POLARIZATION AND FACTIONALIZATION

James Owen Weatherall

Logic and Philosophy of Science, UC Irvine

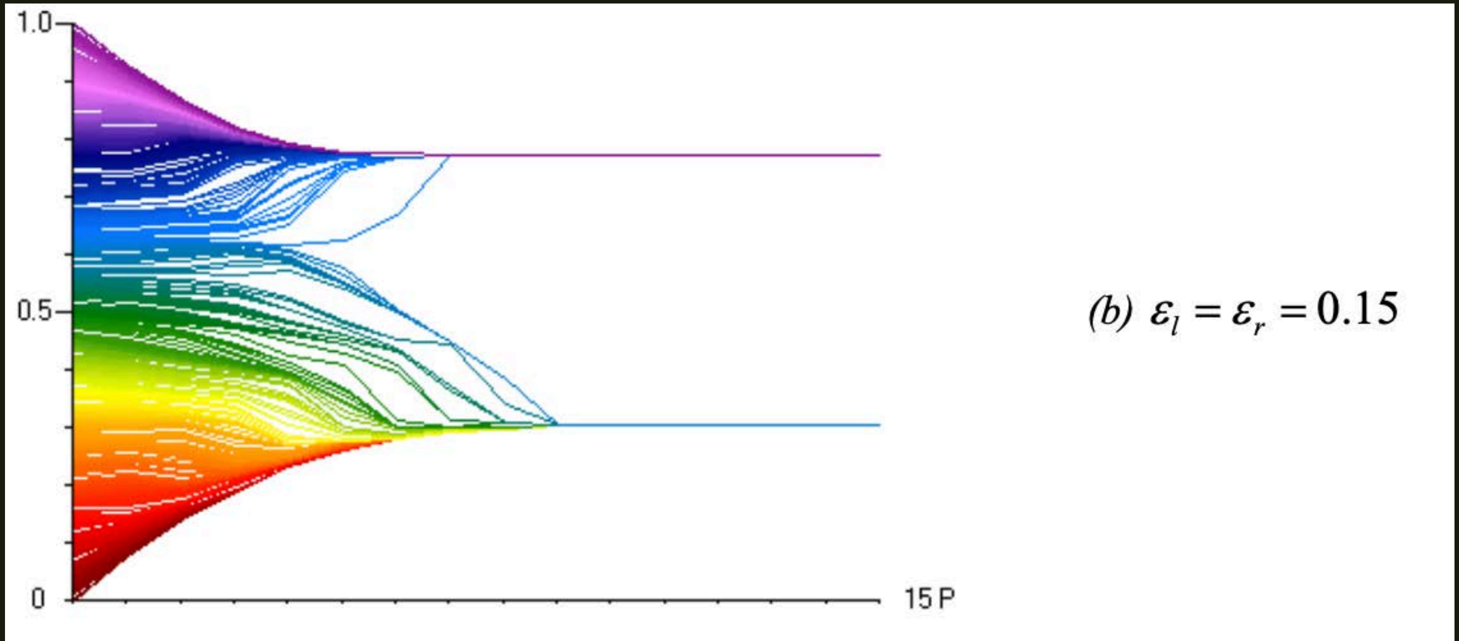
CHRONIC LYME

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.

Example: Hegselmann and Krause (2002)



Today's Plan

This talk will address the following question: why do communities sometimes polarize over matters of **scientific fact**, especially in cases where there is ample evidence?

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.

As we saw, epistemic networks also converge.

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

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 and abortion.

But in at least some cases, polarization occurs over matters of fact between actors who share values.

Our Approach

In this literature, previous work has mostly* focused on opinion dynamics.

In these models actors don’t have good reasons to hold beliefs, outside the social realm.

We look at models better tuned to scientific communities, using the **network epistemology framework**.

Adequacy

Why are these good models for the cases at hand?

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

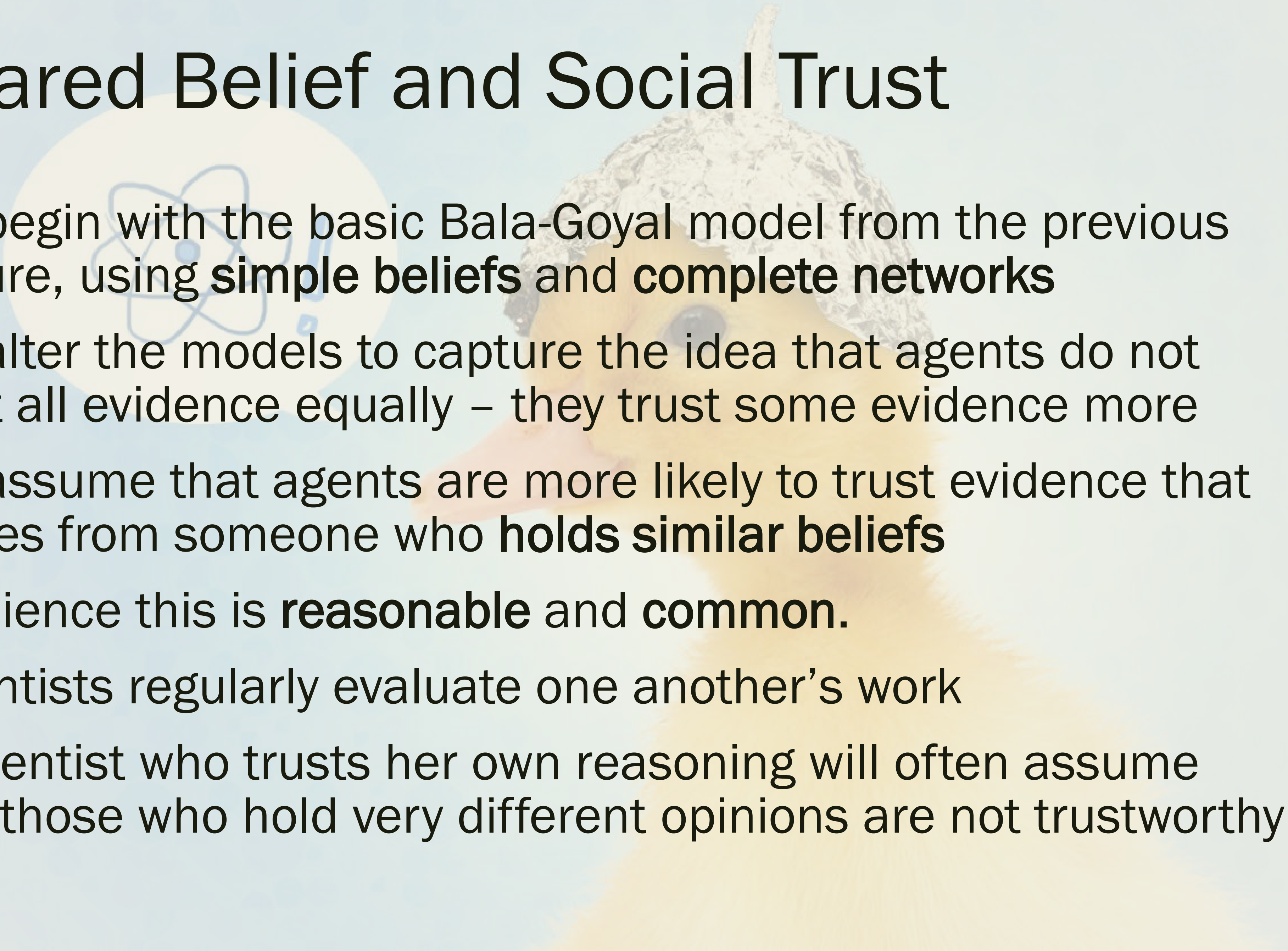
Roadmap

1) Trust Dynamics
2) Polarization in Science
3) Factionalization in Science

Roadmap

1) Trust Dynamics
2) Polarization in Science
3) Factionalization in Science

Shared Belief and Social Trust



We begin with the basic Bala-Goyal model from the previous lecture, using **simple beliefs** and **complete networks**

We alter the models to capture the idea that agents do not treat all evidence equally – they trust some evidence more

We assume that agents are more likely to trust evidence that comes from someone who **holds similar beliefs**

