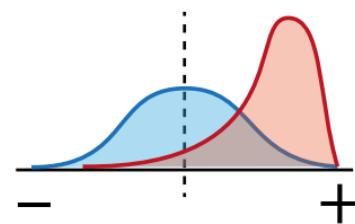
A tool that allows to interactively explore the dynamics of the argument communication model and to compare the evolving opinion distribution to real survey data.

# Collective Dynamics

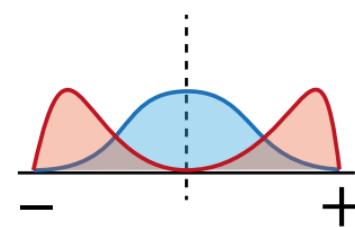
(without noise first)



*moderate  
consensus*



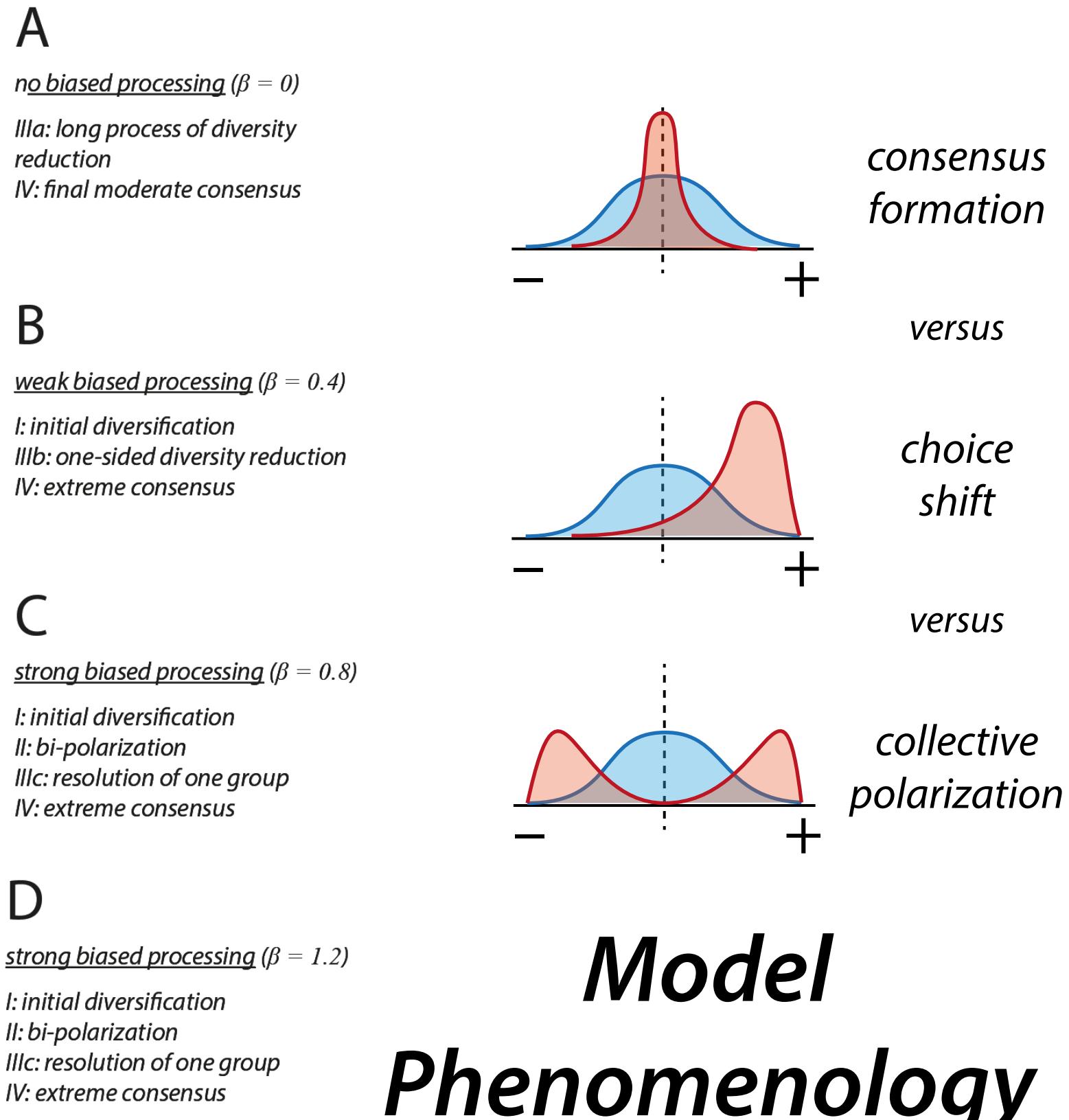
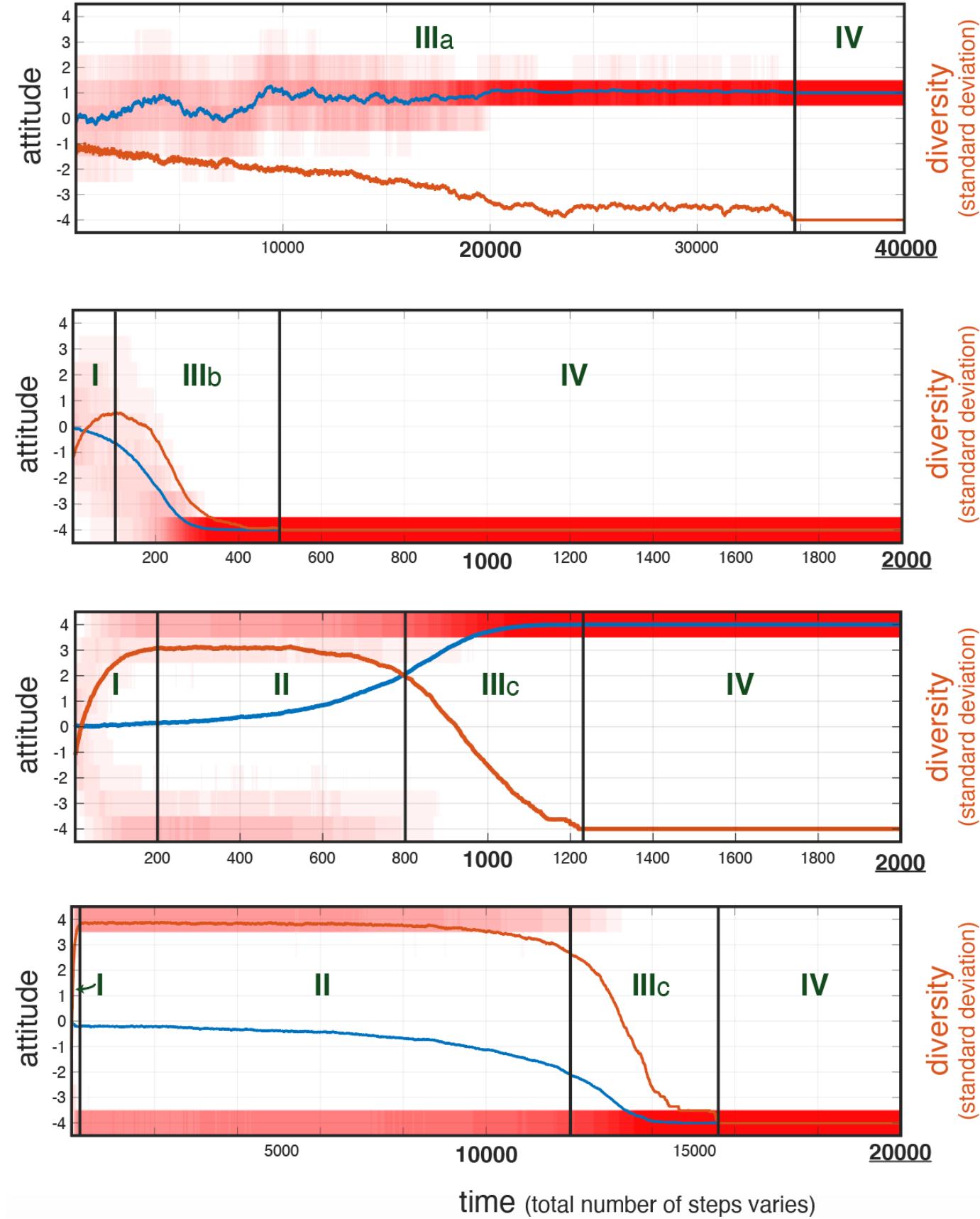
*choice  
shift*



