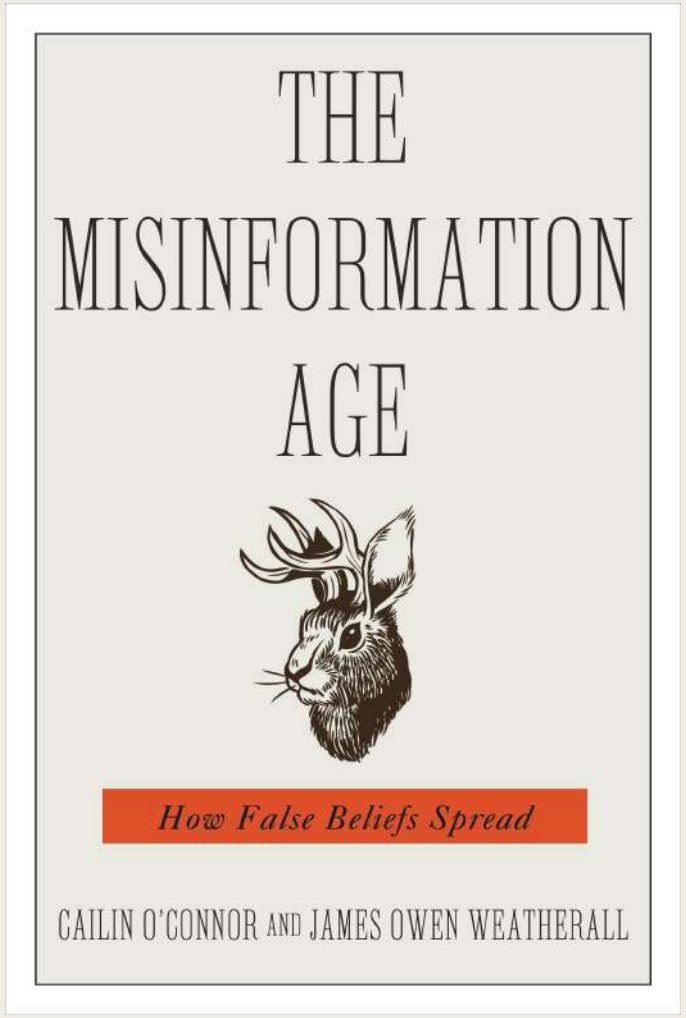
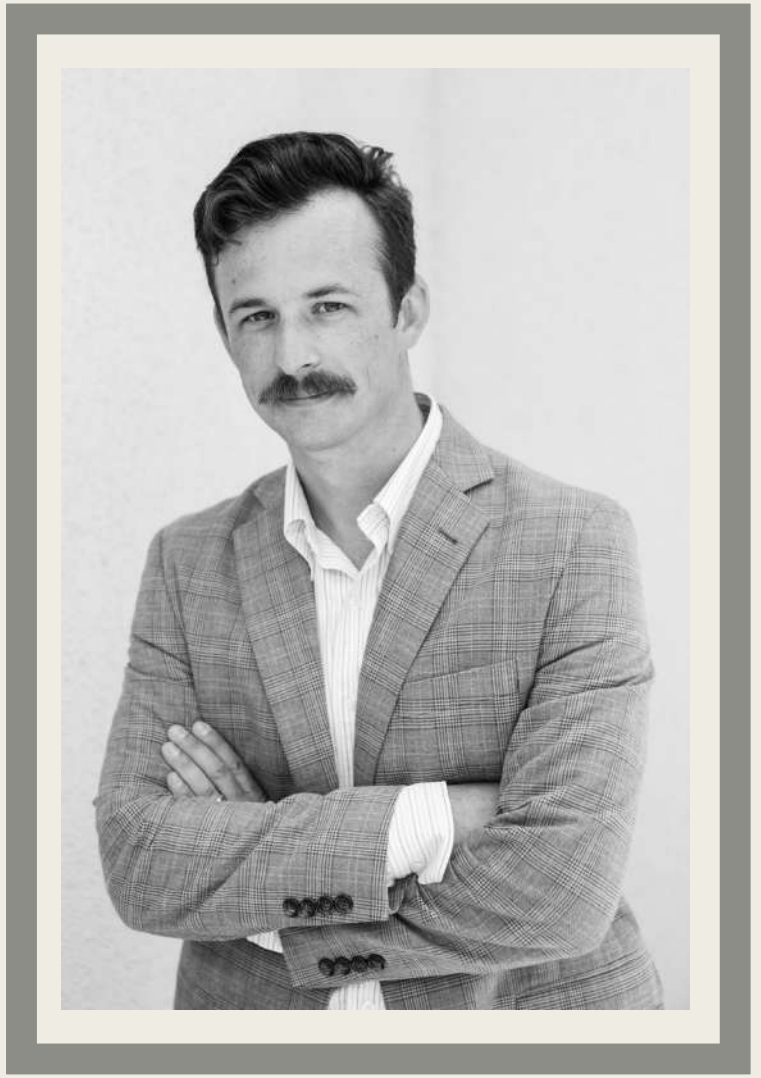
SCIENTIFIC POLARIZATION

Cailin O'Connor, Logic and Philosophy of Science, UC Irvine


THE
MISINFORMATION
AGE
How False Beliefs Spread
CAILIN O'CONNOR AND JAMES OWEN WEATHERALL



Big Picture

This talk will address the following question: why do communities sometimes polarize over matters of **scientific fact**, even in cases where they can gather evidence?

Methodology

In addressing this question, we use agent-based models employing the **network epistemology framework.**

Roadmap

1) Polarization and Network Epistemology
2) The Models
3) Polarization in Science
4) Factionalization in Science

Roadmap

1) Polarization and Network Epistemology
2) The Models
3) Polarization in Science
4) Factionalization in Science

Polarization

By polarization here we refer to situations where subgroups in a society hold **stable, mutually exclusive beliefs**, even in the face of debate and discussion.

Convergence

In the social sciences, polarization of this sort has presented a puzzle. Empirical evidence suggests that interacting people **typically converge** with respect to beliefs.

Then why, in some cases, do communities fail to come to consensus? Or even diverge in belief over time?

Modeling Approaches

Many teams have modeled polarization. These models usually include a feature of the following sort - **similarity of belief/opinion determines level of social influence.**

This is instantiated in different ways – through network connections, belief dynamics, etc.

Polarization and Values

In many cases, polarization occurs over beliefs/opinions that are grounded in **moral, religious, and political values.** Consider, for instance, debates over gun rights.

But in at least some cases, polarization occurs over matters of fact between actors who share values. **Previous models have not been geared at this sort of scenario.**

CHRONIC LYME AND POLARIZATION

NETWORK EPISTEMOLOGY

To build the models in this project, we use the network epistemology framework.

Multi-armed bandit problems

These models involve a **decision problem** and a network.

Agents decide between two actions A and B, which succeed at different rates.