In science this is **reasonable** and **common.**

Scientists regularly evaluate one another’s work

A scientist who trusts her own reasoning will often assume that those who hold very different opinions are not trustworthy

Jeffrey Conditionalization

We incorporate this assumption by changing the update rule

Instead of strict Bayesian updating, agent adopt Jeffrey’s rule

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 fact obtained, **$P_f(E)$**

Their new credence given this evidence is:

$$P_f(H) = P_i(H|E) \cdot P_f(E) + P_i(H|\sim E) \cdot P_f(\sim E)$$

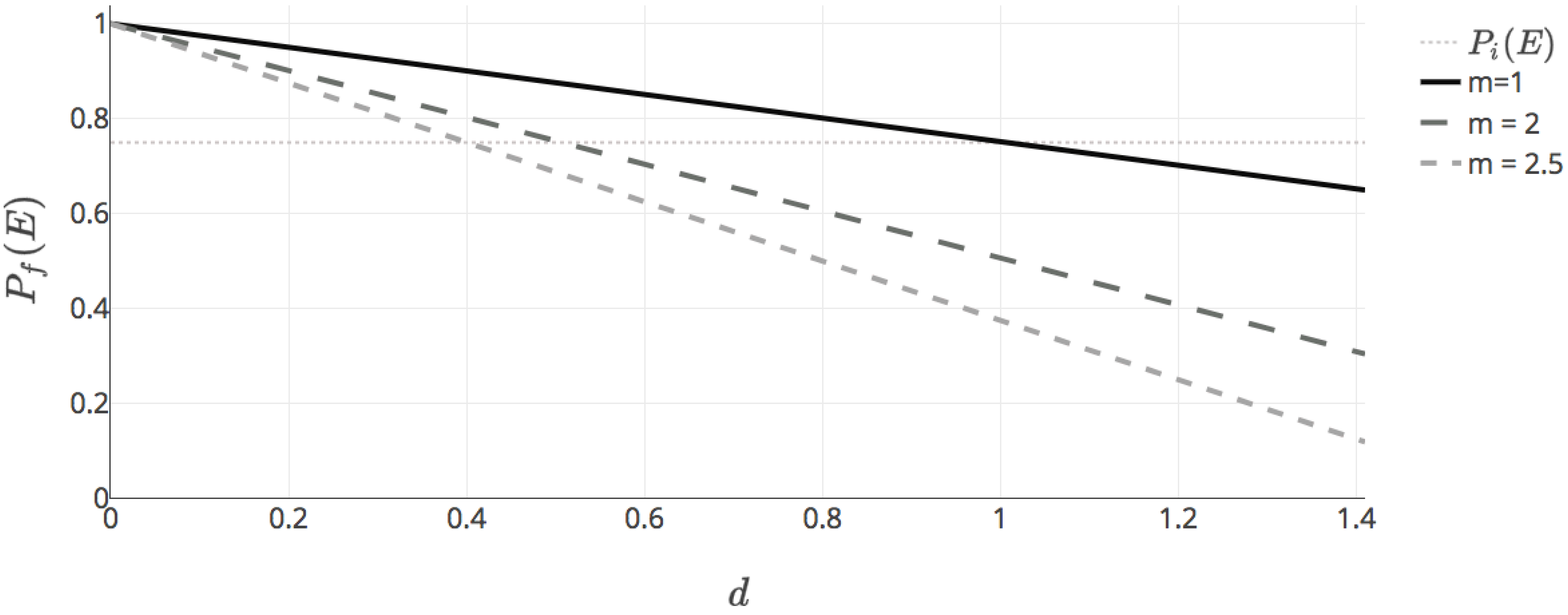
Decreasing Certainty

We model $P_f(E)$ as a **decreasing function** of the (1-dimensional Euclidean) 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.