*collective  
polarization*

*versus*

*versus*



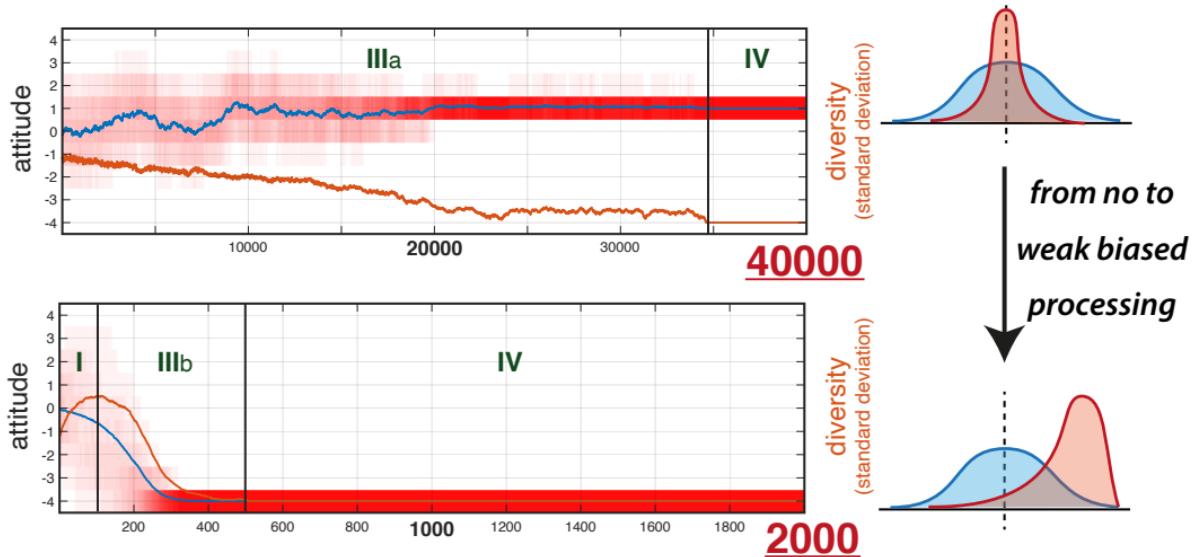
# Model Phenomenology

# First Transition

- *The introduction of weak biased processing speeds up group decision processes by two orders of magnitude*
  - Groups without processing bias may remain in indecision for a long time

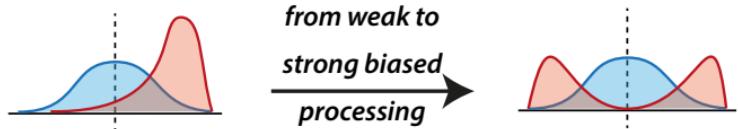
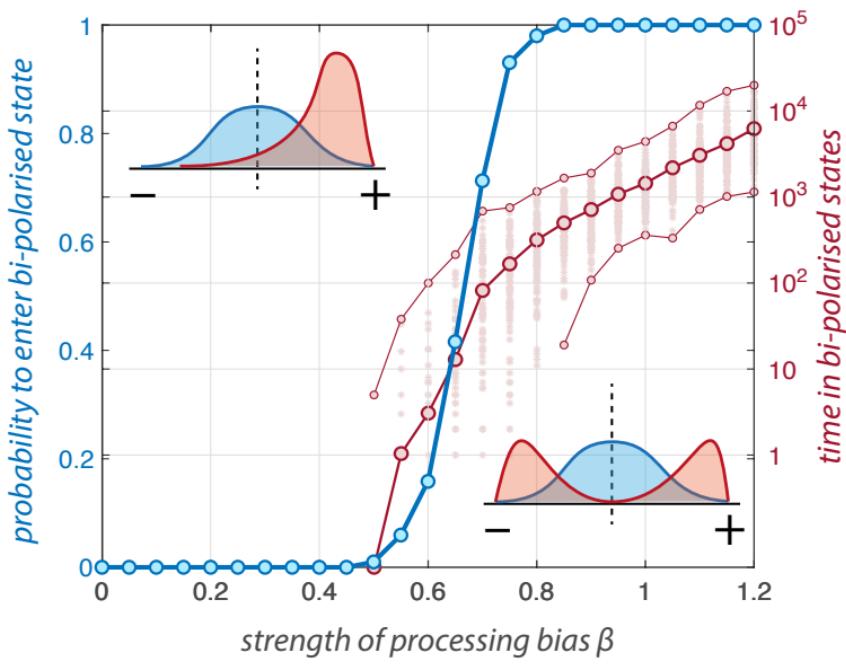
- Reinterpretation of „extreme consensus“ as an effective group decision process

- *Evolutionary origins of biased processing!?*



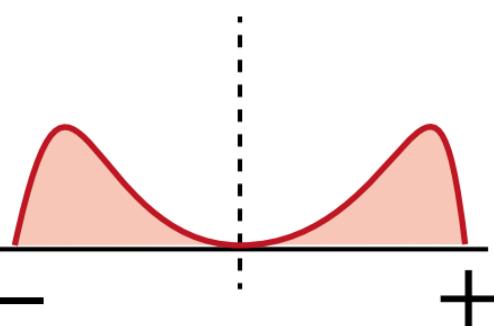
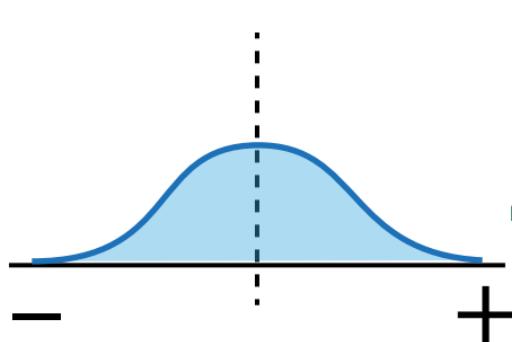
# *Second Transition*

- Strong biased processing may lead to persistent intra-group conflict
  - Biased processing alone is sufficient for collective bi-polarisation (without any assumptions about social composition)
  - Only in the regime of attitude polarisation at individual level
  - Strength of processing bias is what matters! (depends on topic)