For our models, A (all right) is successful with p(A) = .5

B (better) is successful with p(B) = .5 + e

But agents are unsure about whether B is in fact better or whether it is worse (p(B) = .5-e)

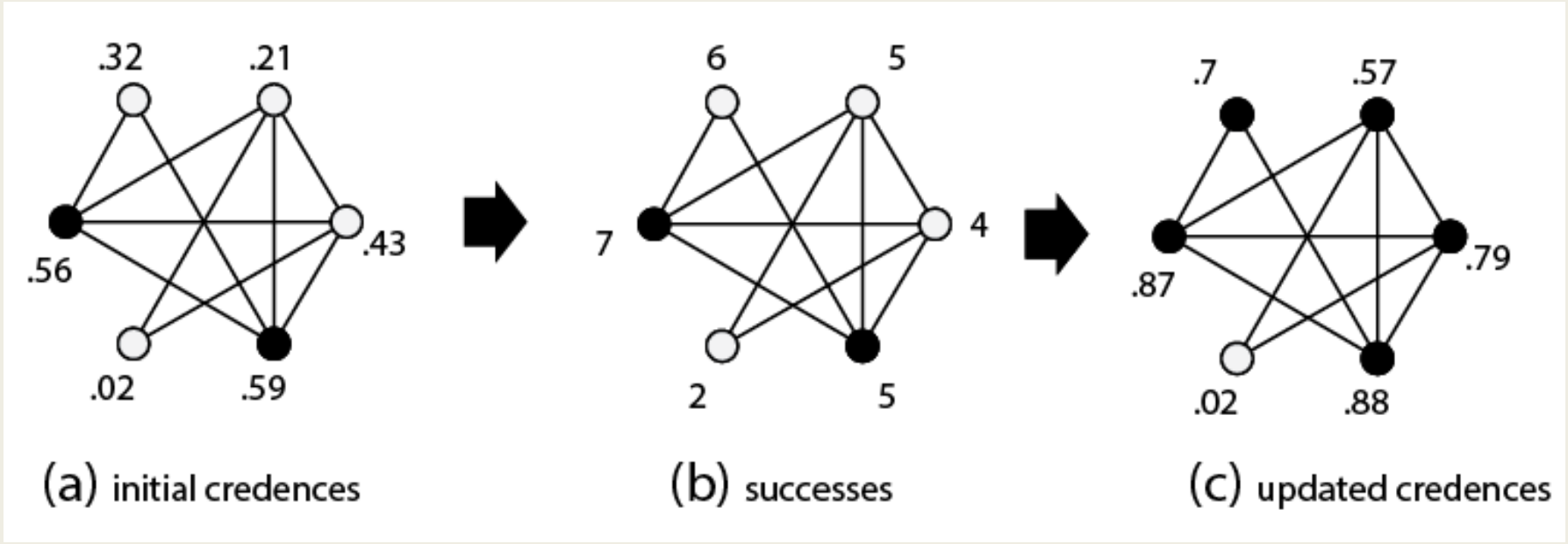
Dynamics

Agents start with a random credence [0,1] about whether B is better.

In successive rounds, agents take the action that they think is better, some number of times *n*.

They update their credences, using Bayes rule, based on the results.

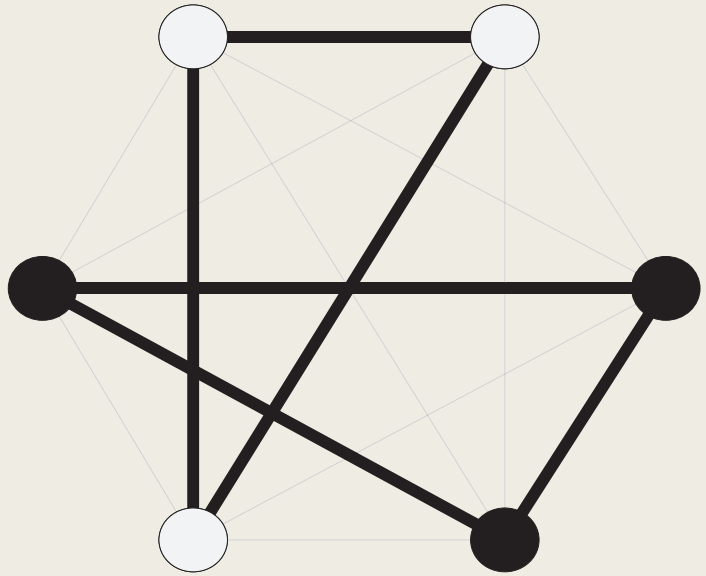
In addition, they update on results gathered by neighbors.


.32
.21
.56
.43
.02
.59
(a) initial credences
6
5
7
4
2
5
(b) successes
.7
.57
.87
.79
.02
.88
(c) updated credences

Network Structure

Variations on this model consider different network structures. We only consider **complete networks.**

As will become clear, our networks can be conceptualized as dynamic, where shared trust influences edge weights



Consensus

In previous work, authors have found that in this base model communities tend towards **consensus.**

Usually all agents converge to high confidence in the better action B.

Sometimes all agents erroneously settle on the worse action, A.

In other words, **we do not see stable polarization.**

Adequacy

Why are these good models for cases of scientific polarization?

1) Agents gather data
2) Data is equivocal, like scientific data
3) Actors are epistemically motivated
4) Actors have reasonable ways of connecting evidence with belief

Roadmap

1) Polarization and Network Epistemology
2) The Models
3) Polarization in Science
4) Factionalization in Science

Shared Belief and Social Trust

The key alteration to the model assumes agents are more likely to trust evidence that comes from someone who **holds similar beliefs.**

In science this is at least somewhat **reasonable** - it is irresponsible to trust data gathered by known quacks. A scientist who trusts her own reasoning might well deduce that those who hold very different opinions are not trustworthy.

Jeffrey Conditionalization

Jeffrey's rule gives a way for actors to do Bayesian updating on evidence that is **uncertain.**