$$P_f(E)(d) = \max(\{1 - d \cdot m \cdot (1 - P_i(E)), 0\})$$

Credibility of Evidence and Distance in Belief



Ignoring vs. Anti-updating

We consider two treatments

In one, if this function yields a value lower than their prior

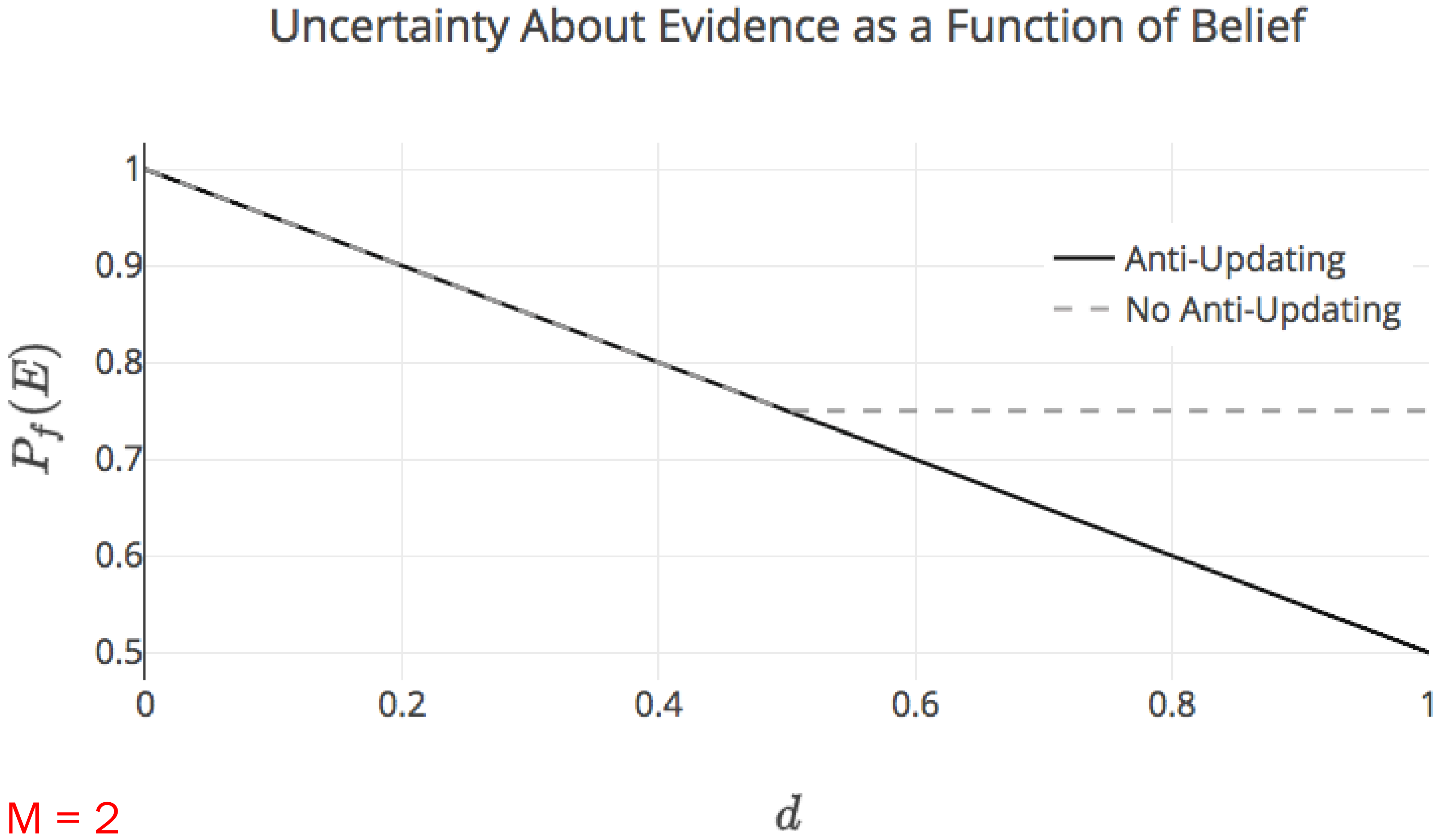
$$P_f(E) < P_i(E)$$

agents **ignore** the data

In the other treatment, they update with Jeffrey’s rule and $P_f(E)$

This pushes credences in the opposite direction: **anti-updating**

This is related to the **backfire effect**



Uncertainty About Evidence as a Function of Belief

M = 2

Roadmap

1) Trust Dynamics
2) Polarization in Science
3) Factionalization in Science

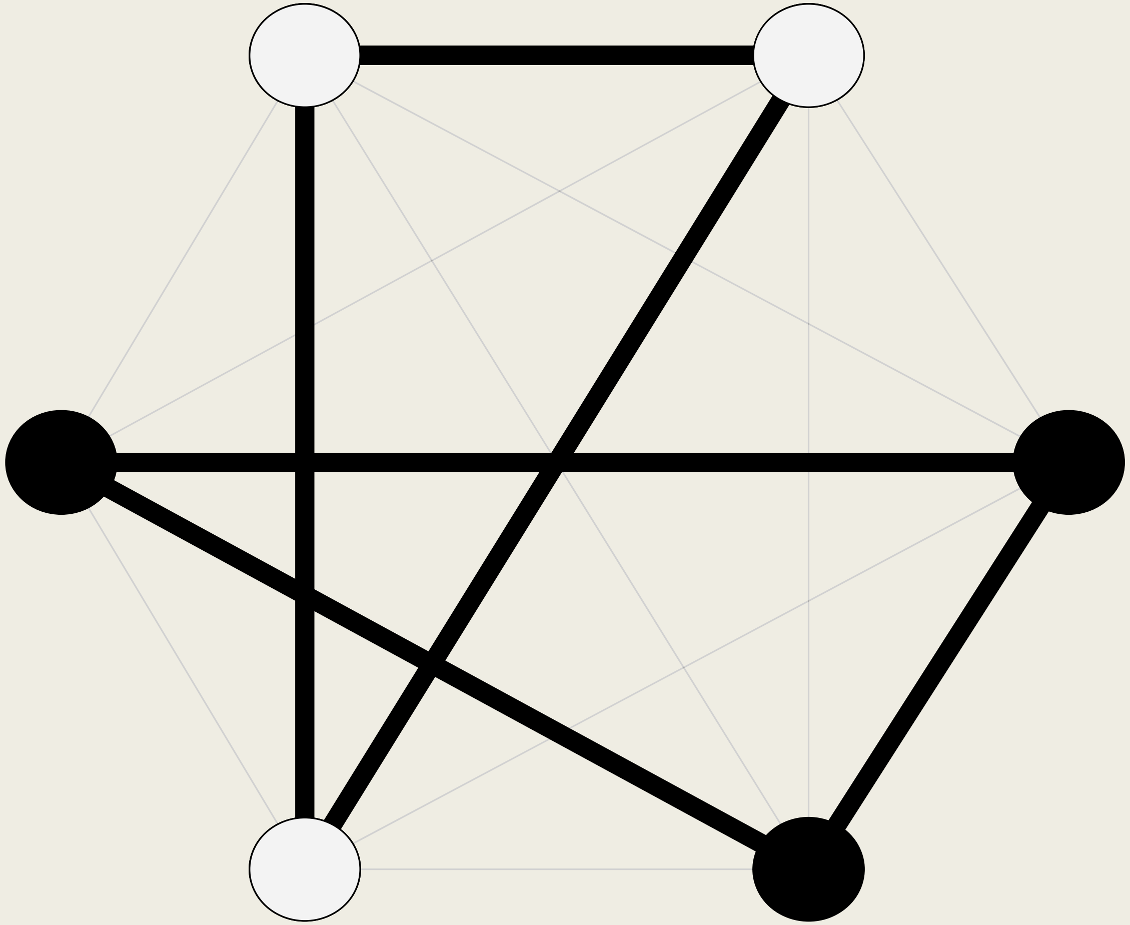
Results: No anti-updating

Adding differential trust generates a **new possibility**

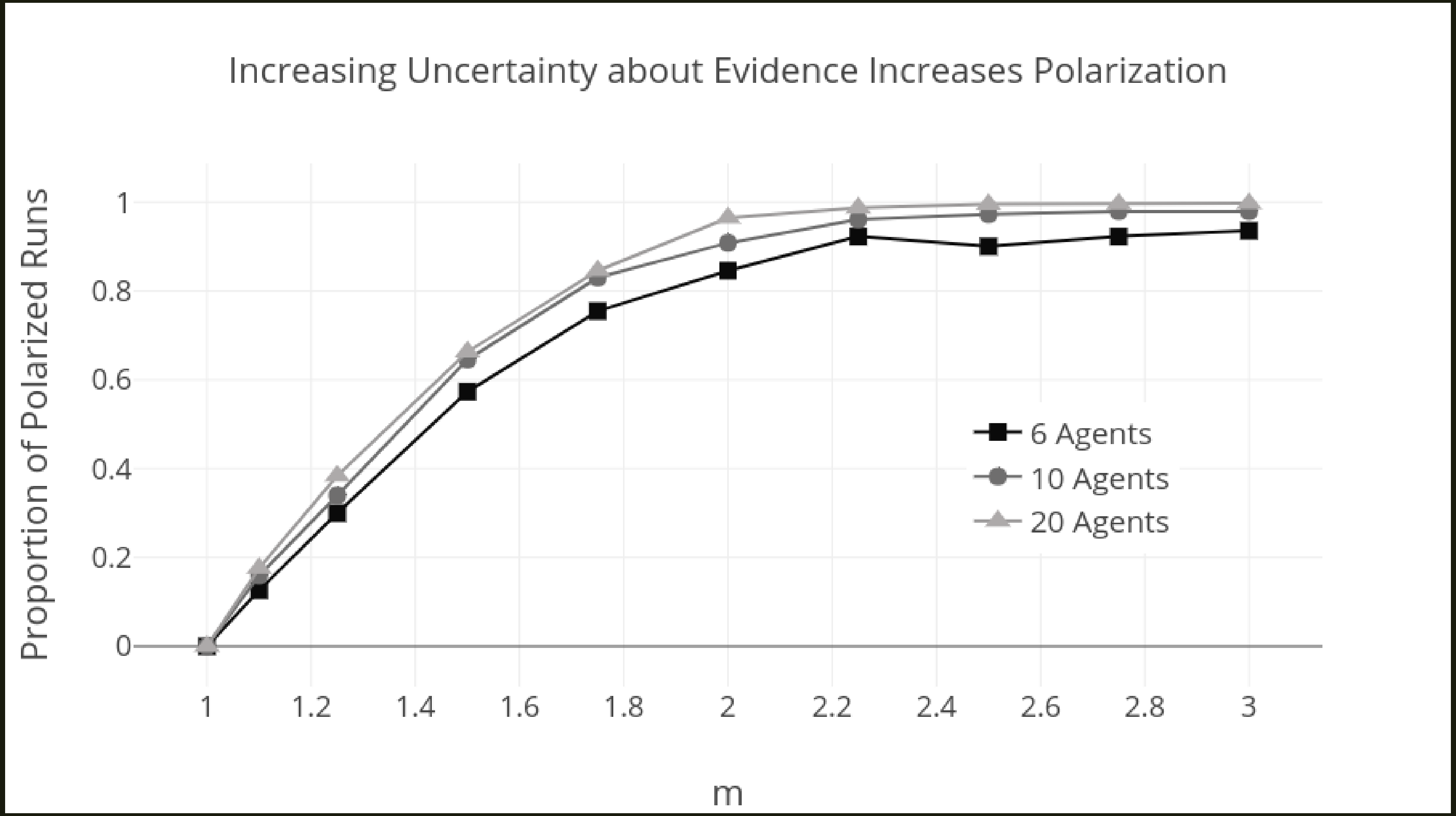
When m <1, all converge **to consensus,** either True (p(B) > .99) or False (p(B)<.5)