# The Puzzle of Polarization

► Biased processing alone can lead to persistent bi-polarization if biased processing is strong



preprint (second revision)

Biased processing and opinion polarisation: experimental refinement of argument communication theory in the context of the energy debate

Sven Banisch<sup>1\*</sup> and Hawal Shamoun<sup>2</sup>

SSRN

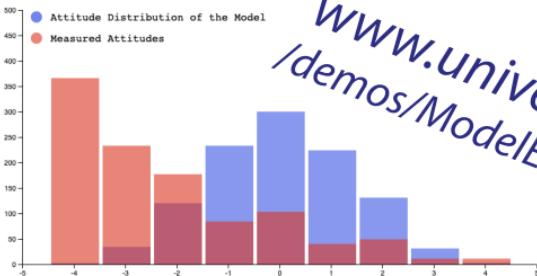
#### Abstract

We combine empirical experimental research on biased argument processing with a computational theory of group deliberation in order to clarify the role of biased processing in debates around energy. The experiment reveals a strong tendency to consider arguments aligned with the current attitude more persuasive and to downgrade those speaking against it. This is integrated into the framework of argument communication theory in which agents exchange arguments about a certain topic and adapt opinions accordingly. We derive a mathematical model that allows to relate the strength of biased processing to expected attitude changes given the specific experimental conditions and find a clear signature of moderate biased processing. We further show that this model fits significantly better to the experimentally observed attitude changes than the neutral argument processing assumption made in previous models. Our approach provides new insight into the relationship between biased processing and opinion polarisation. At the individual level our analysis reveals a sharp qualitative transition from attitude moderation to polarisation. At the collective level we find (i.) that weak biased processing significantly accelerates group decision processes whereas (ii.) strong biased processing leads to a persistent conflictual state of subgroup polarisation. While this shows that biased processing alone is sufficient for polarisation, we also demonstrate that homophily may lead to intra-group conflict at significantly lower rates of biased processing.

#### Keywords

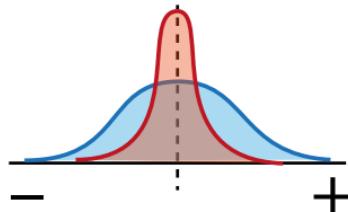
biased processing, attitude change, polarisation, experimental calibration, argument persuasion, group deliberation, opinion dynamics, energy debate

# *Collective Dynamics (with noise)*

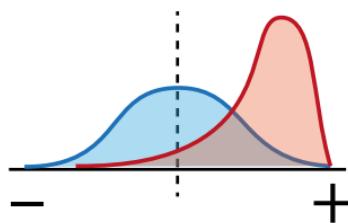


A tool that allows to interactively explore the dynamics of the argument communication model and to compare the evolving opinion distribution to real survey data.

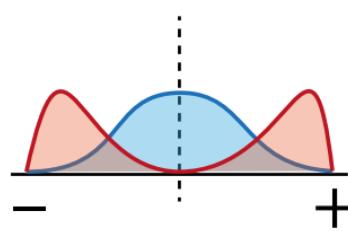
*www.universecity.de  
/demos/ModelExplorer.html*



*moderate  
consensus*



*choice  
shift*



*collective  
polarization*

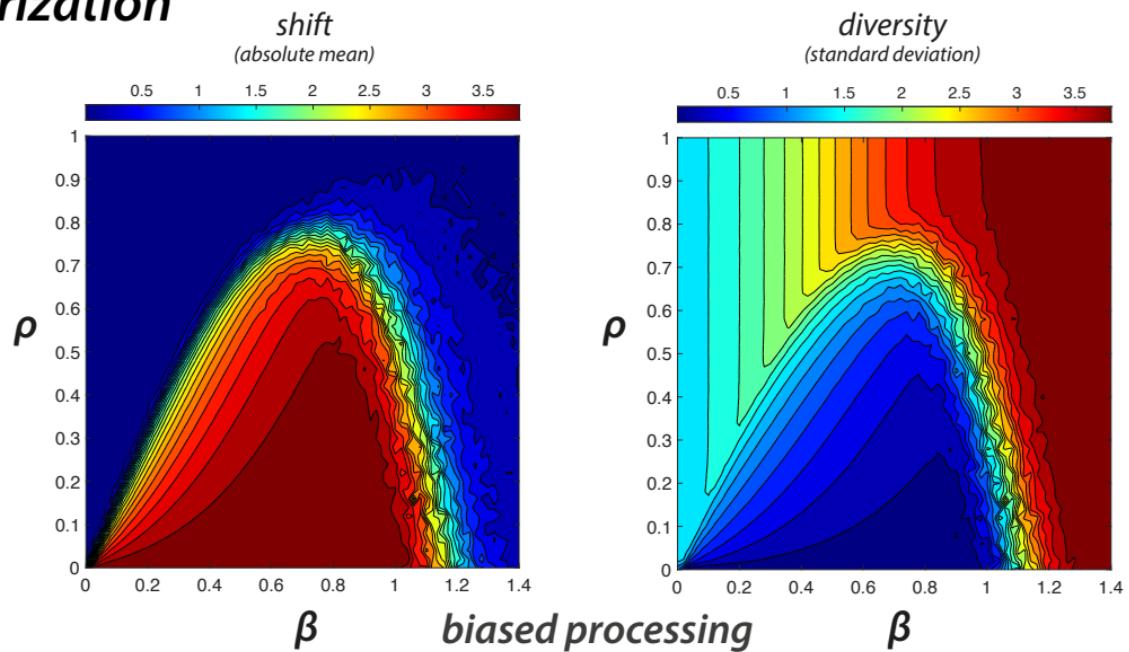
*versus*

*versus*

# *Systematic Computational Analysis*

- ▶ Characterization of emerging opinion distributions
  - to which extent they are **shifted toward one side** and the amount of **diversity or polarization**

