Agent-Based Modeling as a Philosophical Method

Dunja Šešelja

May 5, 2026

Institute for Philosophy II, Ruhr-University Bochum

Introduction

Social Epistemology of Science

Significance of Epistemic Diversity

- Is—and if so why–diversity of inquiries, approaches, hypotheses, values important?
- Advanced by pluralists (e.g., Feyerabend (1975), Longino (1990), Kellert et al. (2006), and Chang (2012))

Another question that intrigued philosophers of science:

- How is such diversity achieved and maintained?

How is such diversity achieved and maintained?

- Complex social dynamics of inquiry
- How does diversity as a *social phenomenon* emerge from *individual inquiries*?
- Requires studying causal factors => *counterfactuals*

Methodological shift

- Formal models
- Focusing on some factors while abstracting away from others

Formal models of scientific inquiry

- Analytical models
- Agent-based models

ABMs:

Computer simulations

From individual to group behavior

Highly idealized models

- Focusing on some causal factors (for example, how scientists conduct research, with whom they share information, etc.)
- Abstracting away from various other factors

What kind of scientific community do they represent?

(Plug in your favorite account of scientific representation)

What do ABMs represent?

- A “generic” scientific community?
- A small/large scientific community?
- Community typical of a specific scientific domain?
- An abstract notion of scientific community?

But ABMs are often linked to historical episodes

- Zollman (2010): peptic ulcer disease
- Holman and Bruner (2015): the use of DES for the prevention of miscarriage
- O'Connor and Weatherall (2018): chronic Lyme disease
- Weatherall et al. (2020): Tobacco Strategy

Problem:

No evaluation that confirms the link between the model and its target.

General take in the literature

These ABMs provide a *how-possibly explanation* of these historical episodes.

In this talk

- I will argue that highly idealized ABMs of science provide only epistemically opaque how-possibly explainations (HPEs)
- Such HPEs cannot be relevantly related to an empirical target
- Strategies for reducing the opaqueness:
  1. Robustness analysis
  2. Empirical validation
  3. Targeted monotonicity analysis (joint work with Daniel Frey and Christian Straßer)

Explanations from Highly-Idealized Models

How-Actually and How-Possibly Explanation (Verreault-Julien, 2019)

HAE

‘p because q and initial conditions c.’,

HPE

‘It is possible that ‘p because q and initial conditions c’.’

Syrjänen (2024):

Conditions c: boundary conditions

- parameters that delineate the predictive scope of a model
- assumptions specifying the ‘where’, ‘when’, and ‘who’ of a theory’s applicability.

HAE – Further Specification

HAE

‘p because q, restricted by the boundary conditions c.’,

- where we know what conditions these are, and we know that they hold for the given empirical target.

For example

ABMs in urban planning (e.g., why a specific transportation investment led to a specific change in a specific city).

Causal HPE

‘It is causally possible that ‘p because q, restricted by the boundary conditions c’’,

- where we know which conditions these are.
- We may not know whether they hold for a specific empirical target,
- Or we know that they do not hold for the given target (but we are interested in counterfactual scenarios).

For example

ABM providing HAE for influenza, used for studying epidemiology of COVID-19.

But sometimes we don’t know under which conditions the given causal relationship holds...

Epistemically opaque HPEs

‘It is causally possible that p because q, restricted by the boundary conditions c.’,

- where we are uncertain what conditions these are, and whether they hold for the given empirical target.
- The boundary conditions are *opaque*.

For example?

Virtually all highly idealized models that haven't been evaluated with respect to a specific empirical target.

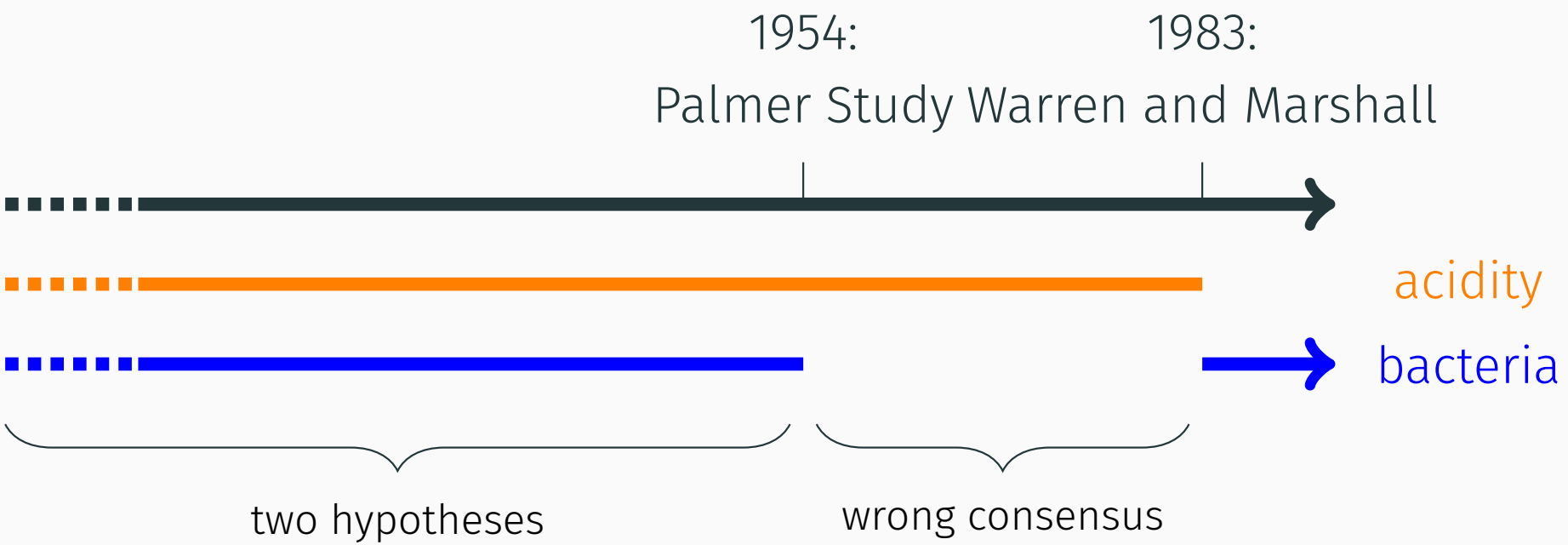
How to reduce opaqueness?