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

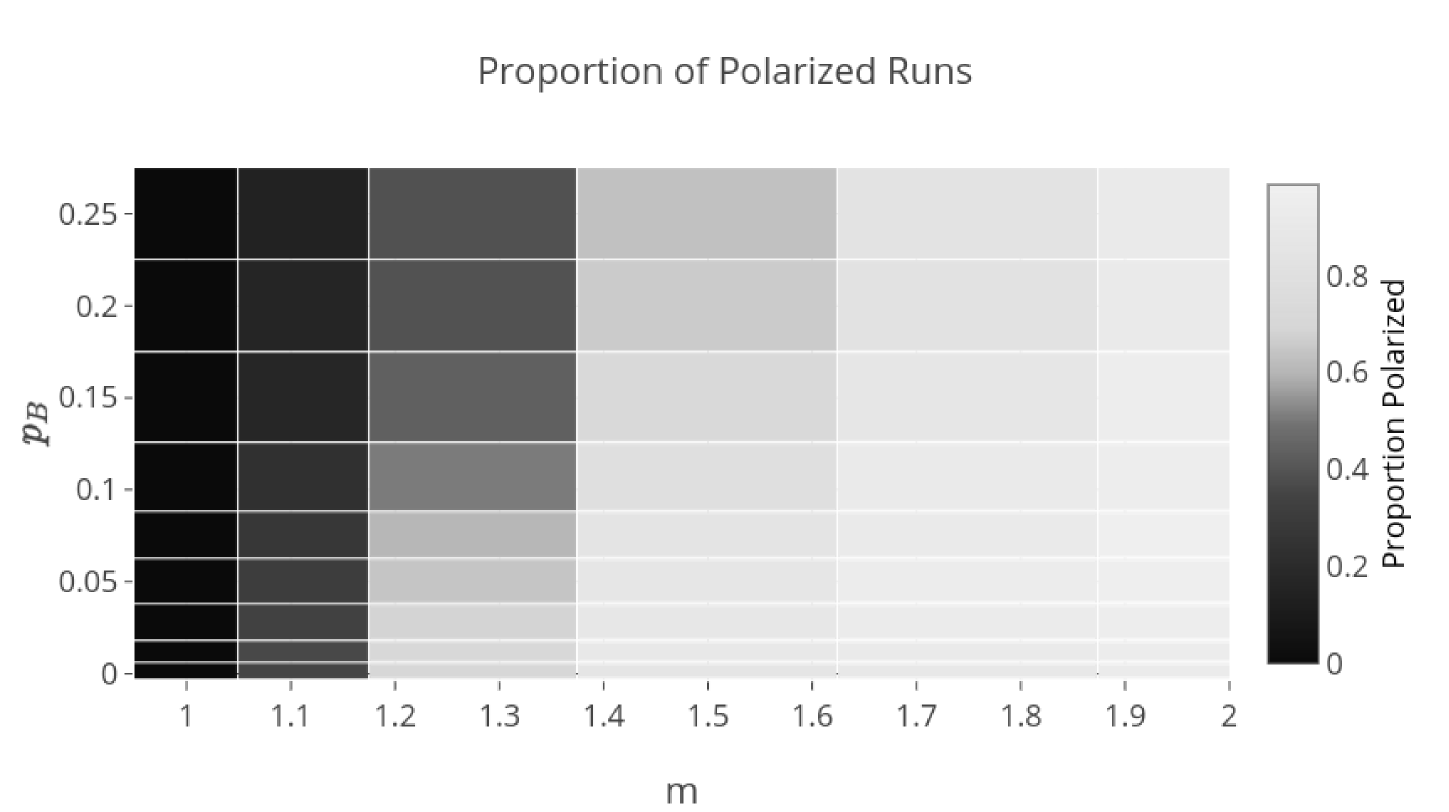
The value of m determines how low one side must be for polarization to be stable.