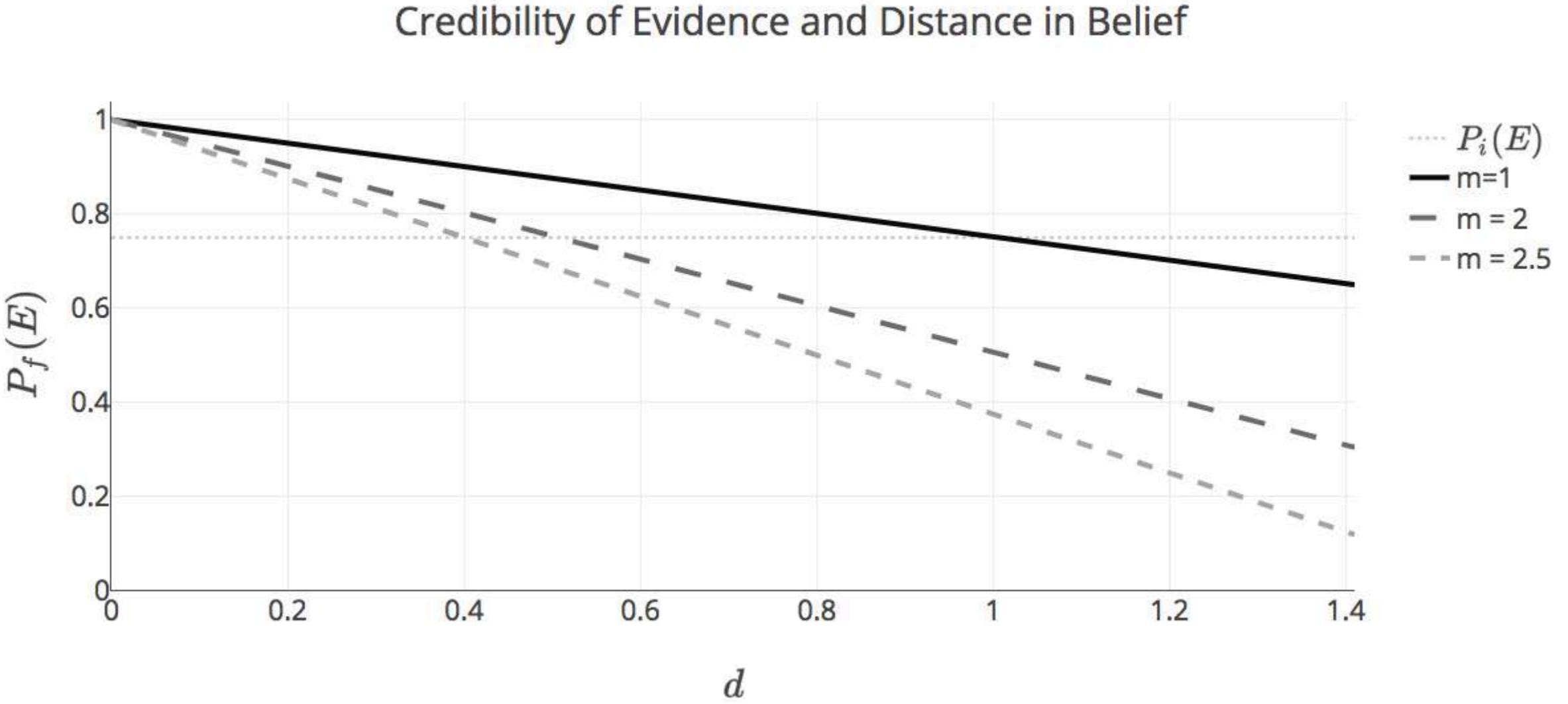
Under this rule, an agent has a credence about how likely it is some set of evidence in face obtained, $\mathbf{P_f(E)}$.

They then update their belief weighing two possibilities – that the evidence is real, and that it isn't. Their new credence takes these possibilities into account.

Decreasing Certainty

We model $P_f(E)$ as a **decreasing function** of the distance between agents' beliefs.

Although we test various functions, we mostly use a linear function, which is scaled by a **multiplier** (m), the **distance between beliefs** (d), and an **agent's prior** belief that the evidence in question occurred.


Credibility of Evidence and Distance in Belief
$P_f(E)$
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
d
$P_i(E)$
m=1
m = 2
m = 2.5

Roadmap

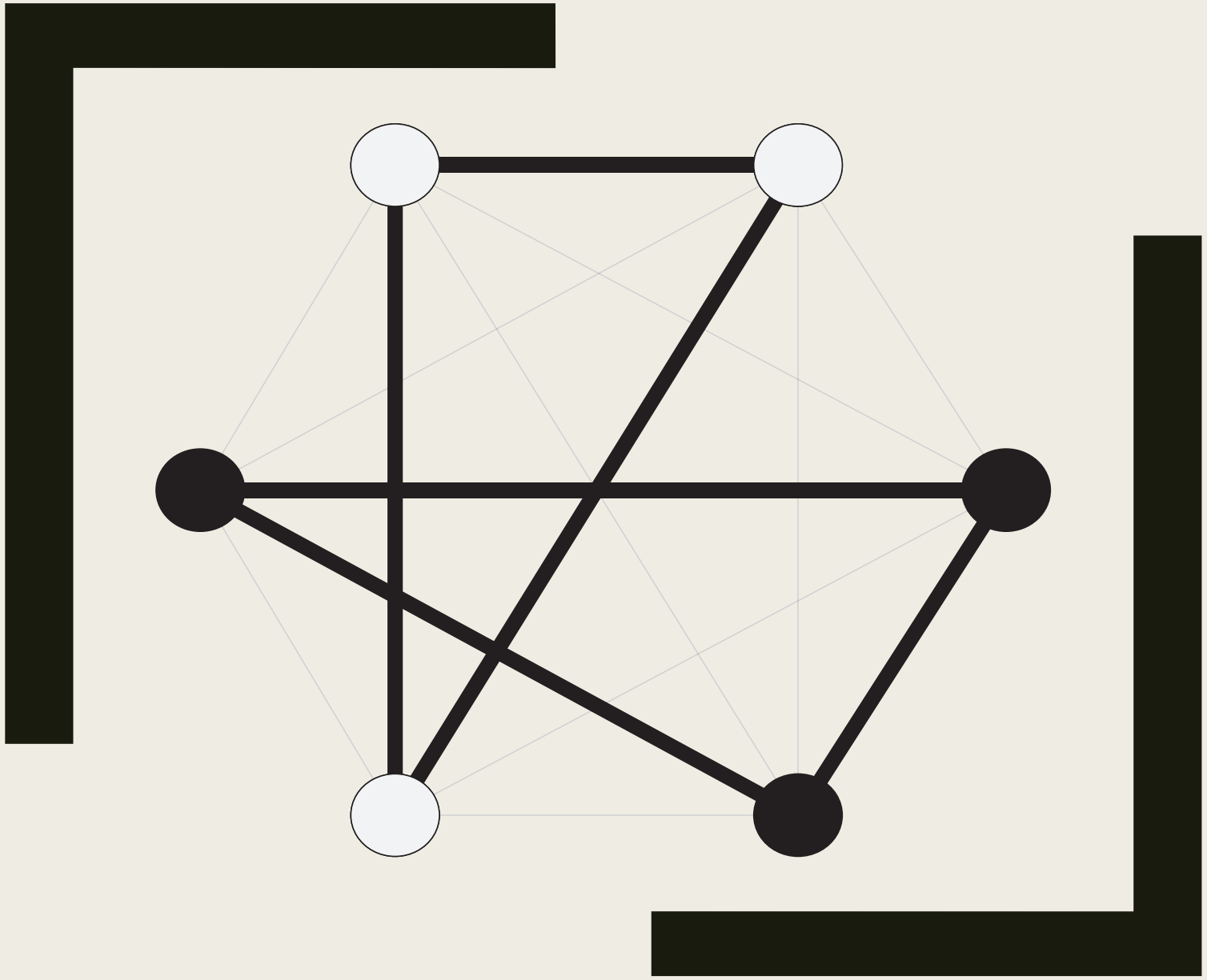
1) Polarization and Network Epistemology
2) The Models
3) Polarization in Science
4) Factionalization in Science