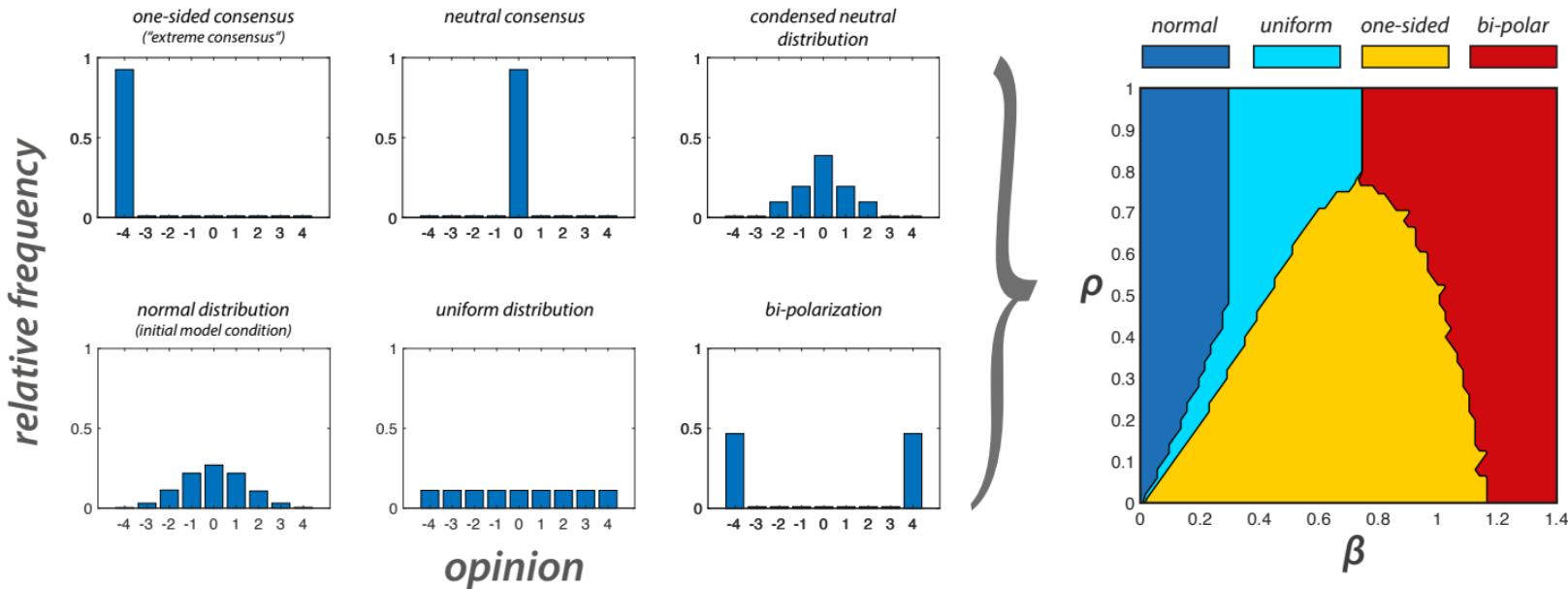
*impact of social influence  
versus external information*



# *Systematic Computational Analysis*

## ► Classification of different opinion regimes

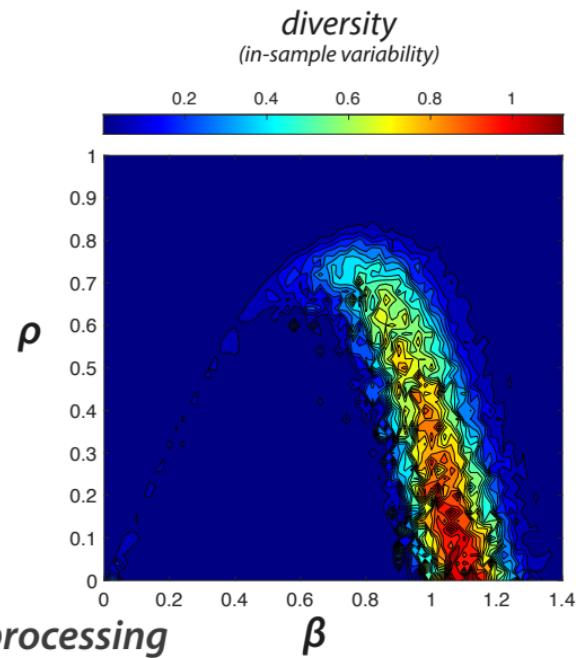
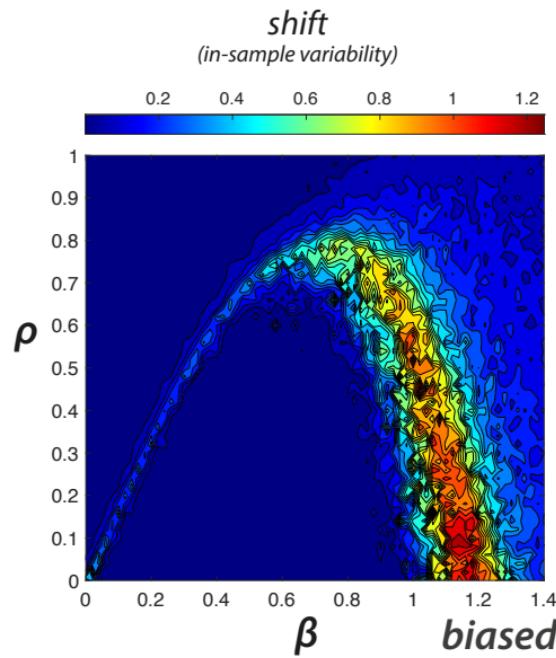
- comparing the outcome distributions to a series of stylized distributions using the Jensen-Shannon divergence



# *Systematic Computational Analysis*

- *High within-sample variance of the observables in the transition regions (based on 25 runs per sample point)*

*impact of social influence  
versus external information*



# *model validation*

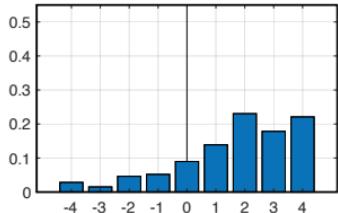
*micro assumptions*  
&  
*macro outcomes*

# Model Validation with Survey Experiments

## Macro Validation

(does the agent-based model generate opinion distributions that match survey data?)

opinion distribution for technology



## Computational Model

agent-based  
model of collective  
opinion formation

micro mechanisms  
of attitude change  
(influence-response  
function)

realistic  
representation of  
social interaction  
and informational  
environment

realistic  
representation of  
experimental  
treatment

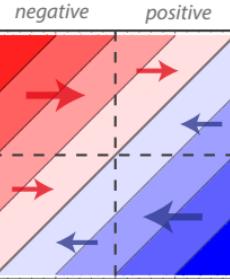
**Survey on opinions**  
(e.g., on different energy generating technologies, other control variable)

**different treatments**  
(exposure to balanced argument on focal energy technology)

## Micro Validation

(does the influence-response function fit data on individual attitude change?)

current opinion  
negative positive



message/information  
positive  
negative

**Post-treatment survey**  
(opinion on focal energy generating technology)

**Survey Experiment**

# Experiment

- Hawal Shamon carefully designed an experiment in which subjects received a **balanced set of pro and con arguments regarding six technologies for energy production.**



Copyright: Privat



Energy Research & Social Science  
Volume 55, September 2019, Pages 106-121



Original research article

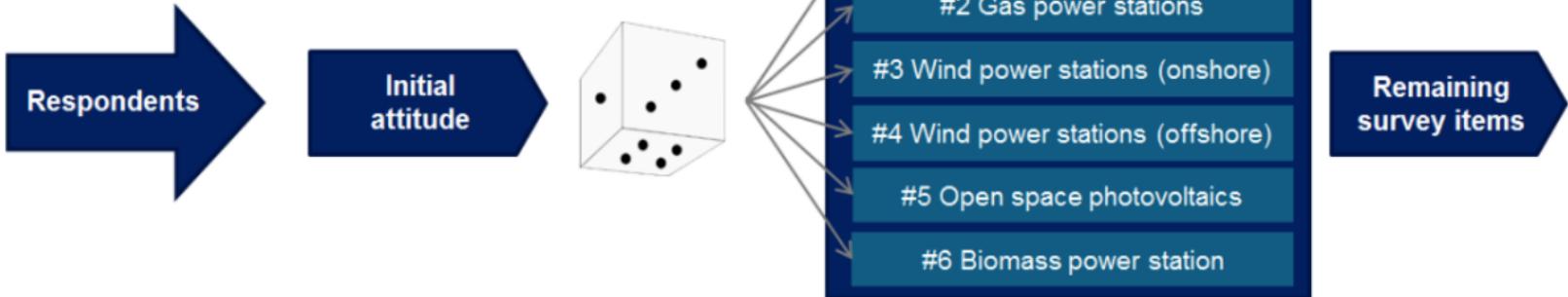
Changing attitudes and conflicting arguments: Reviewing stakeholder communication on electricity technologies in Germany