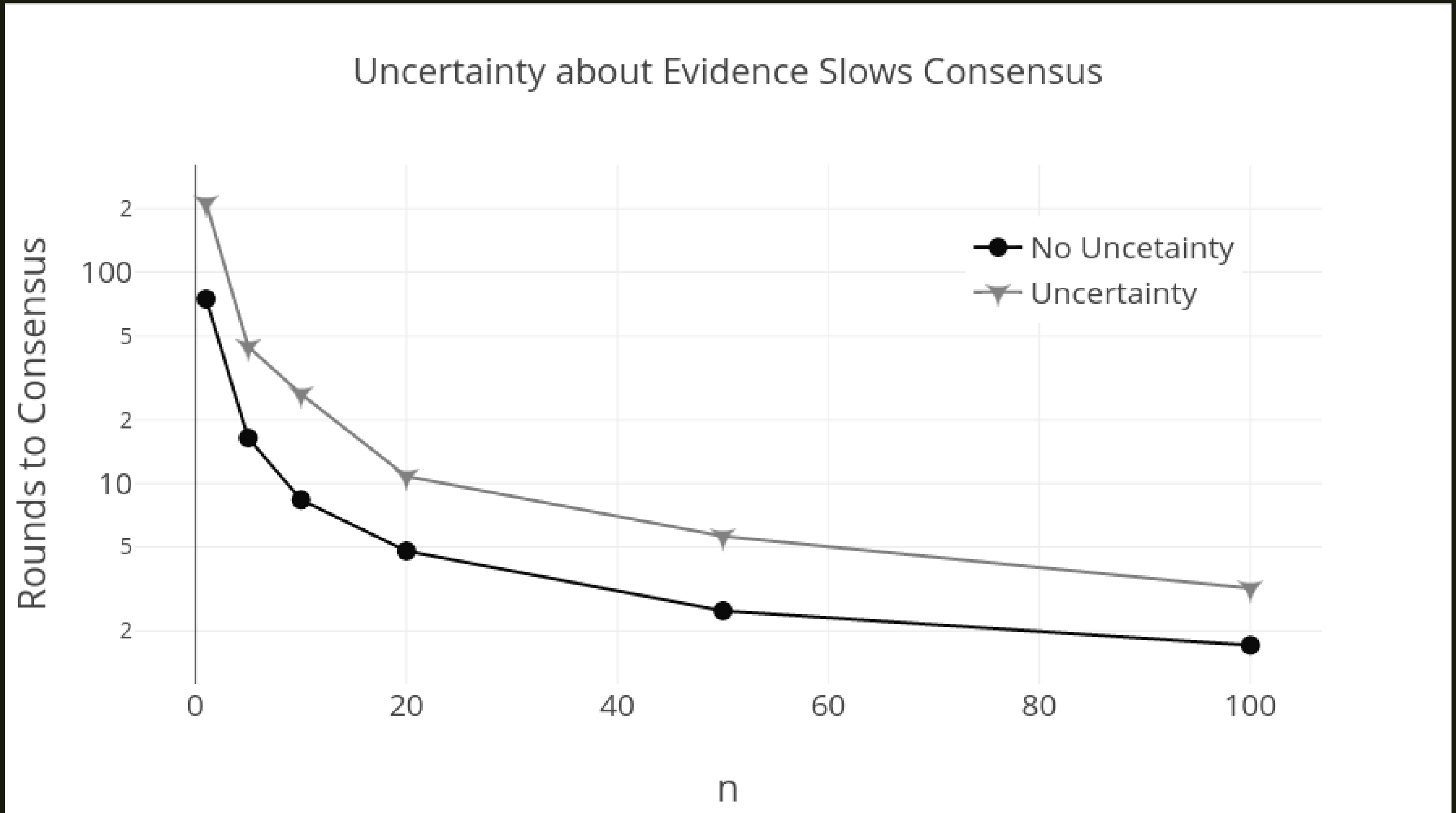
POLARIZATION



Increasing m increases polarization

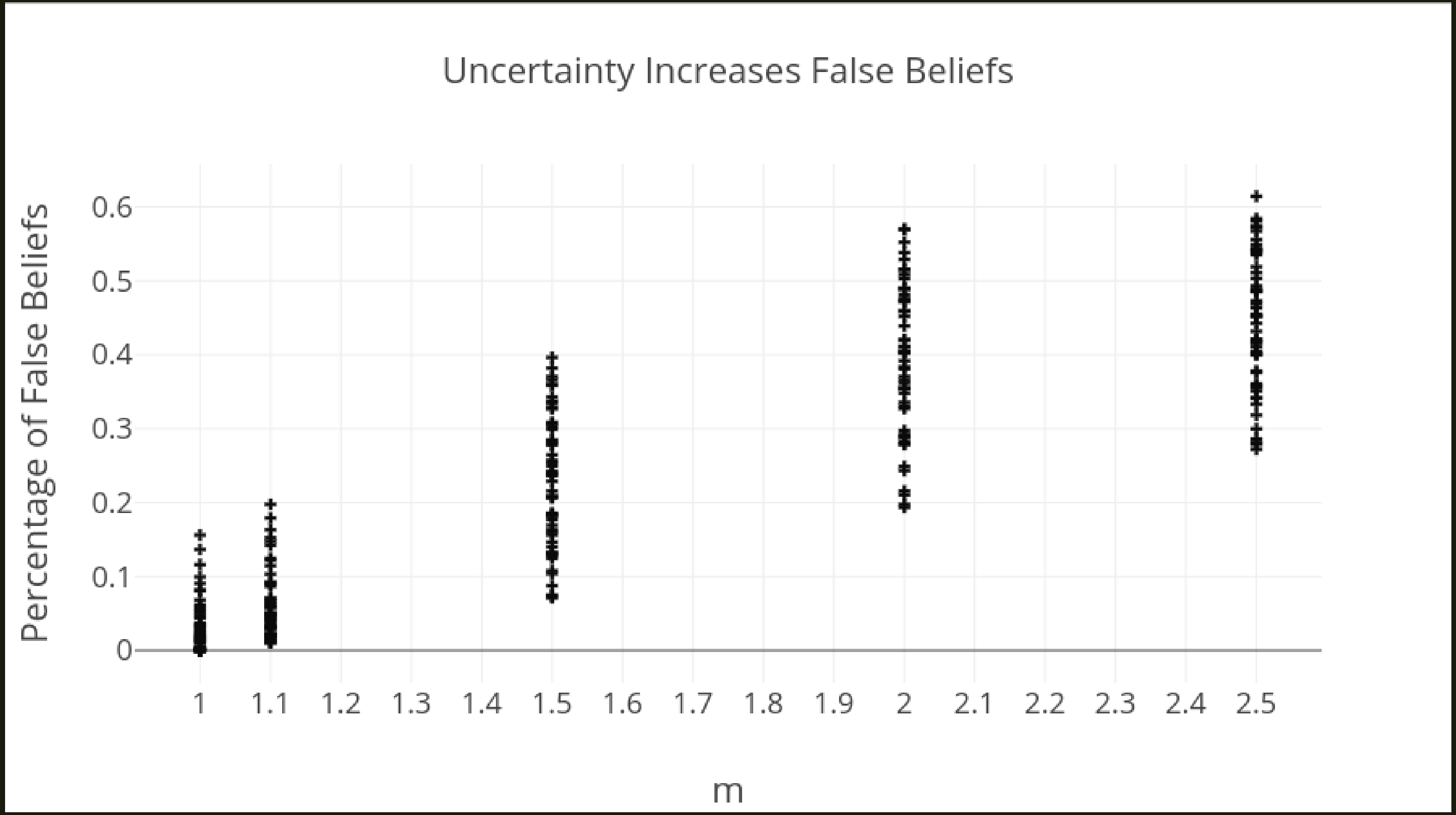


For easier problems, this is mitigated



Uncertainty slows consensus

(m = 0 vs. m = 1)



Uncertainty Increases False Beliefs

Results: Anti-Updating

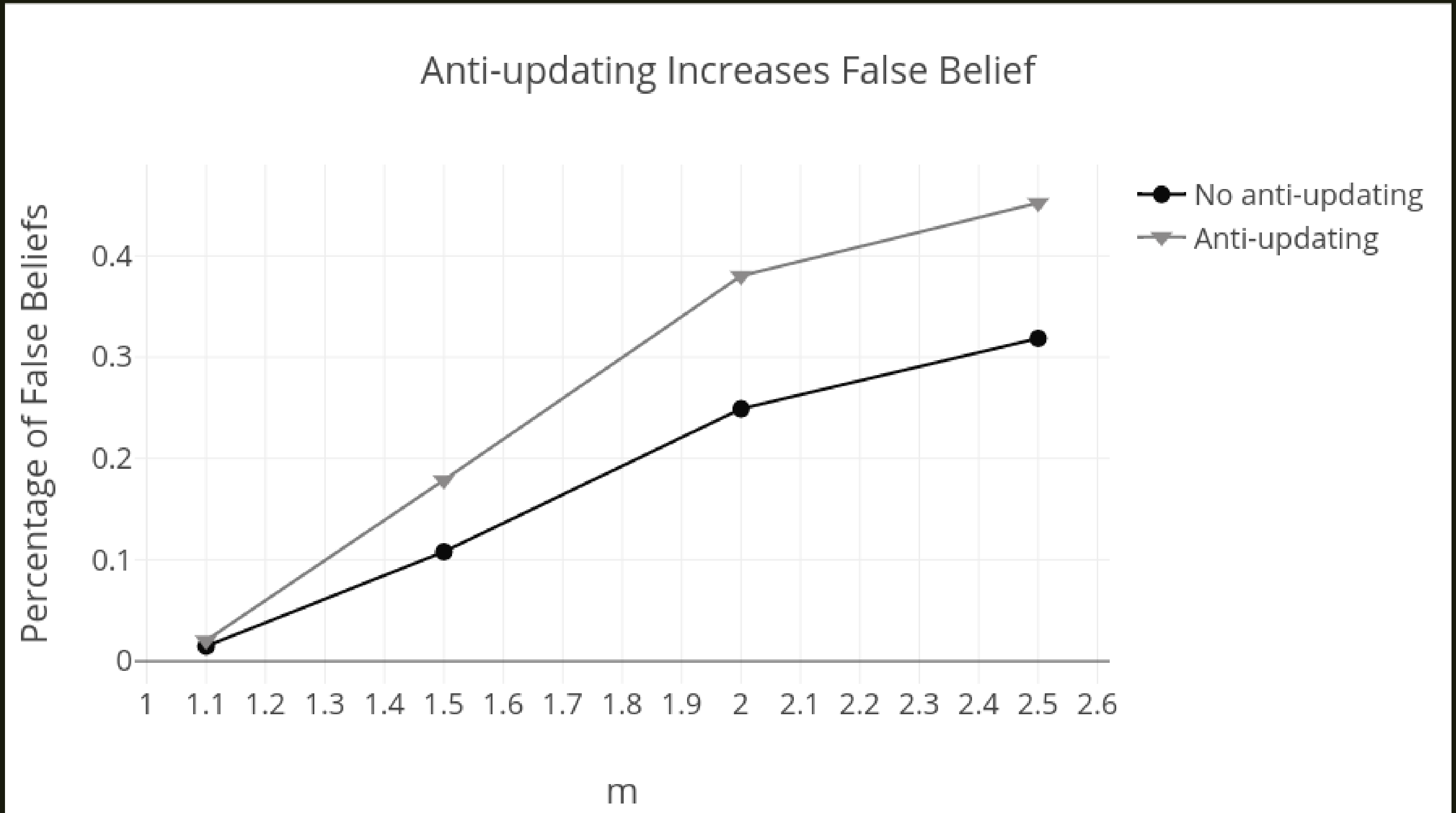
Polarization is also possible with anti-updating

Now polarized outcomes involve stable groups with very high credences ($p(B)<.99$) and very low credences ($p(B) < .01$).

In these models, polarization emerges more often than the no-anti-updating treatment

In addition, actors are **worse at learning the truth**.

Anti-updating leads to worse beliefs



Take-Away

As mentioned, scientists are often skeptical of those who have reached different conclusions

Nonetheless, this sort of skepticism, on a global level, **hurts the knowledge producing capacity of the community**

It leads to stable polarization

In a case like Lyme, we see why actors with similar values could end up with very little trust and stable polarization

Roadmap

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

Factionalization

We often see polarization where actors form **factions with multiple, shared, polarized beliefs**

This can happen even when the beliefs are apparently unrelated

In the US, beliefs about climate change are correlated with beliefs about whether evolutionary theory is correct and about gun safety

Ideology and Explanation

Previous authors have explained this via appeal to **shared ideology.**

George Lakoff claims that 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?

Can such bundles can **emerge endogenously** as a result of social trust grounded in shared beliefs?

Previous Work

Axelrod (1997) presents a model that involves a grid of “cultures”, with variants represented by a list of integers.

Shared integers in each “spot” increase the chances that a variant will spread from one culture to another.