Results

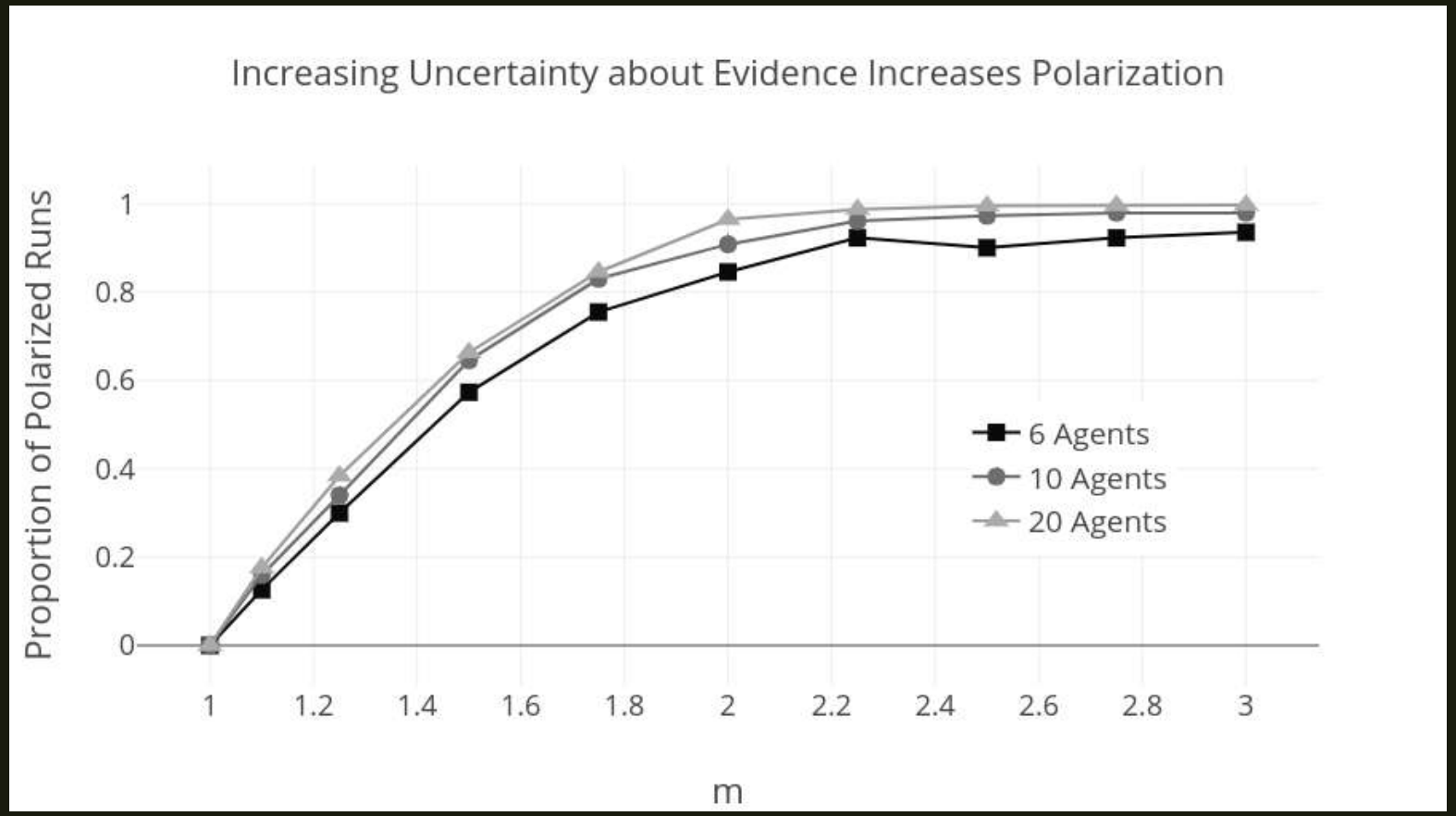
With this alteration there is a **new possibility** for the model.

When m <1, all simulations **go to consensus,** either True (p(B) > .99) or False (p(B)<.5). This is because agents always influence each other at least a bit.

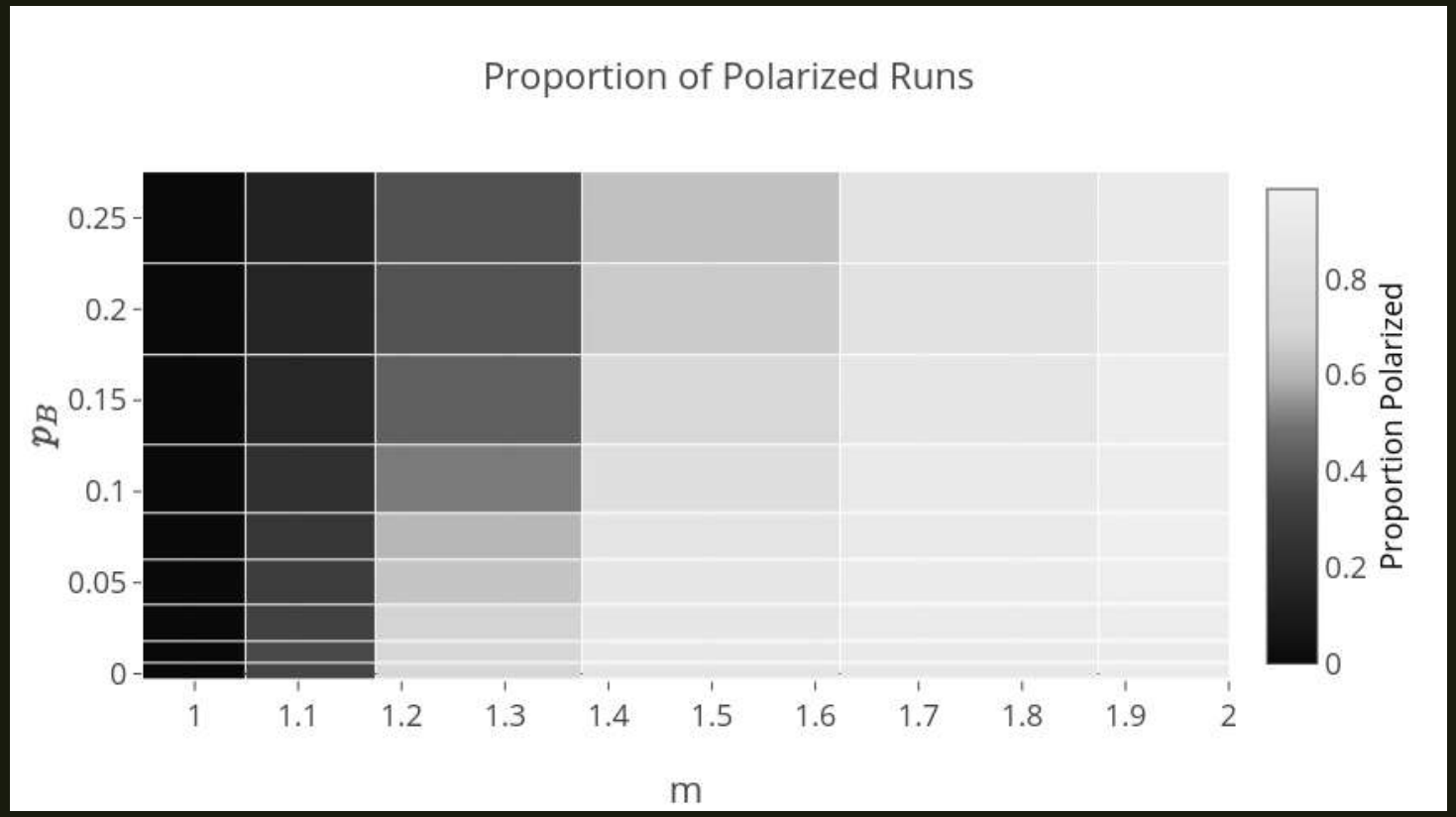
When m>1, we see **polarization** – stable outcomes where some agents have high credences, and some low.