1. Robustness analysis: does only so much (especially if the results are positive)
2. Empirical validation: challenging for most ABMs of science.
3. Targeted monotonicity analysis

Targeted Monotonicity Analysis

Zollman's (2010) ABM and the case of peptic ulcer disease

The case of peptic ulcer disease (PUD)

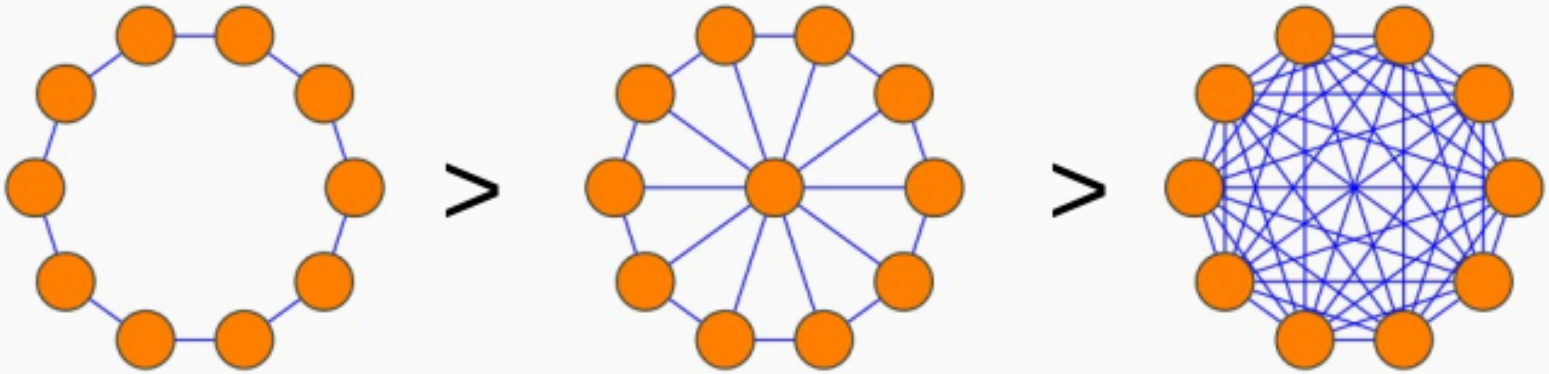


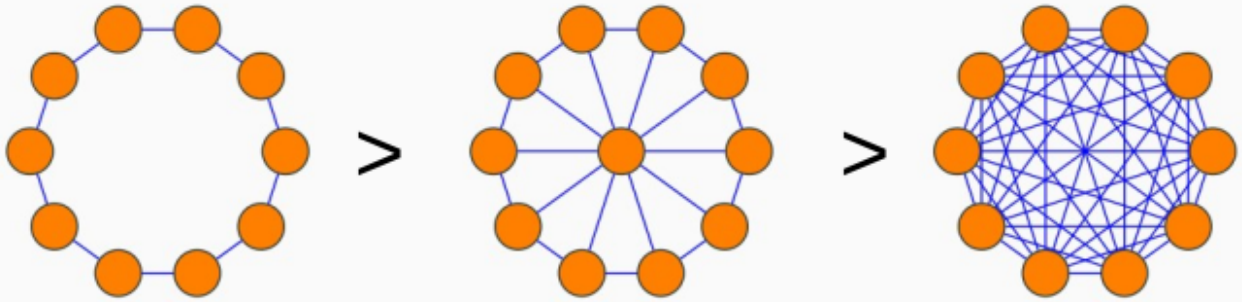
Zollman's (2010) ABM

Modeling science by 'bandit problems'

Zollman's (2010 ABM)

Unrestricted information flow appears to be harmful





	cycle	complete
Number of all runs	309932	310010
Runs with wrong consensus	53004	134629