The end result is patches of stable cultures with entirely shared variants

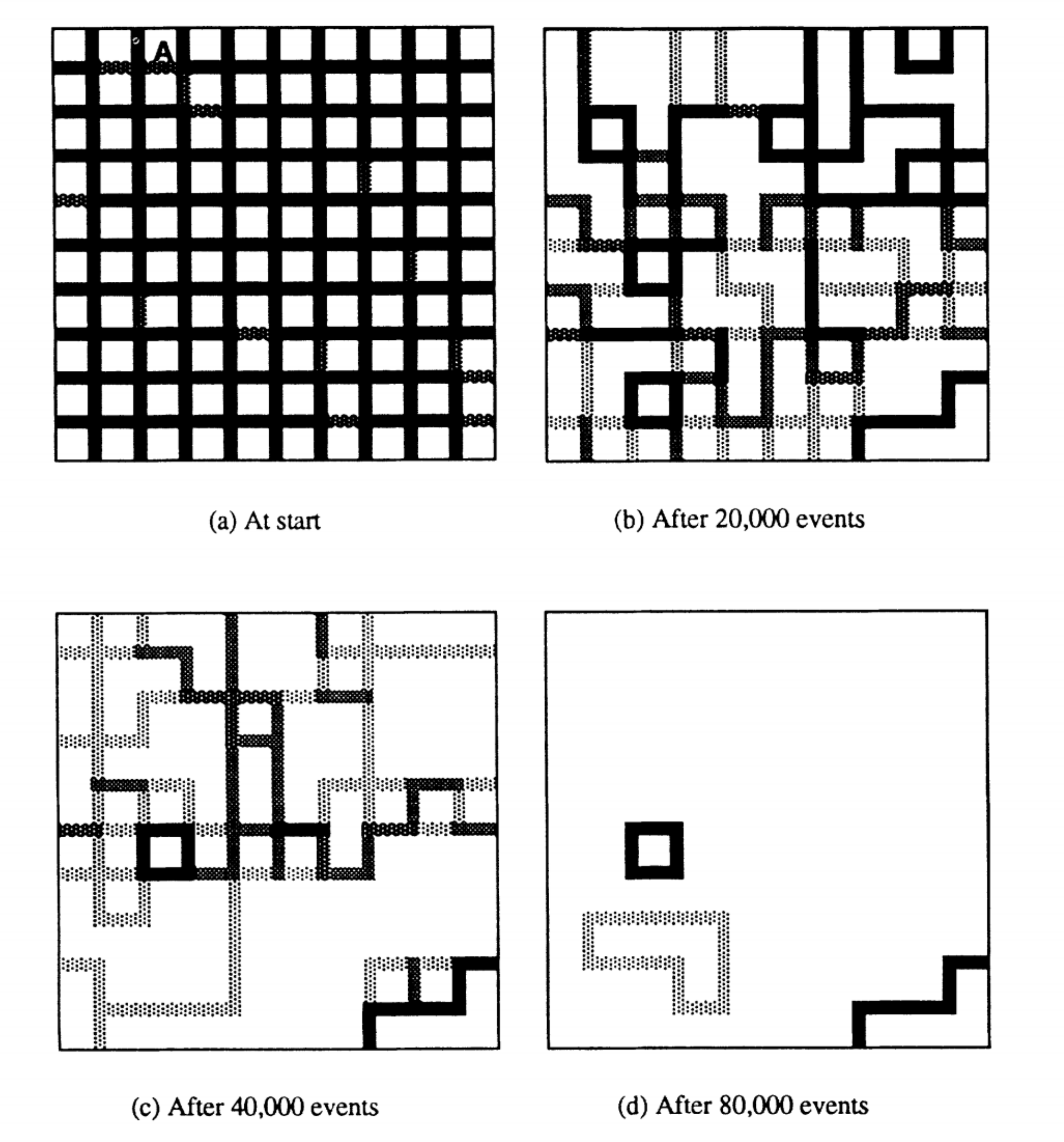
There is no sense in which actors have good reasons to hold these variants and there is no evidence

TABLE 1

A Typical Initial Set of Cultures

74741	87254	82330	17993	22978	82762	87476	26757	99313	32009
01948	09234	67730	89130	34210	85403	69411	81677	06789	24042
49447	46012	42628	86636	27405	39747	97450	71833	07192	87426
22781	85541	51585	84468	18122	60094	71819	51912	32095	11318
09581	89800	72031	19856	08071	97744	42533	33723	24659	03847
56352	34490	48416	55455	88600	78295	69896	96775	86714	02932
46238	38032	34235	45602	39891	84866	38456	78008	27136	50153
88136	21593	77404	17043	39238	81454	29464	74576	41924	43987
35682	19232	80173	81447	22884	58260	53436	13623	05729	43378
57816	55285	66329	30462	36729	13341	43986	45578	64585	47330

SPREAD OF CULTURE



(a) At start

(b) After 20,000 events

(c) After 40,000 events

(d) After 80,000 events

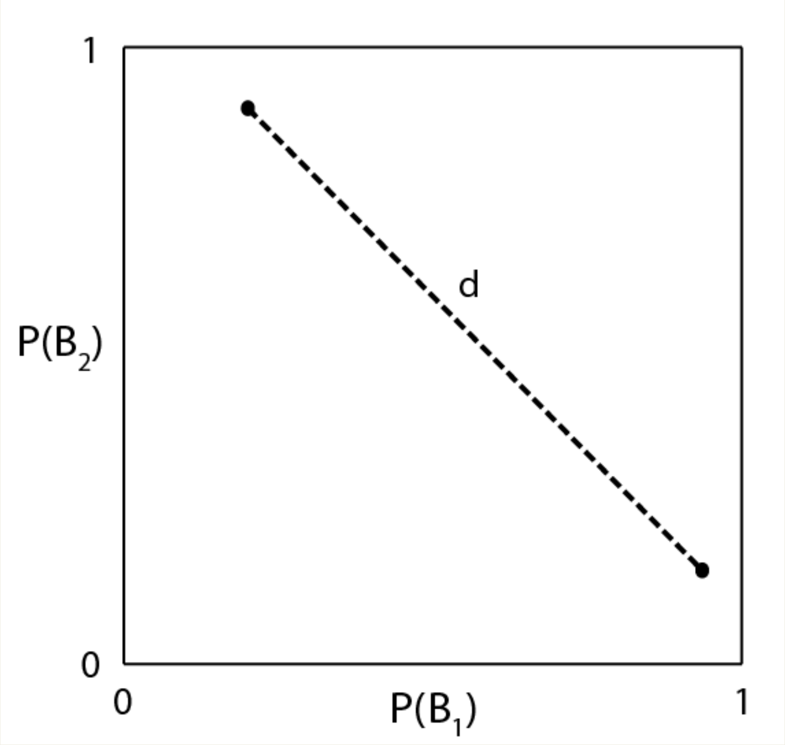
Our Model

We consider a variant on the trust model, where actors hold **multiple beliefs.**

We use the same formula to determine certainty in evidence

Now the distance is a (higher-dimension) Euclidean distance between their beliefs

For two dimensions, this distance can range from 0 to 1.41.