Hawal Shamon , Diana Schumann, Wolfgang Fischer, Stefan Vögele, Heidi U. Heinrichs, Wilhelm Kuckshinrichs

Institute of Energy and Climate Research – Systems Analysis and Technology Evaluation  
(IEK-STE), Forschungszentrum Jülich, Wilhelm-Johnen-Straße, Jülich 52425, Germany

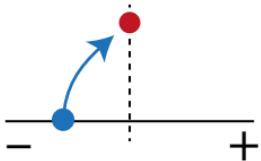
# Experiment

- ▶ Attitudes measured before and after exposure to 7 pro and 7 con arguments
- ▶ Survey data for 6 different energy sources ( $N = 1078$ )
- ▶ Attitude change data on these topics ( $N \geq 170$ )

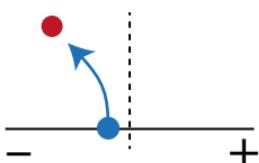


# The Psychological Puzzle

individuals  
(attitude change)  
(argument persuasion experiments)



attitude moderation  
versus  
attitude polarisation



- Humans process information in a biased way:  
**attitude-congruent arguments are favoured over incongruent ones!**
- A lot of experimental research has been invested on the question whether biased processing implies attitude polarisation when subjects are exposed to conflicting arguments.
- Attitude polarisation refers to an intra-individual effect of becoming more extreme after exposure
- **Empirical evidence is mixed!**

## *micro-level calibration*

*how well does the microscopic model of  
attitude change (i.e. biased processing) fit  
experimental data?*

# *Expected Attitude Change (virtual experiment)*

- ▶ How would artificial agents react to the same experimental treatment? What is their expected attitude change?

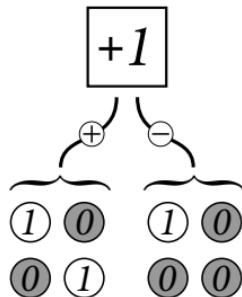
▶  $M = 4$  pro and con arguments (to match the 9-point attitude scale)

▶ How many new arguments are adopted?

▶ What is the respective attitude change?

*computational implementation  
(of balanced argument treatment)*

$T_1$



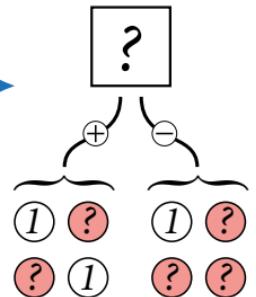
4 pro  
arguments

(1) (1)  
(1) (1)

4 con  
arguments

(1) (1)  
(1) (1)

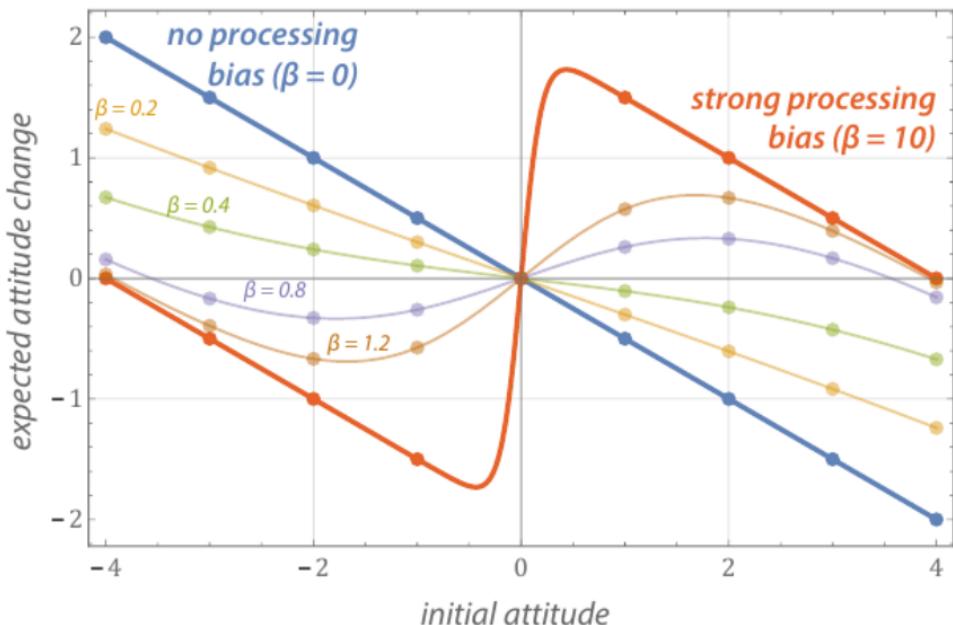
$T_2$



# *Expected Attitude Change (virtual experiment)*

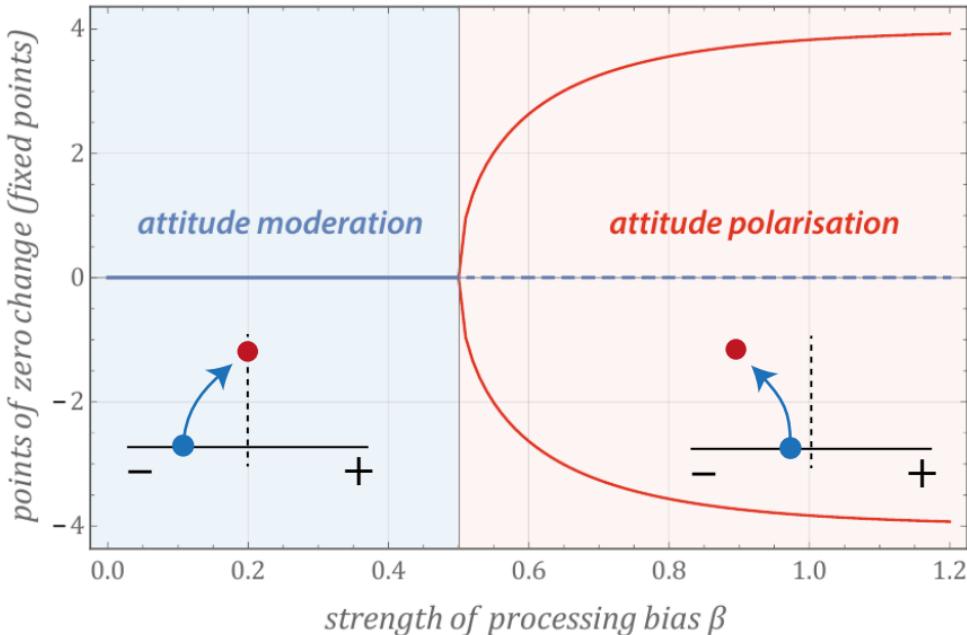
- ▶ **Attitude moderation** if bias  $\beta = 0$ . Negative opinions become more positive and vice versa.
- ▶ **This is what current models assume!**
- ▶ **Attitude polarisation** if  $\beta$  is large. Negative opinions become more negative and vice versa.

$$\mathbb{E}[\Delta o|o] = 2 \tanh\left(\frac{\beta o}{2}\right) - \frac{2o}{M}$$