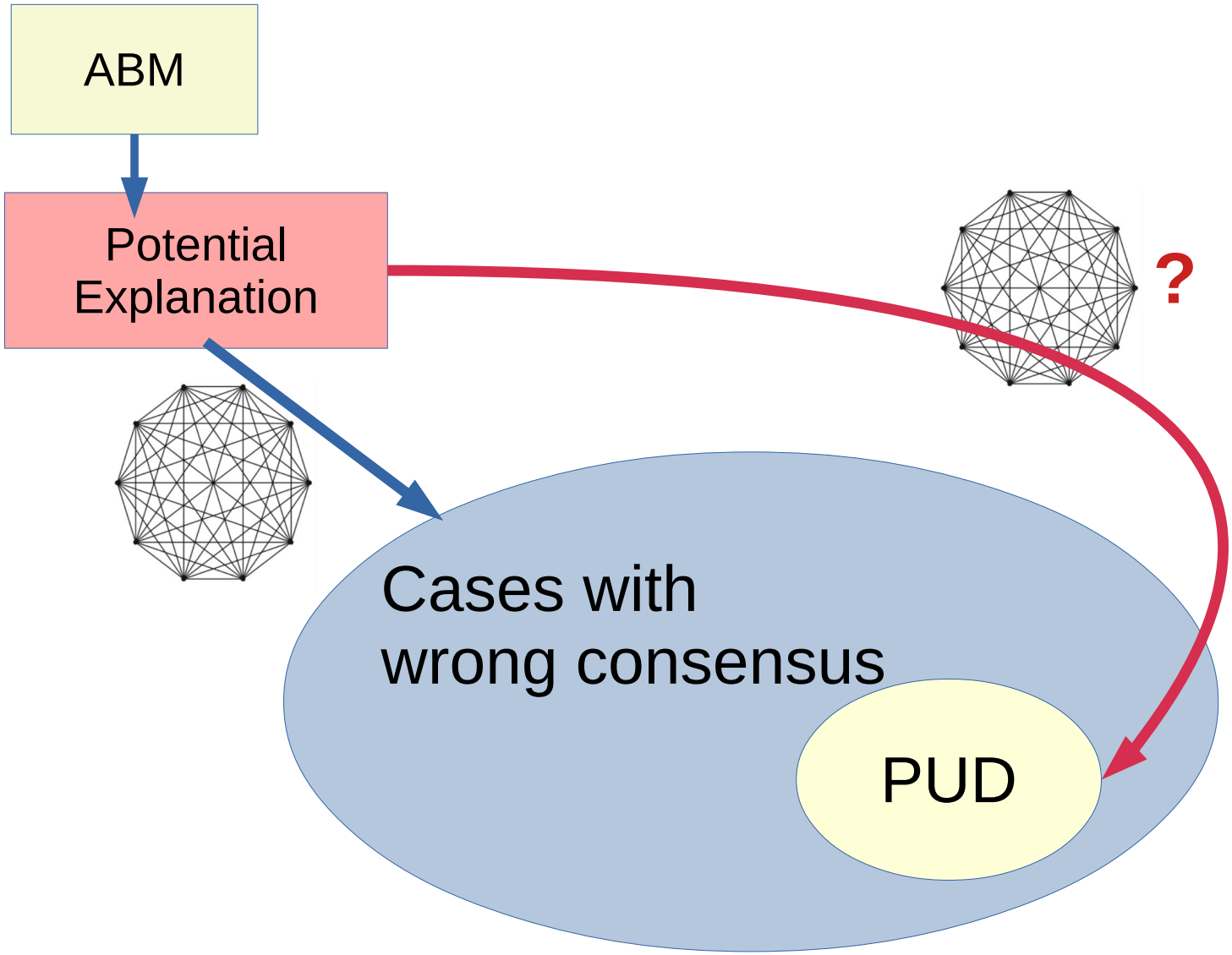
Application to the PUD case

Zollman, 2010

”For PUD, I have suggested that things might have been better had Palmer’s result not been communicated so widely or had people been sufficiently extreme in their beliefs that many remained unconvinced by his study.” (p. 33)

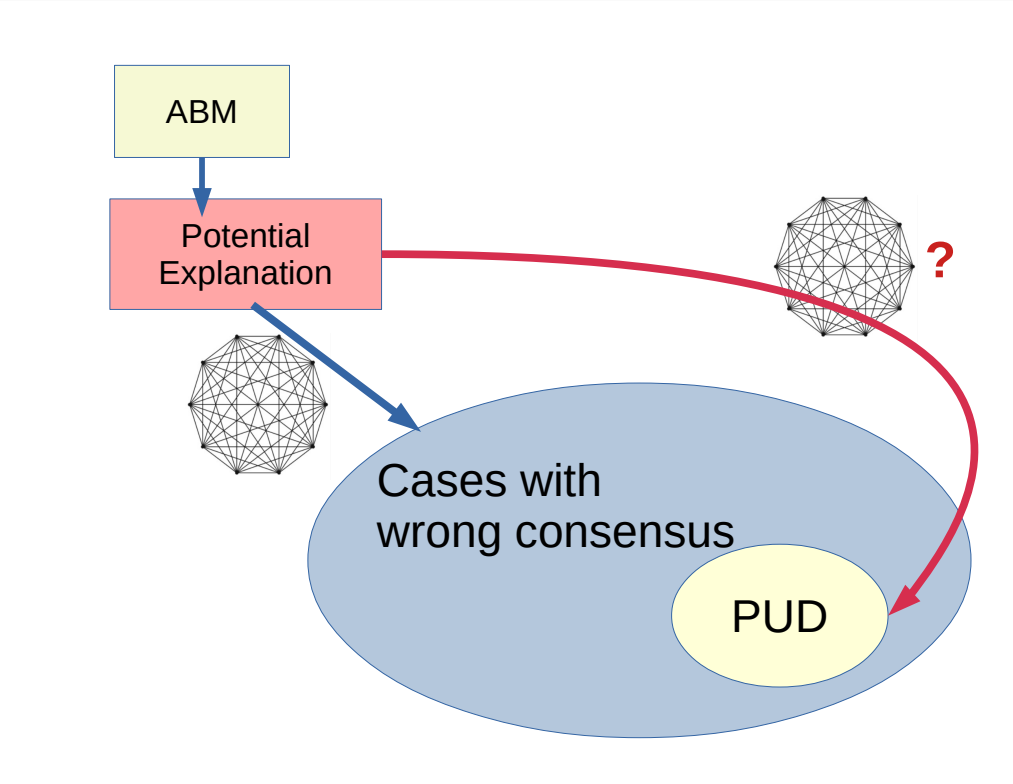
Potential explanation of the PUD case?