Outcomes

Now outcomes are **slightly more complicated.**

We see all combinations of True, False, and Polarized beliefs in all 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

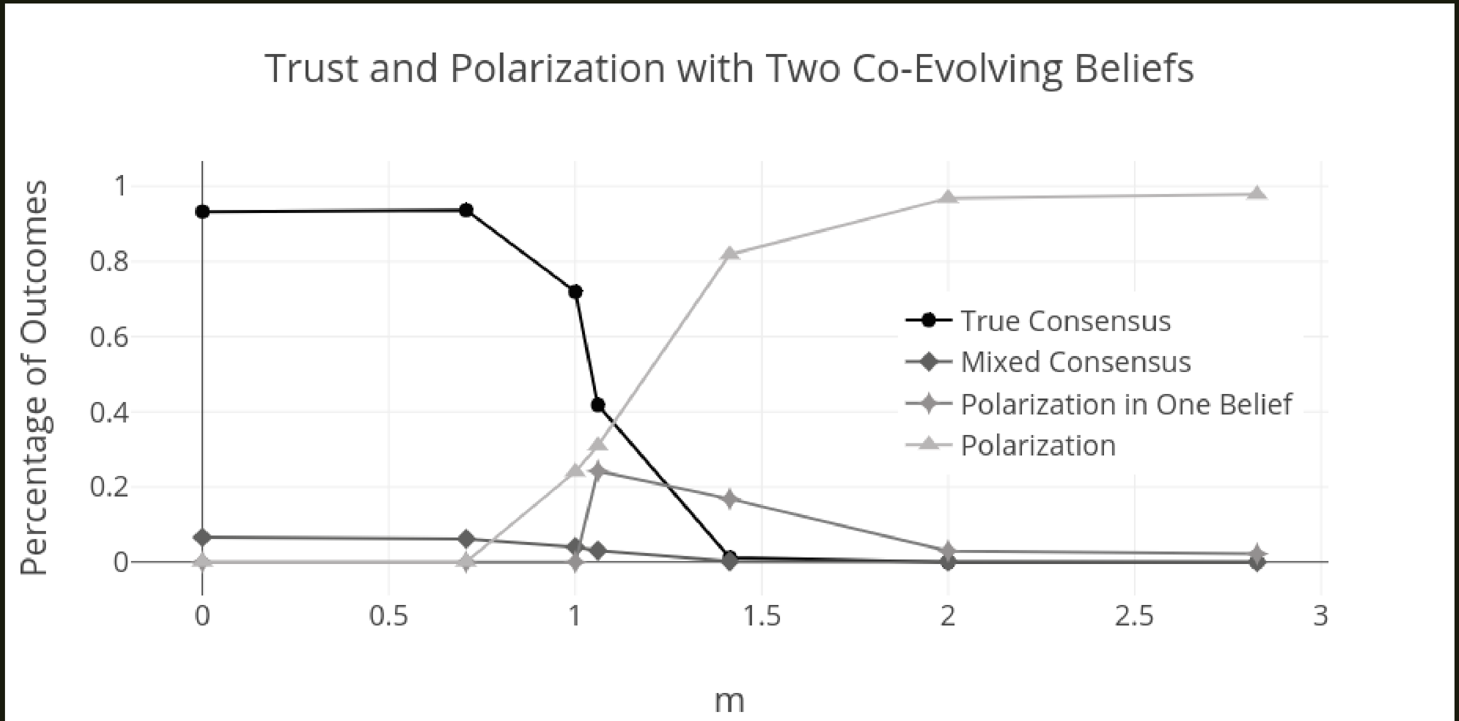
We look at three treatments.

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

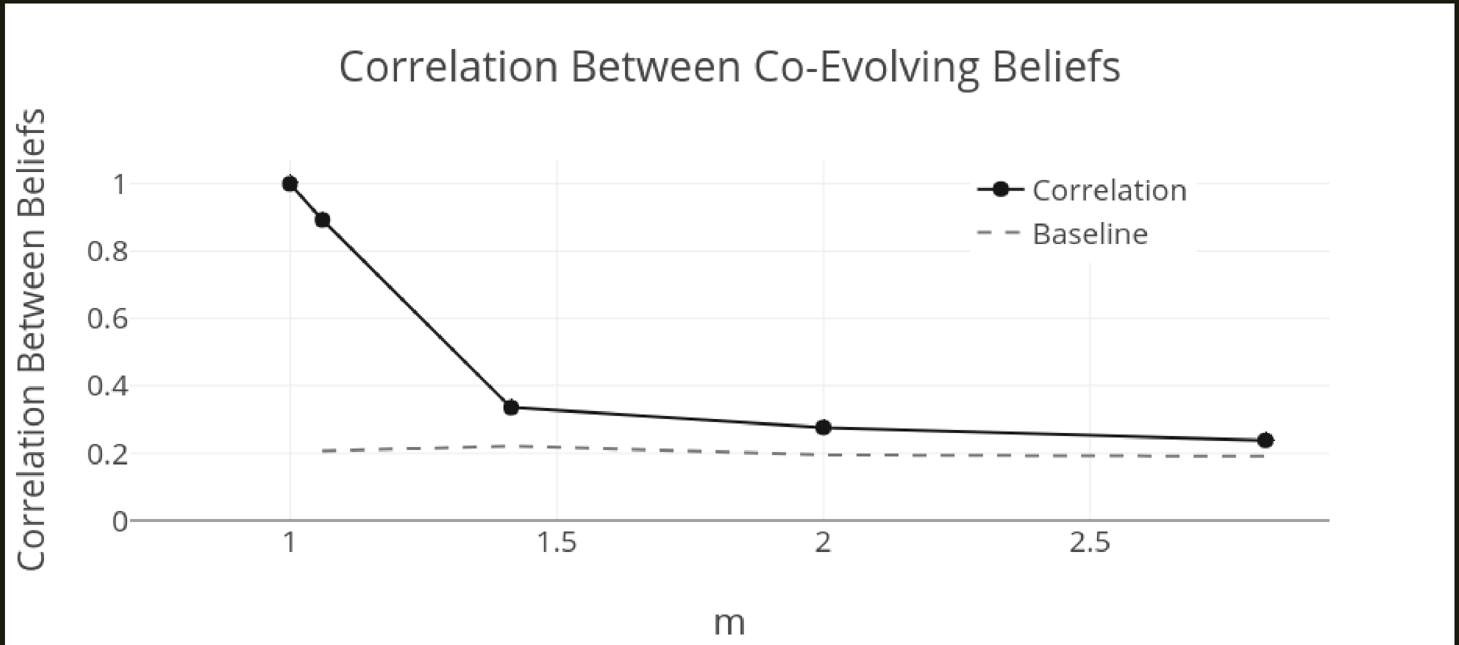
Coevolving Beliefs

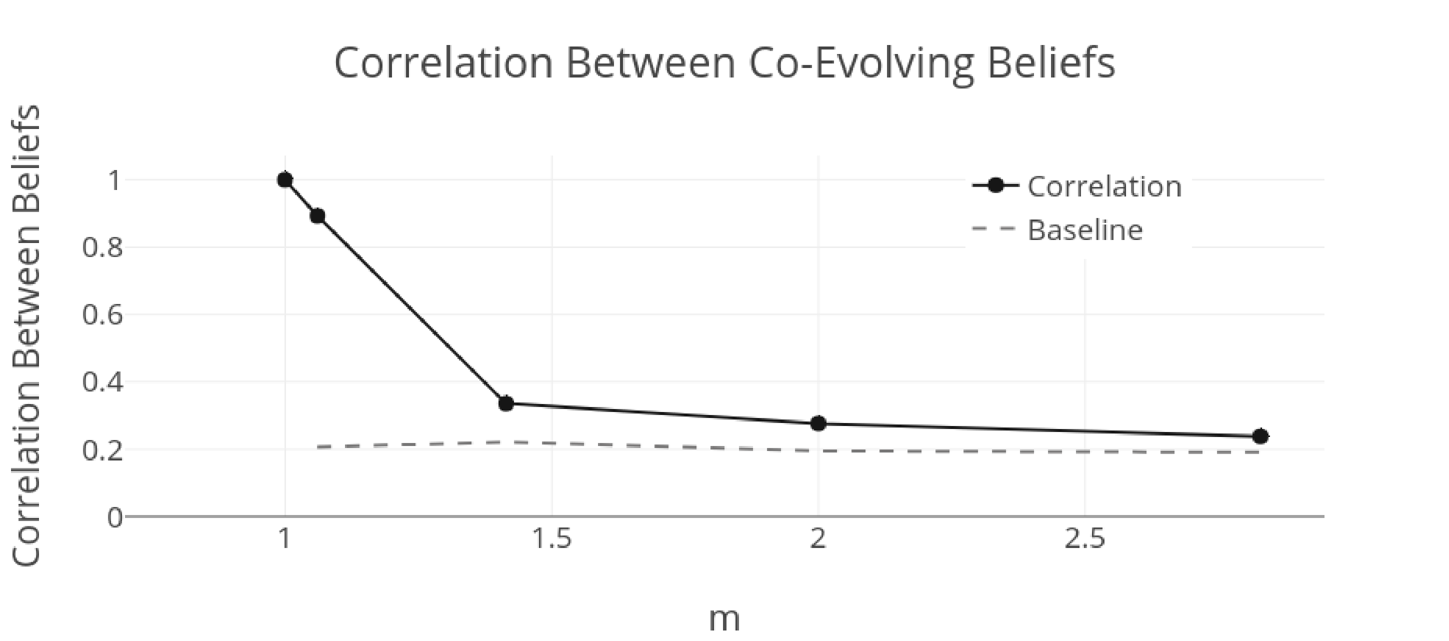
Since results are qualitatively similar across these three treatments, we look at the models where beliefs co-evolve from the start.