# *The psychological puzzle: attitude moderation versus attitude polarisation*

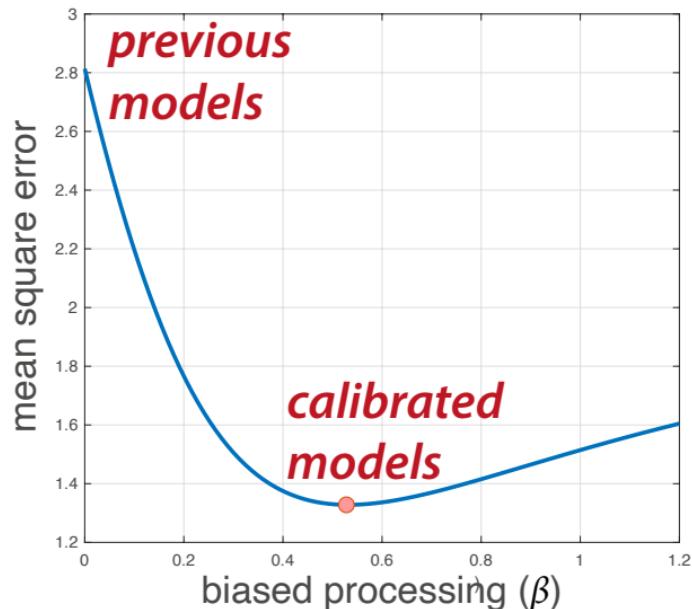
$$\mathbb{E}[\Delta o|o] = 2 \tanh\left(\frac{\beta o}{2}\right) - \frac{2o}{M}$$



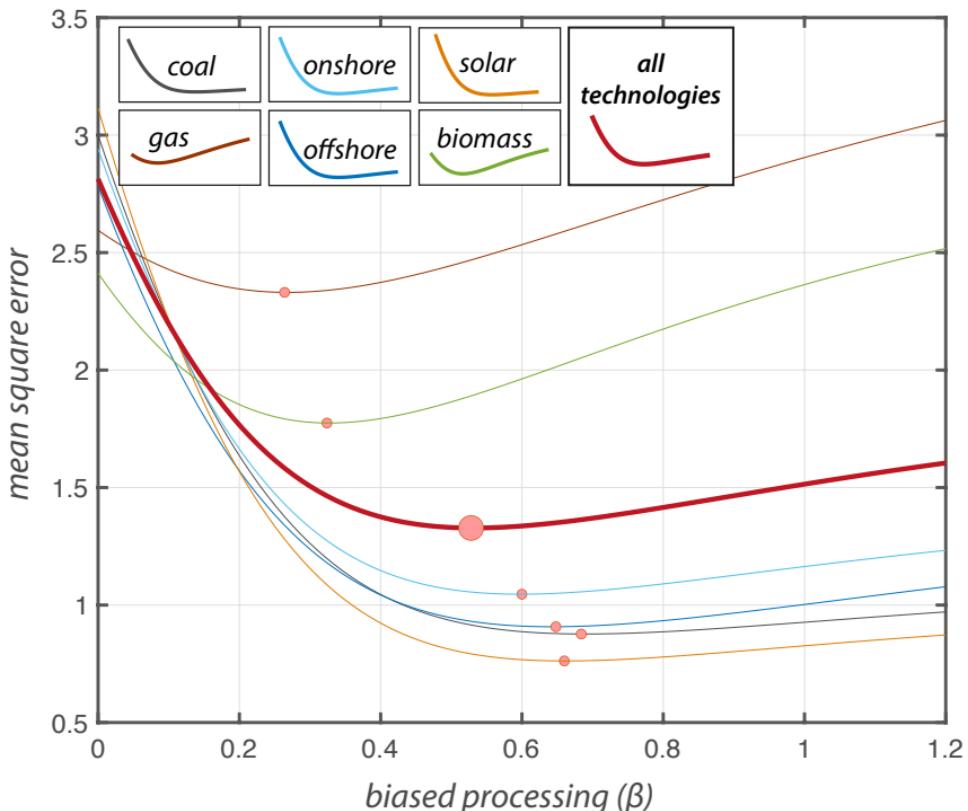
- ▶ **Bifurcation** from attitude moderation to attitude polarisation if  $\beta$  crosses a critical value of  $1/2$
- ▶ Biased processing leads to attitude polarisation only if the bias is strong
- ▶ This strength may vary across topics that experiments have addressed

# Calibration with Experimental Data

- ▶ Assuming that agents adapt by argument exchange (as in ACT models), what is the adequate choice for the processing bias  $\beta$ ?
- ▶ Least square error of theoretical prediction of mean attitude changes to the data ( $N = 1078$ )
- ▶ Best match for  $\beta$  close to the critical point 1/2!
- ▶ Clear signature that the refined model with moderate biased processing improves ACT



# Calibration: assessment of processing bias $\beta$



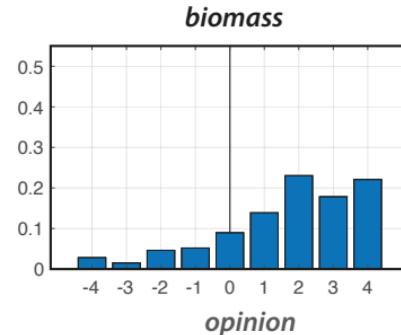
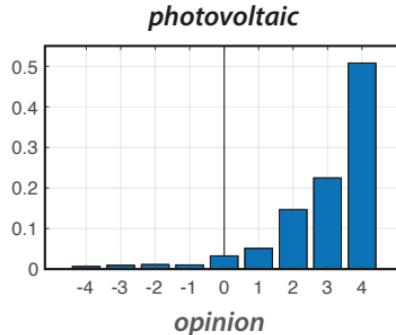
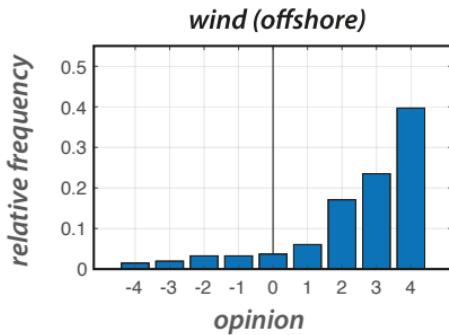
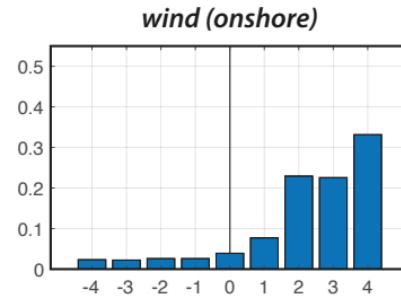
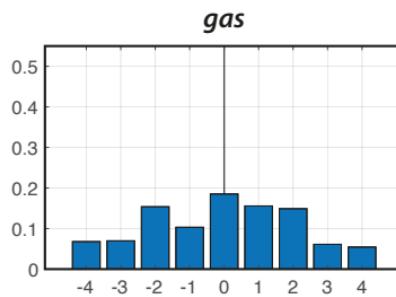
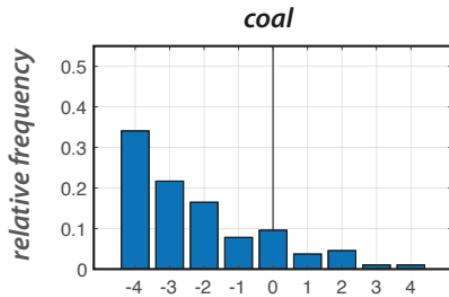
- Least square error of theoretical prediction of mean attitude changes to the data ( $N \approx 170$  per topic)
- Indication that bias is weak ( $< 0.5$ ) for new issues (biomass, gas) and stronger for long-debated issues ( $> 0.5$ )
- Attitude moderation in the former cases, attitude polarisation in the latter ones