Targeted robustness analysis



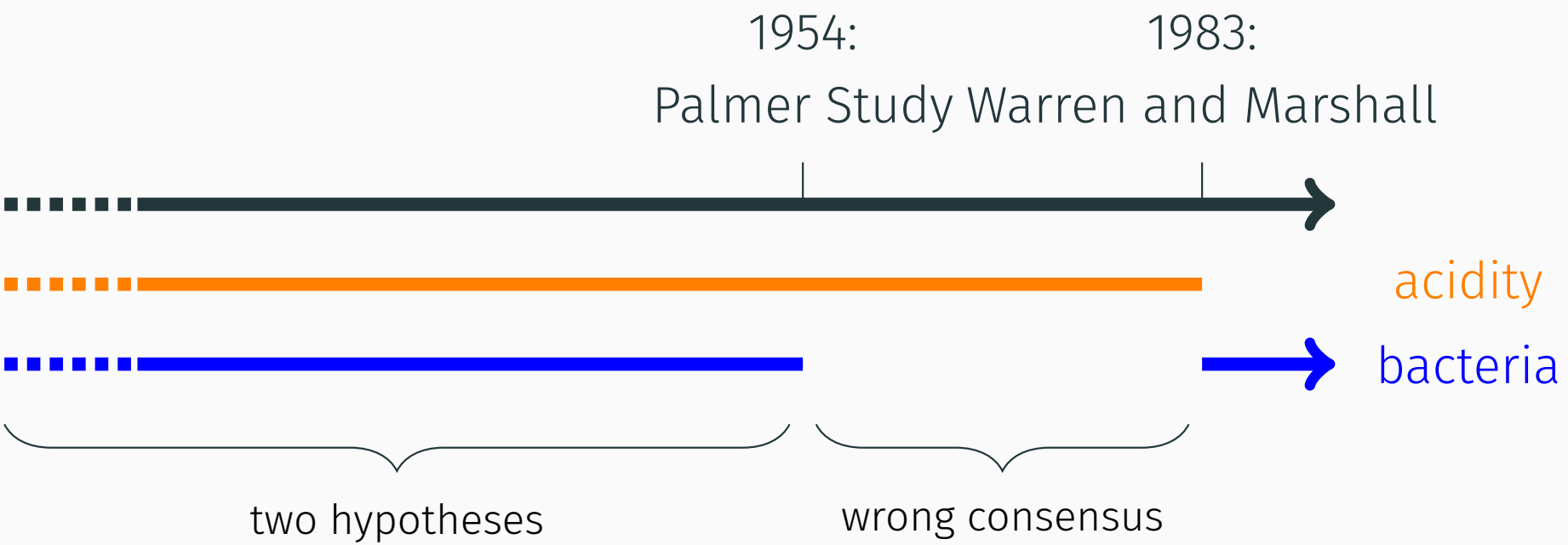
The plan:

- Specify characteristic features of the PUD episode.
- Extract the results of the model which conform to this dynamic.
- Measure the explanatory power of the proposed hypothesis in view of the obtained results.

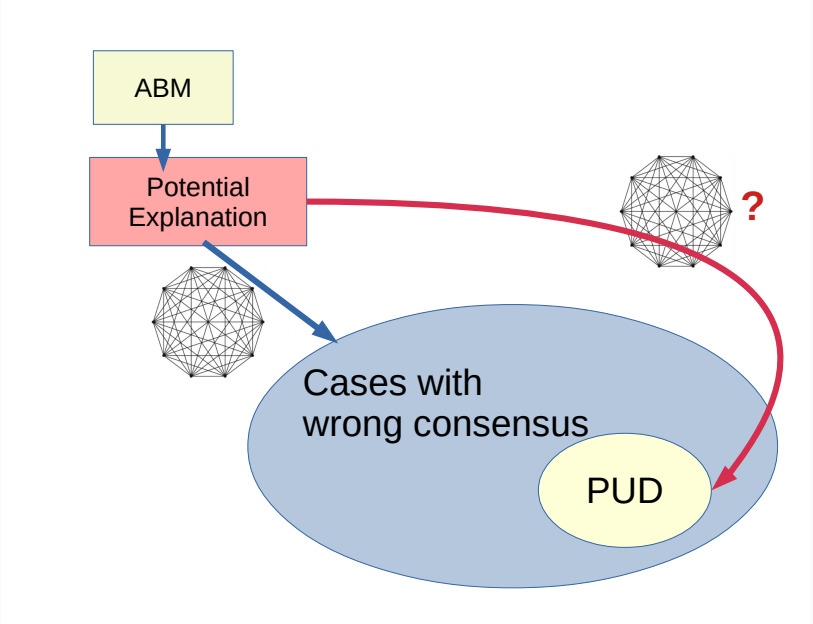


Specifying PUD-like cases

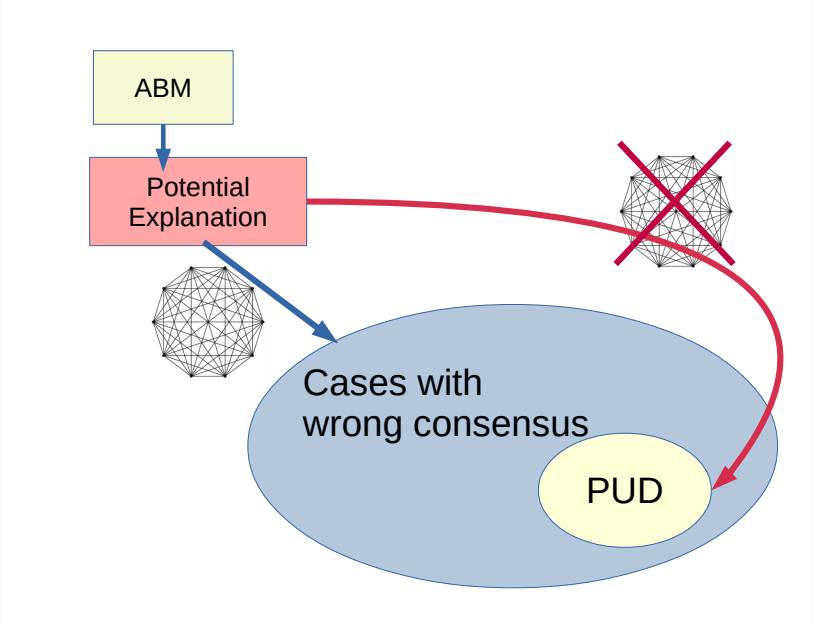
PUD case: some features



Our target: inquiries with wrong consensus, in which diversity was preserved for a large portion of the time span