POLARIZATION

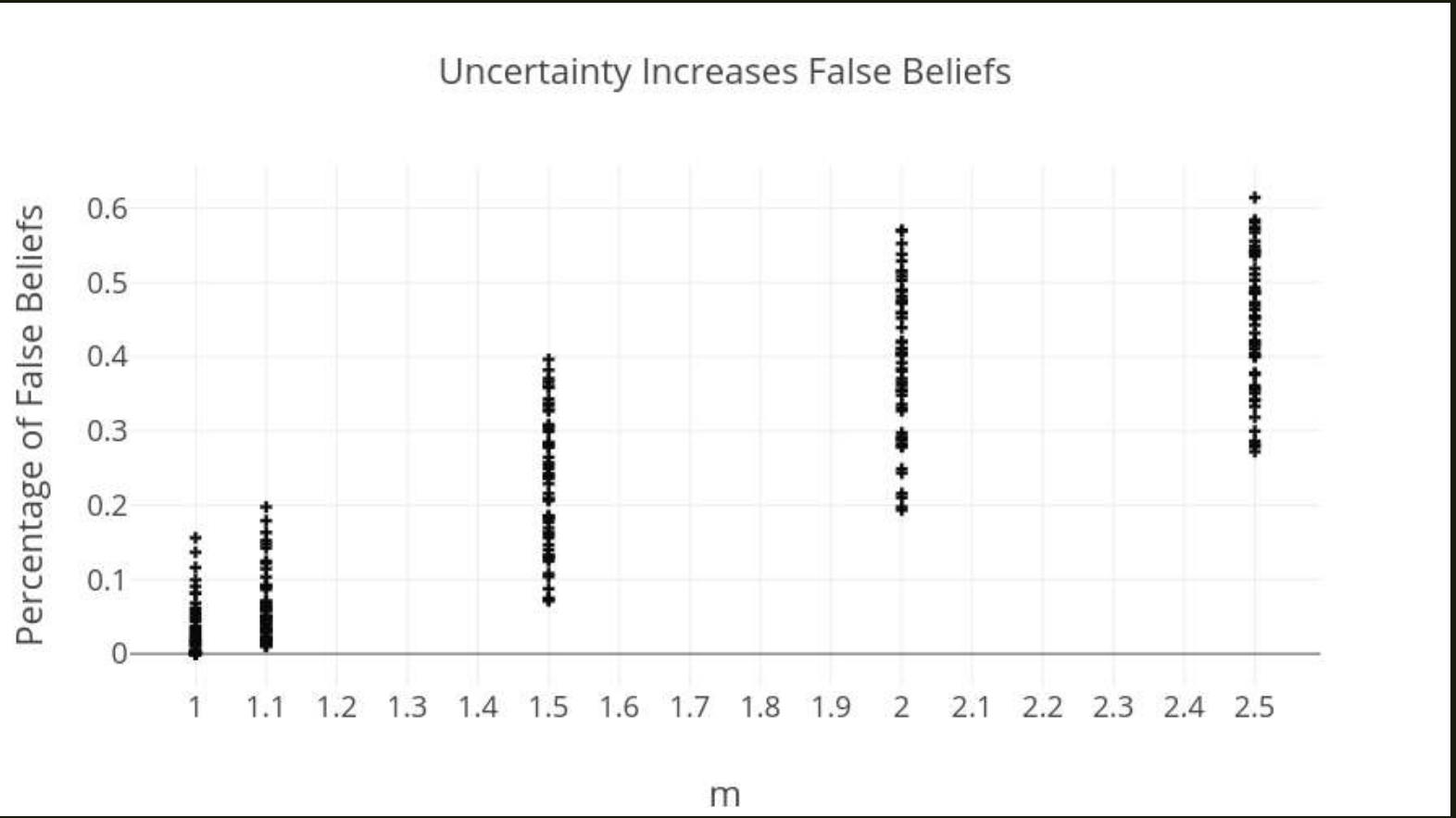


Increasing m increases polarization



When B is better, this is mitigated

Uncertainty Increases False Beliefs



Take-Away

As mentioned, there is something reasonable about a scientist who is skeptical of those who have reached different conclusions.

Nonetheless, we show that this sort of skepticism, on a global level, **hurts the knowledge producing capacity of the community**. And leads to stable polarization.

In a case like Lyme, we see why actors with good intentions and similar values could end up in a situation where researchers share very little trust, and are not headed towards consensus.

Roadmap

1) Polarization and Network Epistemology
2) The Models
3) Polarization in Science
4) Factionalization in Science

Factionalization

In some cases, we see polarization where actors form **factions with multiple, shared, polarized beliefs.** These often fall along political party lines.

For example, in the US beliefs about whether climate change is occurring are correlated with beliefs about whether evolutionary theory is correct.

Ideology and Explanation

Many previous authors have explained this via appeal to **shared ideology**.