# *macro-level fit*

*how well do model outcomes (i.e. opinion distributions) fit the surveyed distributions of opinions?*

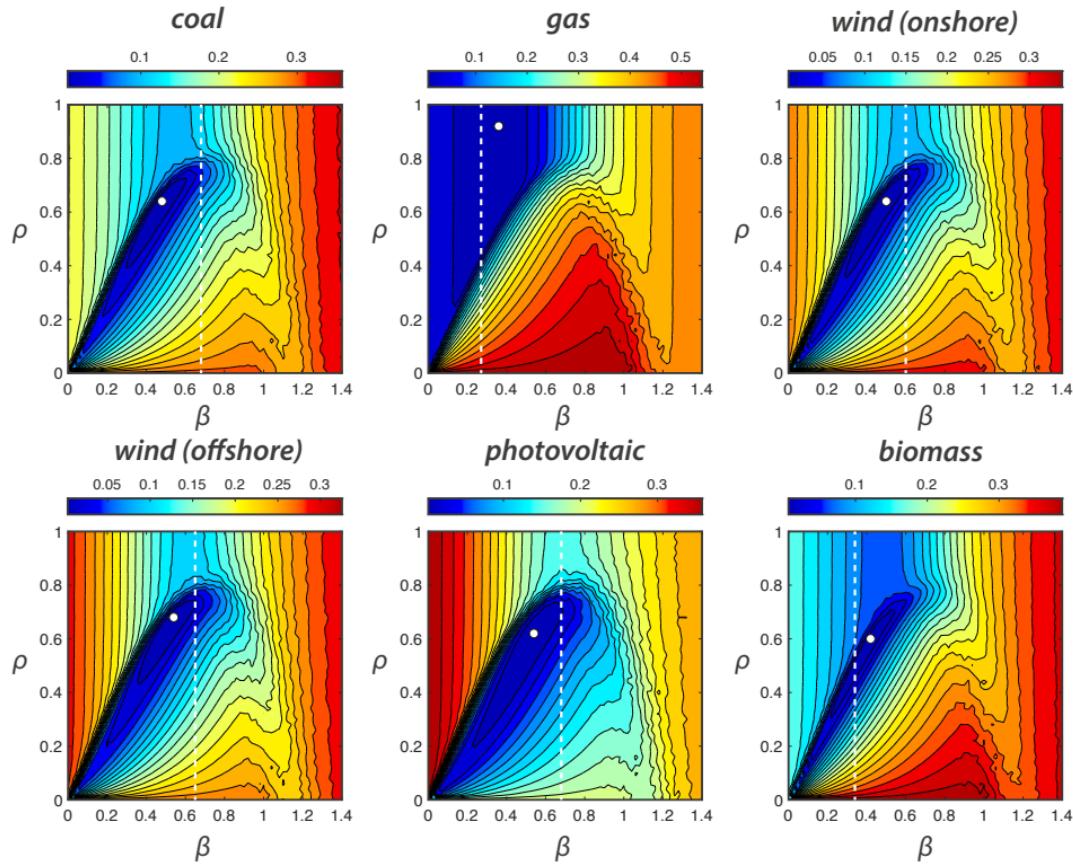
# *Empirical target distributions*

- In the survey experiment all 1078 subjects reported their opinions on the six different energy generating technologies.



# Comparing model outcomes to data

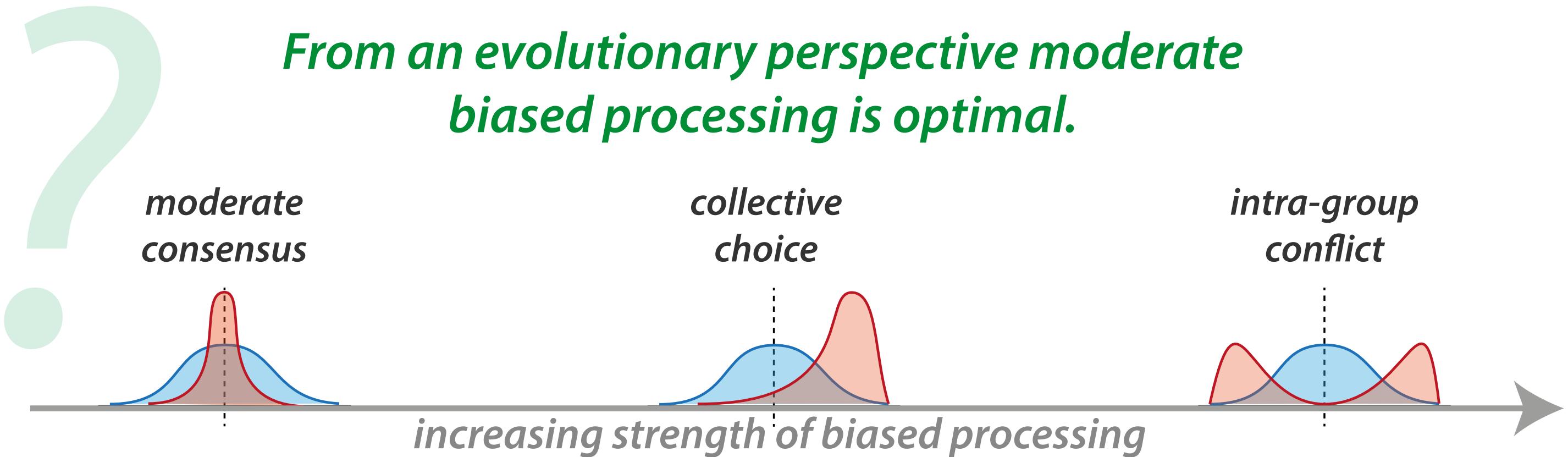
- ▶ Comparing the outcome distributions to empirical distributions using the Jensen-Shannon divergence



# Conclusions

1. *Weak biased processing speeds up group decision processes by orders of magnitude*
2. *Strong biased processing is sufficient for persistent intra-group polarisation*

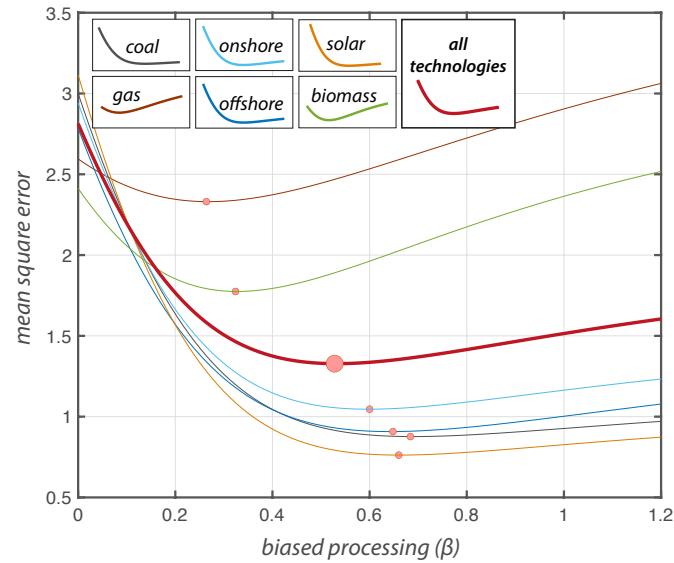
*From an evolutionary perspective moderate biased processing is optimal.*