	cycle	complete
Number of all runs	309932	310010
Runs with wrong consensus	53004	134629



	cycle	complete
Number of all runs	309932	310010
Runs with wrong consensus	53004	134629
Runs fitting PUD case	21705	1802

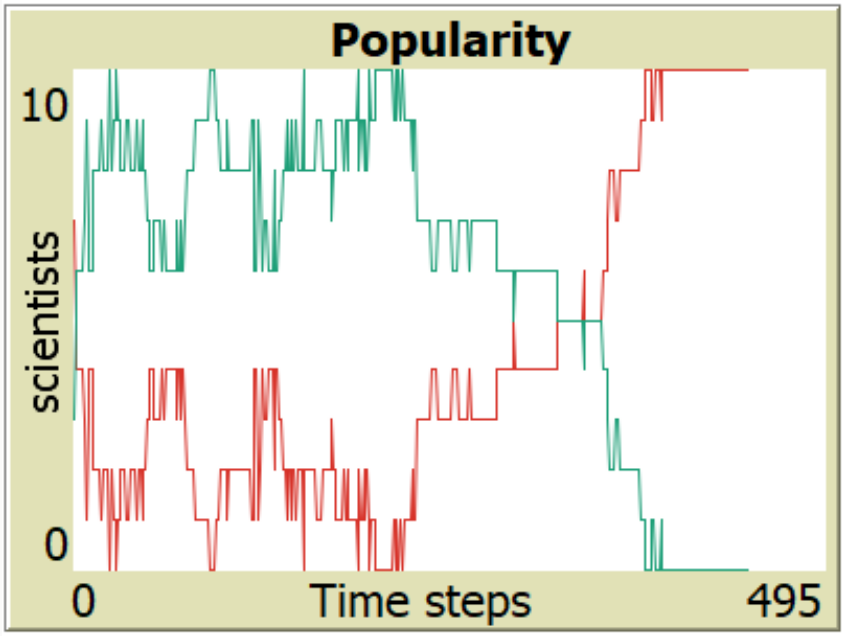


Figure 1: A cycle run

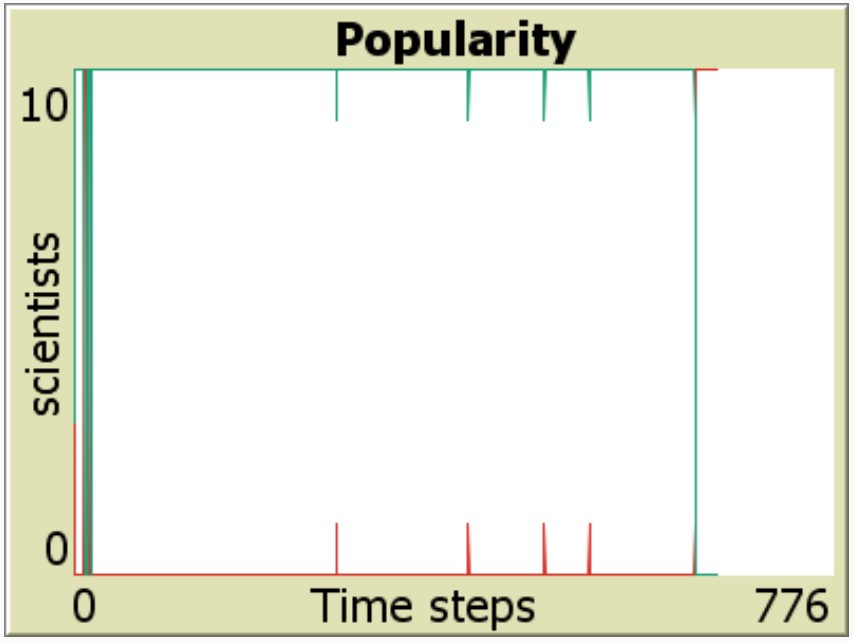


Figure 2: A complete graph run

Explanatory Power

Schupbach, J. N., & Sprenger, J. (2011). The logic of explanatory power. Philosophy of Science, 78(1), 105–127.

h explains e with power $\mathcal{E}(e, h)$:

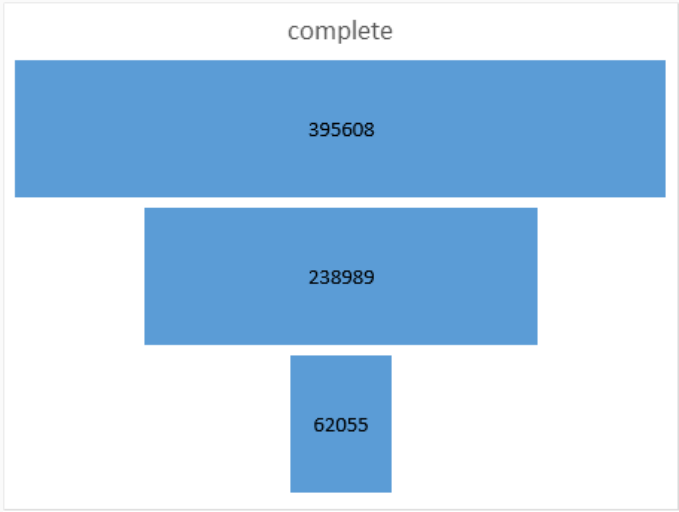
$$\mathcal{E}(e, h) = \frac{P(h \mid e) - P(h \mid \neg e)}{P(h \mid e) + P(h \mid \neg e)}$$