George Lakoff, for example, claims that in the US conservatives hold to a '**strict father**' model, and liberals to a '**nurturant parent**' model:

"the role of government, social programs, taxation, education, the environment, energy, gun control, abortion, the death penalty, and so on… are ultimately not different issues, but manifestations of a single issue: strictness versus nurturance" (Lakoff, 2010, x)

Scientific Factions

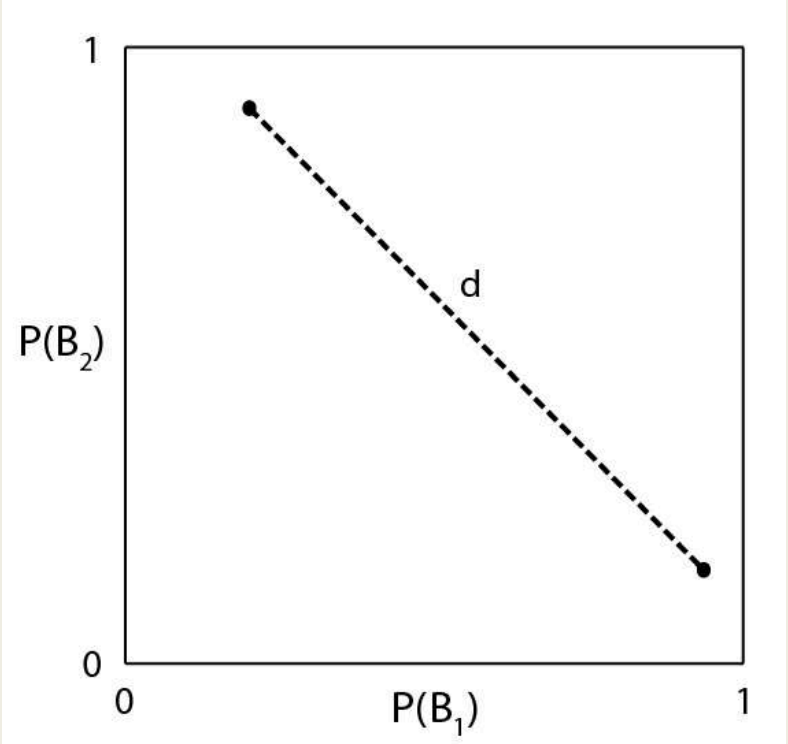
But what about cases where these bundles of beliefs seem to share no ideological grounding?

In this paper, we ask: can such bundles can **emerge endogenously** as a result of social trust grounded in shared beliefs?

Our Model

We consider a variant on our polarization model, where actors hold beliefs in **two arenas.**

We use the same formula to determine certainty in evidence, but now the distance is a Euclidean distance between their two beliefs.



Outcomes

Now outcomes are **slightly more complicated.** We see all combinations of True, False, and Polarized beliefs in both arenas.

When agents polarize on both beliefs, there can be **varying levels of correlation** between these.

Measuring Correlation

Do polarized beliefs end up correlating because of agent trust?

To measure correlation, we calculate the absolute value of the **Pearson correlation** between true and false beliefs in the two arenas for each agent.

We can then compare this with the level of correlation expected to emerge without mistrust in both beliefs.

	Belief X	Belief Y
Agent 1	0	1
Agent 2	1	0
Agent 3	1	0

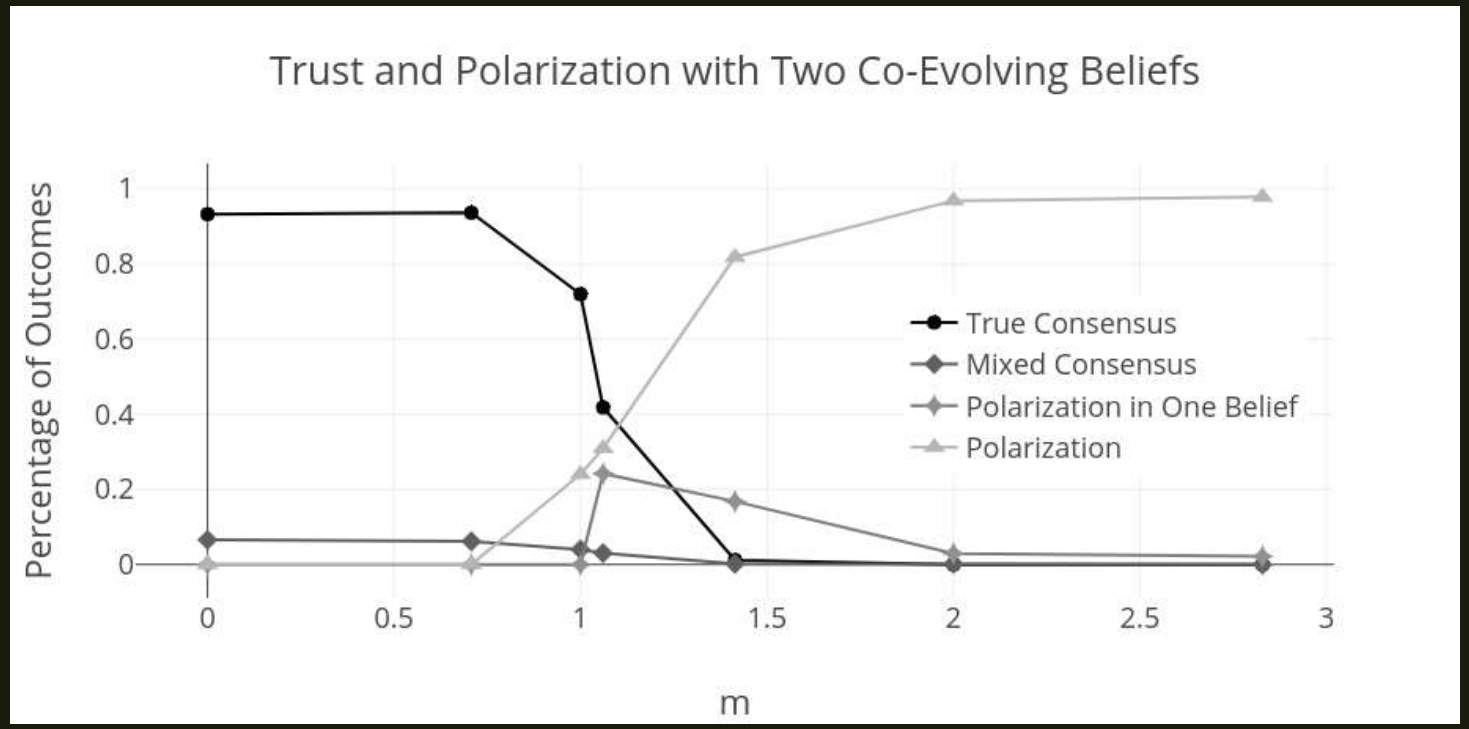
Three treatments

We look at three treatments.

1) Agents are pre-polarized in belief X, and develop credences about belief Y
2) Agents co-evolve credences about both beliefs
3) Agents co-evolve credences about three beliefs