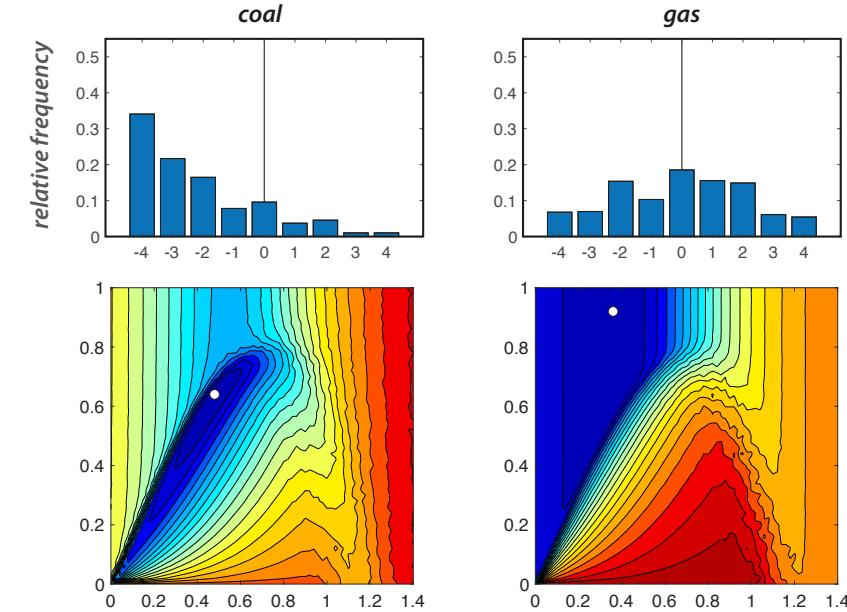
# Conclusions

1. Weak biased processing speeds up group decision processes by orders of magnitude
2. Strong biased processing is sufficient for persistent intra-group polarisation

3. Moderate biased processing is consistent with micro-level data on opinion change.



4. Moderate biased processing is also consistent with macro-level data on opinions.



*From an evolutionary perspective moderate biased processing is optimal.*

moderate consensus

collective choice

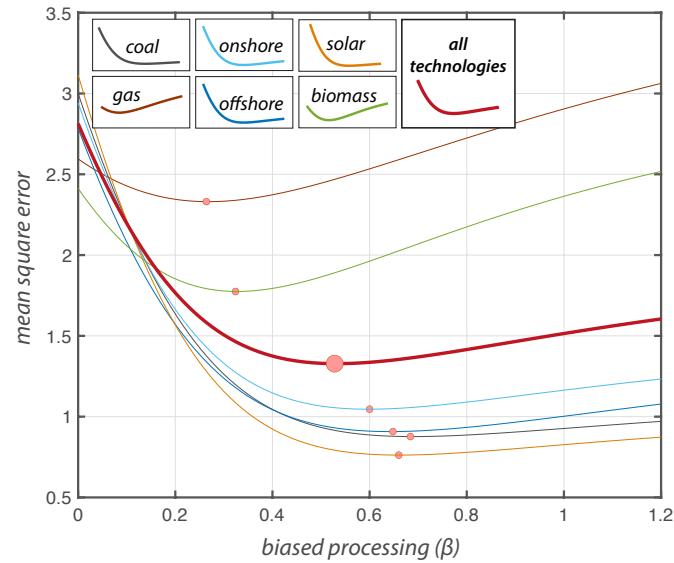
intra-group conflict

increasing strength of biased processing

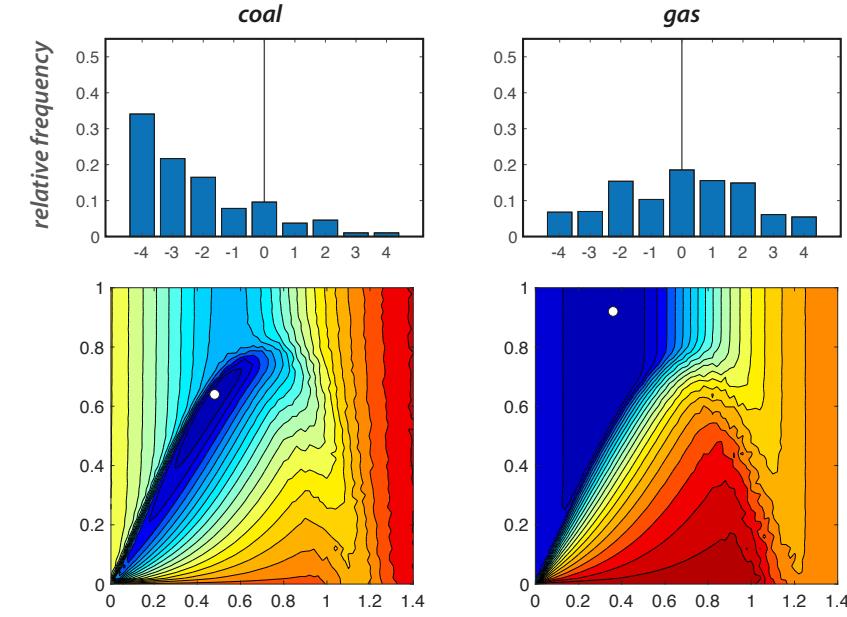
# Conclusions

1. Weak biased processing speeds up group decision processes by orders of magnitude
2. Strong biased processing is sufficient for persistent intra-group polarisation

3. Moderate biased processing is consistent with micro-level data on opinion change.



4. Moderate biased processing is also consistent with macro-level data on opinions.

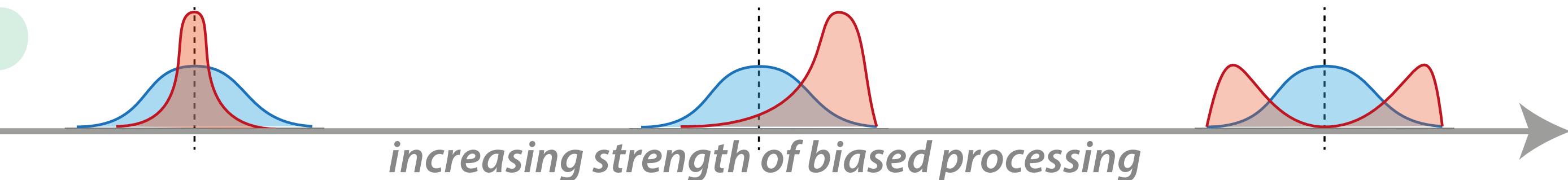


## Thanks!

moderate  
consensus

collective  
choice

intra-group  
conflict



# Topic A

## *at the summer school (14/06)*

*characterization of convergence times  
(analytical/numerical)  
for different  $\beta$*