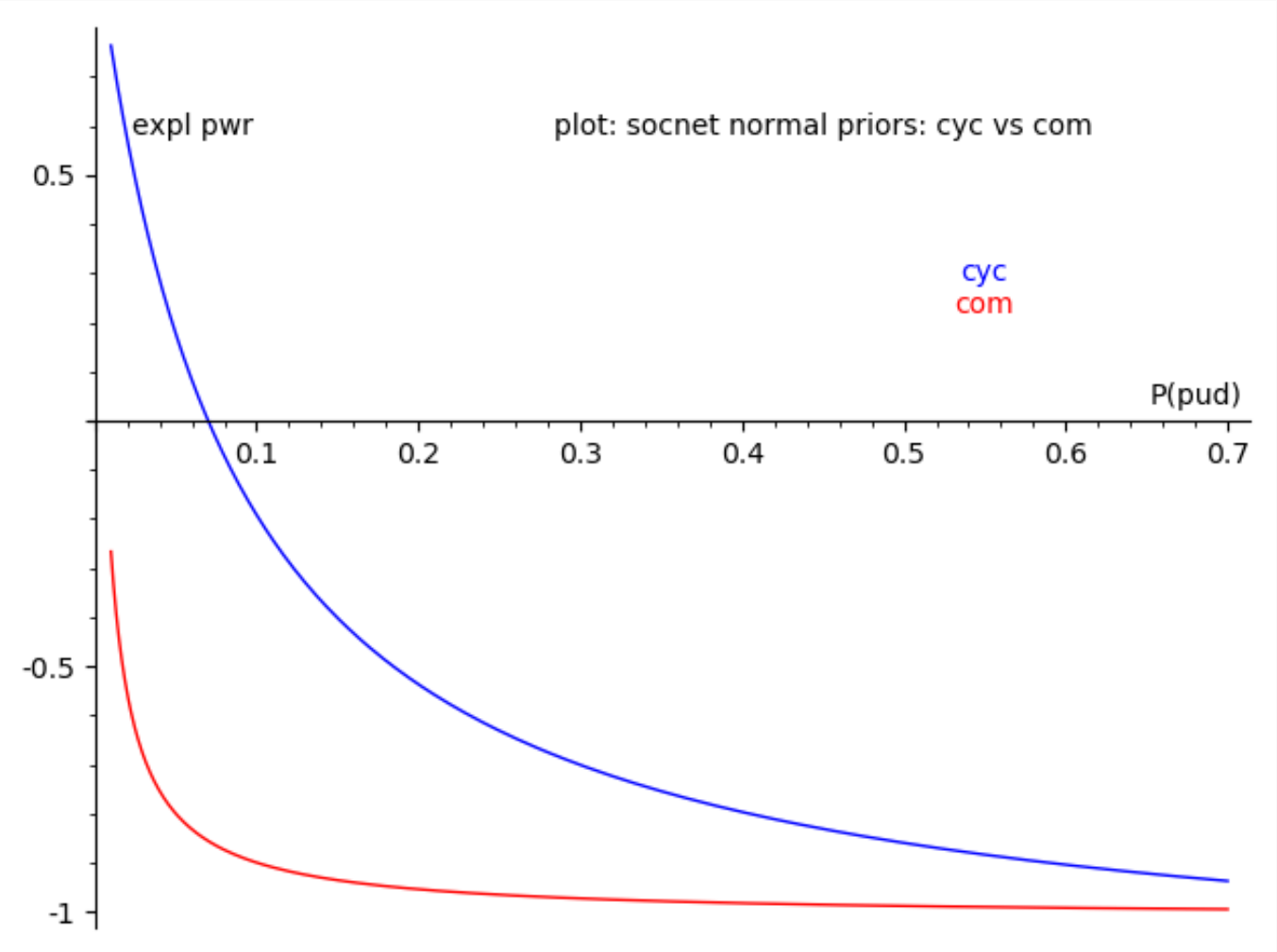
We plug in our data ...

$$\mathcal{E}(\text{pud} \mid \text{net}) = \frac{P(\text{net} \mid \text{pud}) - P(\text{net} \mid \neg\text{pud})}{P(\text{net} \mid \text{pud}) + P(\text{net} \mid \neg\text{pud})}$$