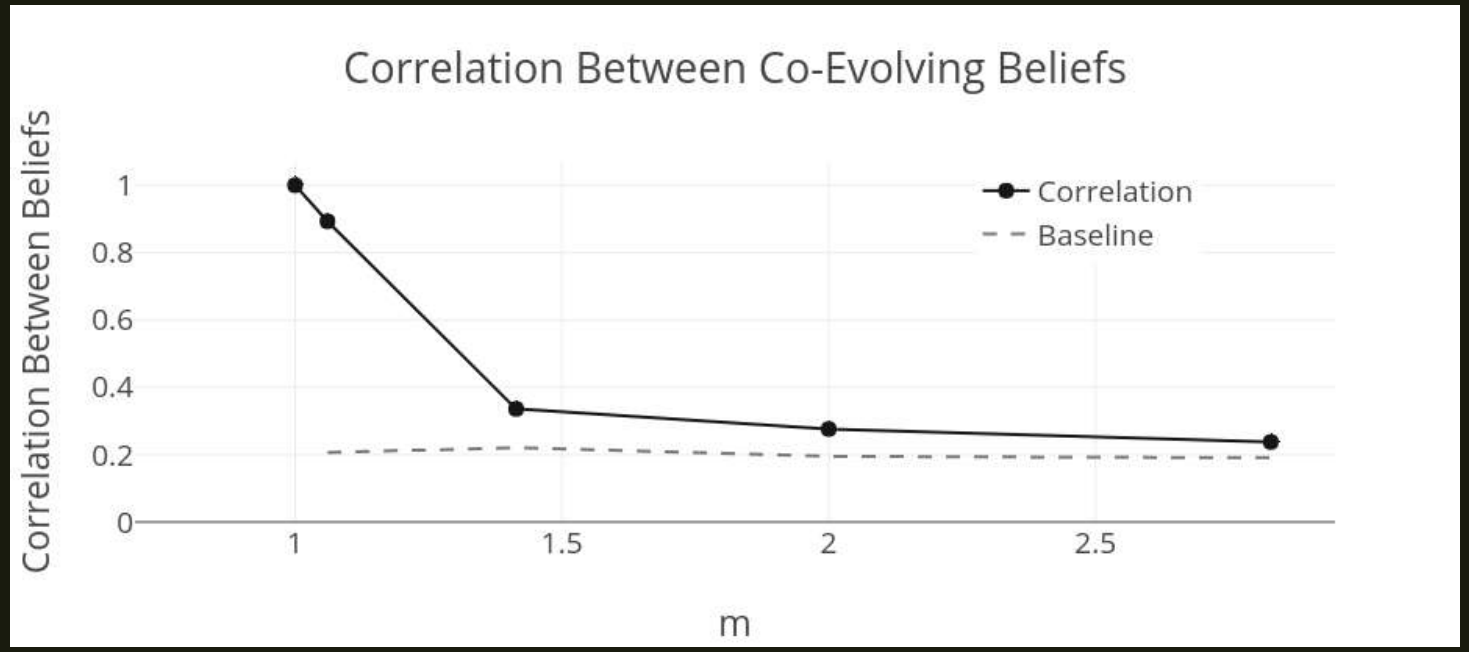
Coevolving Beliefs

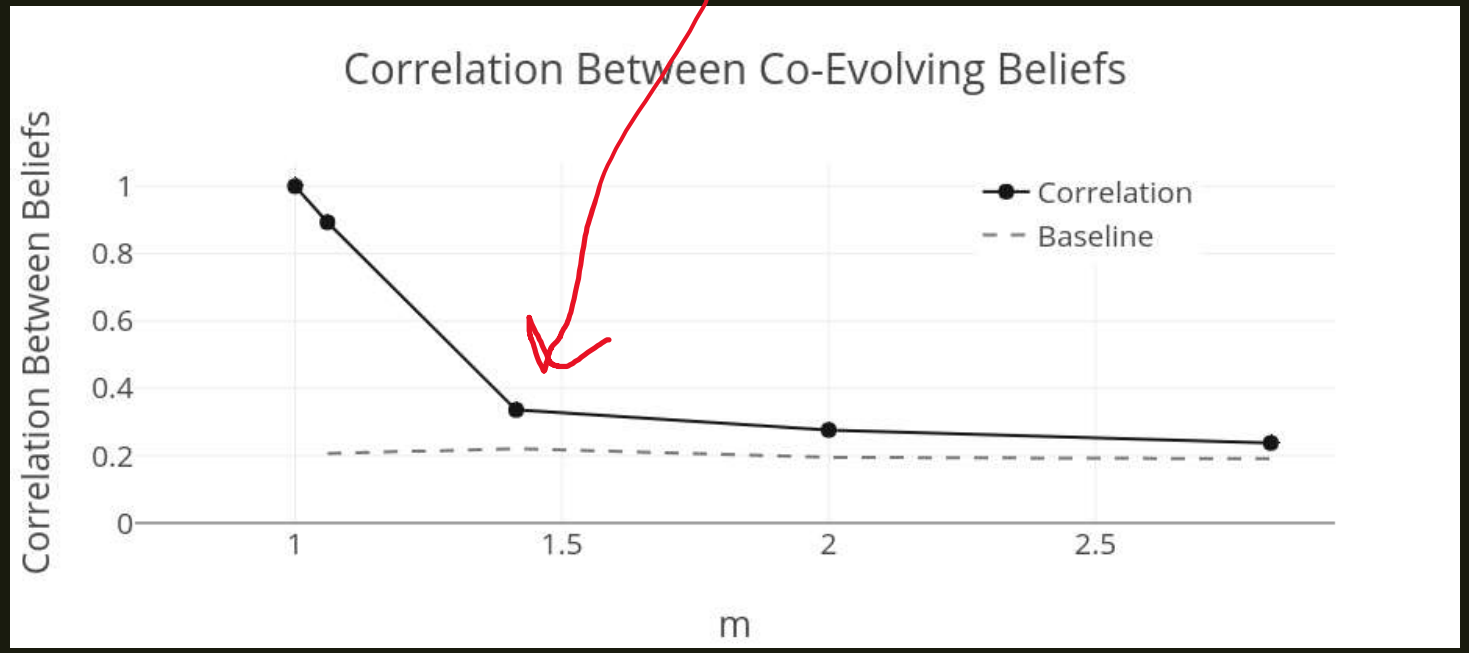
Since results are qualitatively similar across these three treatments, let’s look at the models where beliefs **co-evolve** from the start.



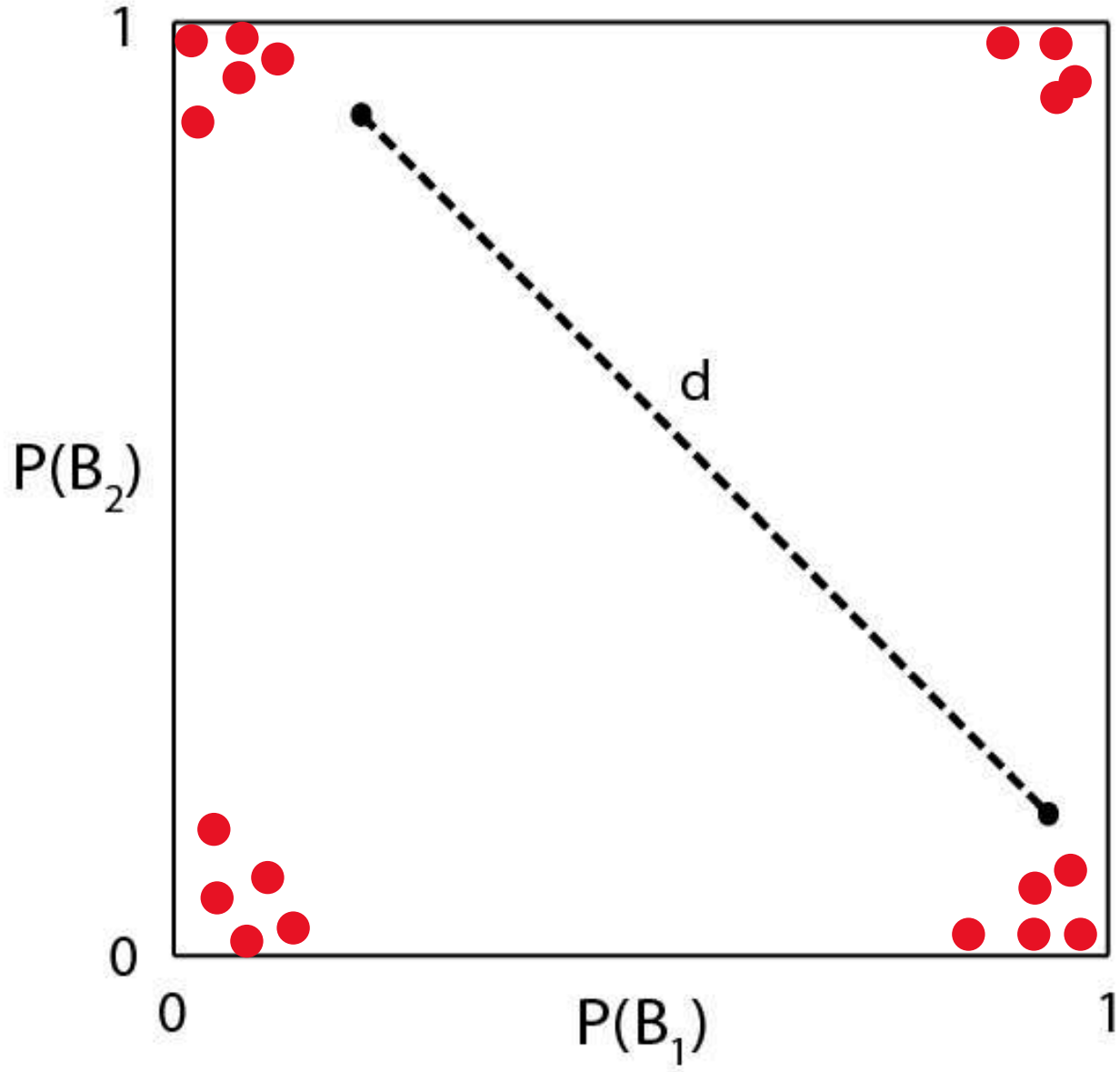
Mistrust increases multiple-arena polarization

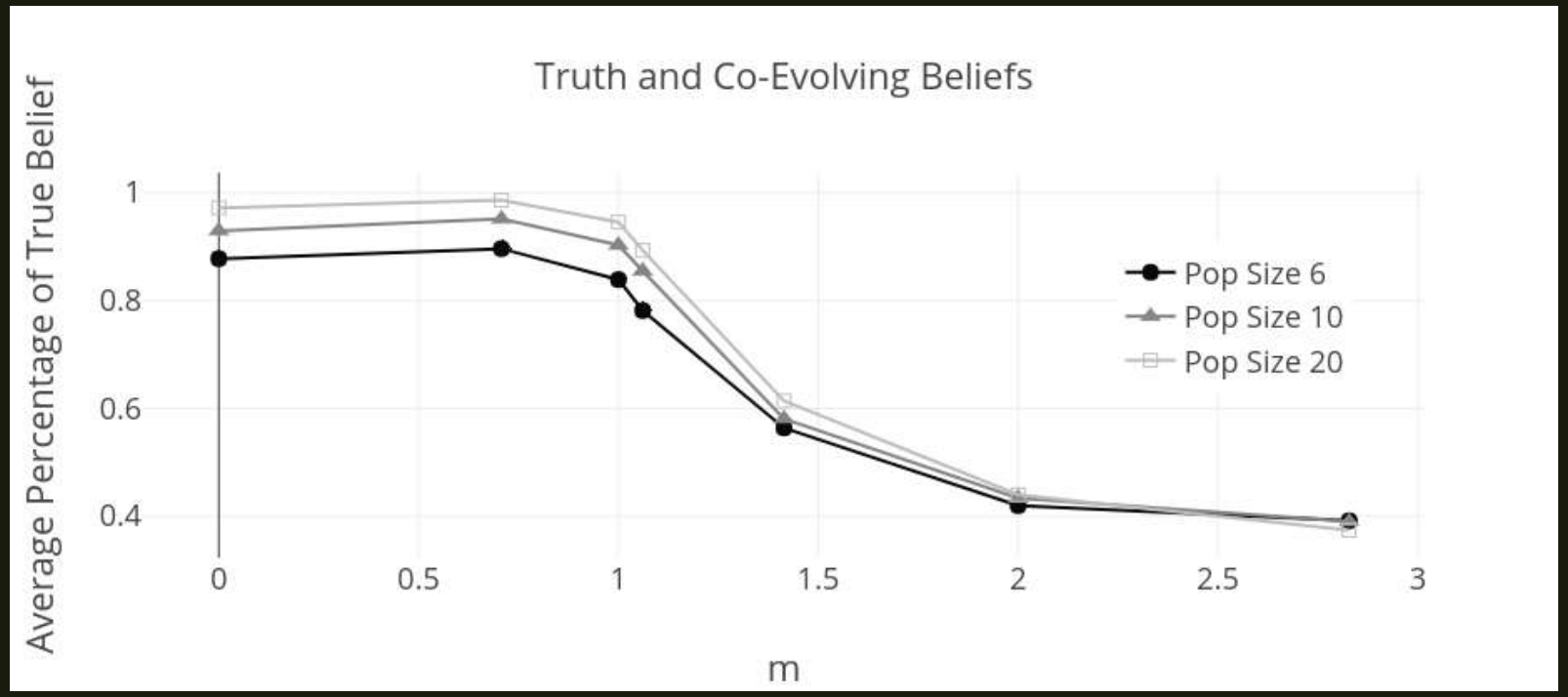
Beliefs correlate





Beliefs correlate


1
P(B2)
d
0
0
P(B1)
1



Mistrust leads to worse beliefs

Take-Aways

Correlation between completely unrelated beliefs can **emerge endogenously** when actors ground trust in shared belief.

This is happening in models where there is **no ideology or group identity** influencing this correlation.

Limitation

One major limitation of these models for explanatory purposes is that they put agents in **very good epistemic situations** – all actors are good at testing the world, and no actors actively try to mislead peers.

Limitation

One major limitation of these models for explanatory purposes is that they put agents in **very good epistemic situations** – all actors are good at testing the world, and no actors actively try to mislead peers.

In models that relax these assumptions, it can be good to disconnect, or mistrust other agents. But basing this trust on statistical properties is better than on shared beliefs.

Summary

1) Mistrust of evidence can lead to polarization, even among ‘scientific’ agents
2) When agents use multiple beliefs to determine trust, this can create endogenous epistemic factions
3) Mistrust decreases the epistemic success of communities*

Thank you!