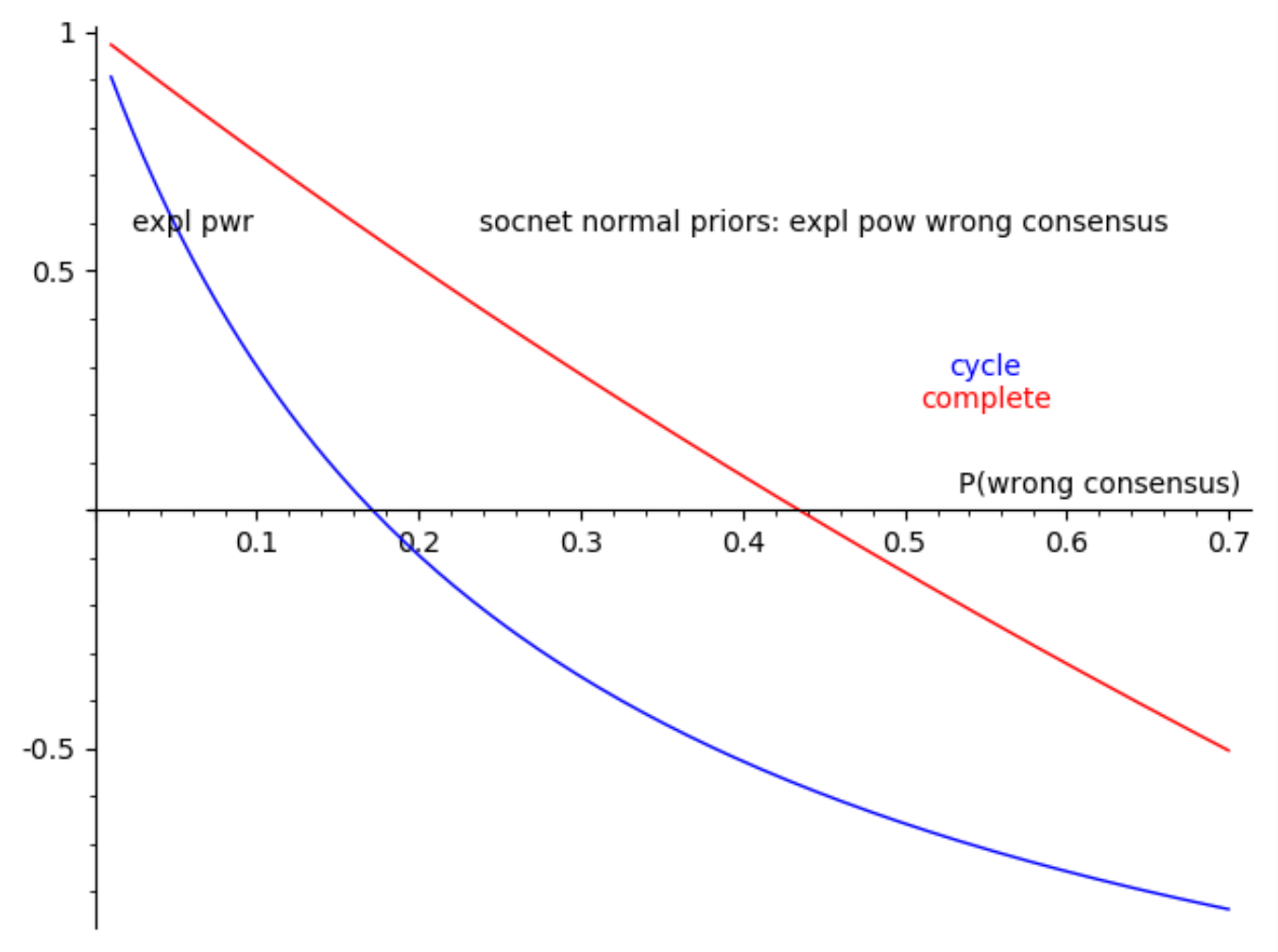
$$P(\text{net} \mid \text{pud}) = \frac{P(\text{net}) \cdot P(\text{pud} \mid \text{net})}{P(\text{pud})}$$

$$\mathcal{E}(\text{pud} \mid \text{net}) = \frac{\frac{P(\text{pud}|\text{net})}{P(\text{pud})} - \frac{P(\neg\text{pud}|\text{net})}{1-P(\text{pud})}}{\frac{P(\text{pud}|\text{net})}{P(\text{pud})} + \frac{P(\neg\text{pud}|\text{net})}{1-P(\text{pud})}}$$




expl pwr
plot: socnet normal priors: cyc vs com
0.5
cyc
com
P(pud)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
-0.5
-1

In contrast: explaining cases with wrong consensus



Zollman's model, while explanatory of cases with wrong consensus, is not explanatory of the PUD-like cases.

In fact...

The received historical narrative is wrong (Radomski et al., 2021).

Robustness analysis and empirical validation

Epistemically opaque HPE

The model shows that:

It is causally possible that a scientific community prematurely abandons a better of two rivaling hypotheses because of a high degree of information flow among the scientists, though we don’t know under which conditions that’s the case.

Robustness analysis?

Sensitivity analysis

Rosenstock et al. (2017): the result holds only for a small portion of the relevant parameter space (the context of *difficult inquiry*).

Robustness analysis?

Sensitivity analysis

Rosenstock et al. (2017): the result holds only for a small portion of the relevant parameter space (the context of *difficult inquiry*).

Derivational robustness

results confirmed by Lazer and Friedman (2007), Grim (2009), Grim et al. (2013), and Derex et al. (2018)

Empirical output validation

Mason et al. (2008) and Derex and Boyd (2016)

computer-based experiments

- Participants linked via different communication networks confronted with problem-solving situations.
- Less connected groups outperformed the more connected ones.

Common across the above studies: the assumption of a trade-off between exploitation and exploration

Relaxing the exploration/exploitation trade-off

Derivational robustness

Higher connectivity appears rewarding:

- Adding exploratory agents to 'bandit models': Kummerfeld and Zollman (2016),
- Argumentation-based ABM: Borg et al. (2017), Borg et al. (2018), and Borg et al. (2019)

Empirical output validation

Higher connectivity appears rewarding:

- an experimental study by Mason and Watts (2012)
- the exploitation of existing ideas does not necessarily lock-in the participants on the local maxima.

Alternative mechanisms of diversity

Focused on difficult inquiry

Frey and Šešelja (2020) If scientists have ‘rational inertia’: higher connectivity is rewarding

Removing epistemic opaqueness

In sum, the boundary conditions of Zollman's result:

- no extreme priors (bias towards the hypothesis)
- *myopic* explorers
- difficult inquiry
- there is a trade-off between exploration and exploitation (pursuing a hypothesis doesn't lead to insights into its rivals).

Conclusion

In this talk

I have argued that:

- For most ABMs in philosophy of science, we don't know their boundary conditions.
- Consequently, they give us only epistemically opaque how-possibly explanations.
- To reduce opaqueness, we can use several methods:
  - Robustness analysis
  - Empirical validation
  - Targeted monotonicity analysis

Thank you!

References

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer (2017). **"Examining network effects in an argumentative agent-based model of scientific inquiry".** In: *Logic, Rationality, and Interaction: 6th International Workshop, LORI 2017, Sapporo, Japan, September 11-14, 2017, Proceedings*. Ed. by Alexandru Baltag, Jeremy Seligman, and Tomoyuki Yamada. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 391–406.

Bibliography ii

- Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer (2018). **"Epistemic effects of scientific interaction: approaching the question with an argumentative agent-based model".** In: *Historical Social Research* 43.1, pp. 285–309.
- — (2019). **"Theory-choice, transient diversity and the efficiency of scientific inquiry".** In: *European Journal for Philosophy of Science* 9.2, pp. 1–25.
- Chang, Hasok (2012). ***Is Water H2O? Evidence, Pluralism and Realism.*** Dordrecht: Springer.

Bibliography iii

- Derex, Maxime and Robert Boyd (2016). **"Partial connectivity increases cultural accumulation within groups".** In: *Proceedings of the National Academy of Sciences* 113.11, pp. 2982–2987.
- Derex, Maxime, Charles Perreault, and Robert Boyd (2018). **"Divide and conquer: intermediate levels of population fragmentation maximize cultural accumulation".** In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1743, p. 20170062.
- Feyerabend, Paul K. (1975). ***Against Method*.** London: New Left Books.

Bibliography iv

- Frey, Daniel and Dunja Šešelja (2020). **"Robustness and idealization in agent-based models of scientific interaction".** In: *British Journal for the Philosophy of Science* 71, pp. 1411–1437. URL: `https://doi.org/10.1093/bjps/axy039`.
- Grim, Patrick (2009). **"Threshold Phenomena in Epistemic Networks.".** In: *AAAI Fall Symposium: Complex Adaptive Systems and the Threshold Effect*, pp. 53–60.
- Grim, Patrick, Daniel J Singer, Steven Fisher, Aaron Bramson, William J Berger, Christopher Reade, Carissa Flocken, and Adam Sales (2013). **"Scientific networks on data landscapes: question difficulty, epistemic success, and convergence".** In: *Episteme* 10.04, pp. 441–464.

Bibliography v

- Holman, Bennett and Justin Bruner (2015). **"The problem of intransigently biased agents".** In: *Philosophy of Science* 82.5, pp. 956–968.
- Kellert, Stephen H., Helen E. Longino, and C. Kenneth Waters, eds. (2006). ***Scientific pluralism.*** Minneapolis: University of Minnesota Press.
- Kummerfeld, Erich and Kevin JS Zollman (2016). **"Conservatism and the scientific state of nature".** In: *The British Journal for the Philosophy of Science* 67.4, pp. 1057–1076.
- Lazer, David and Allan Friedman (2007). **"The network structure of exploration and exploitation".** In: *Administrative science quarterly* 52.4, pp. 667–694.

Bibliography vi

- Longino, Helen E. (1990). ***Science as Social Knowledge: Values and Objectivity in Scientific Inquiry.*** Princeton, New Jersey: Princeton University Press.
- Mason, Winter and Duncan J Watts (2012). **"Collaborative learning in networks".** In: *Proceedings of the National Academy of Sciences* 109.3, pp. 764–769.
- Mason, Winter A, Andy Jones, and Robert L Goldstone (2008). **"Propagation of innovations in networked groups.".** In: *Journal of Experimental Psychology: General* 137.3, p. 422.
- O'Connor, Cailin and James Owen Weatherall (2018). **"Scientific polarization".** In: *European Journal for Philosophy of Science* 8.3, pp. 855–875.

Bibliography vii

- Radomski, Bartosz, Dunja Šešelja, and Kim Naumann (2021). **"Rethinking the History of Peptic Ulcer Disease and its Relevance for Network Epistemology".** In: *History and Philosophy of the Life Sciences* 43.113. DOI: https://doi.org/10.1007/s40656-021-00466-8.
- Rosenstock, Sarita, Cailin O'Connor, and Justin Bruner (2017). **"In Epistemic Networks, is Less Really More?"** In: *Philosophy of Science* 84.2, pp. 234–252.
- Syrjänen, Pekka (2024). **"Dimensions of predictive success".** In: *The British Journal for the Philosophy of Science*. DOI: https://doi.org/10.1086/733193.

Bibliography viii

- Verreault-Julien, Philippe (2019). **"How could models possibly provide how-possibly explanations?"** In: *Studies in History and Philosophy of Science Part A* 73, pp. 22–33.
- Weatherall, James Owen, Cailin O'Connor, and Justin P. Bruner (2020). **"How to Beat Science and Influence People: Policymakers and Propaganda in Epistemic Networks".** In: *The British Journal for the Philosophy of Science* 71.4, pp. 1157–1186. DOI: 10.1093/bjps/axy062. eprint: https://doi.org/10.1093/bjps/axy062. URL: https://doi.org/10.1093/bjps/axy062.
- Zollman, Kevin J. S. (2010). **"The epistemic benefit of transient diversity".** In: *Erkenntnis* 72.1, pp. 17–35.