Mistrust increases multiple-arena polarization

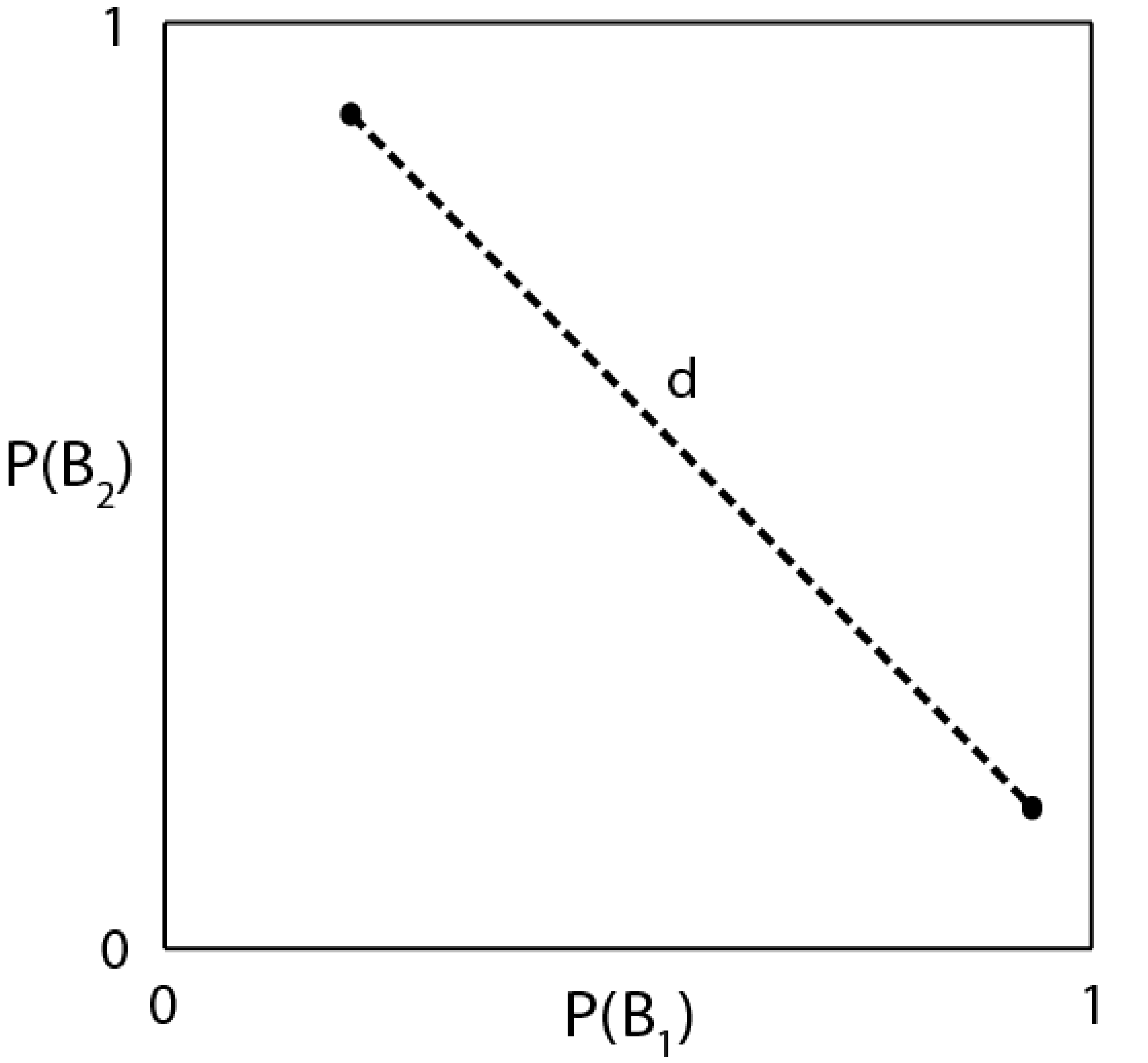


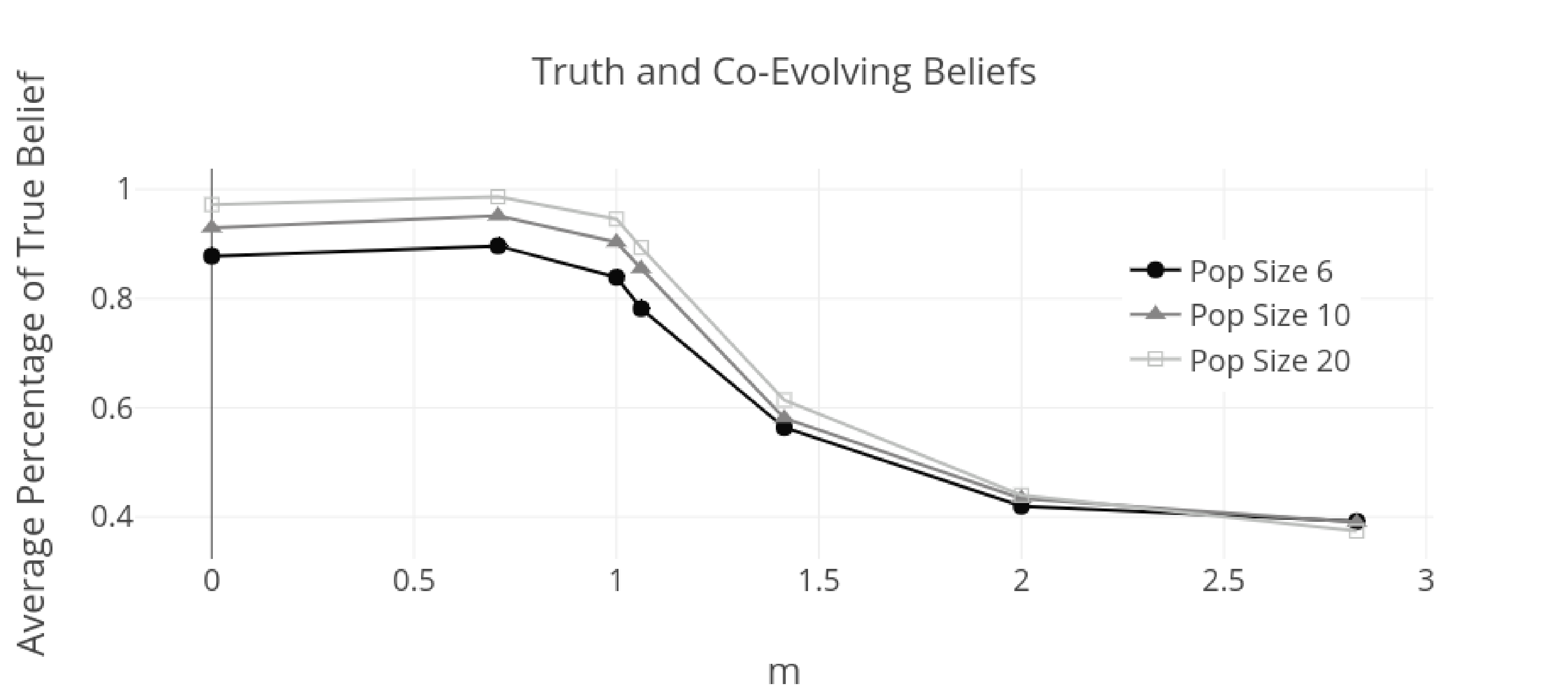
Beliefs correlate





Beliefs correlate


1
P(B$_2$)
0
d
0
P(B$_1$)
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.

Limitations

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

Thus, even ideal agents can polarize under these dynamics

With non-ideal agents, it can be good to disconnect, or mistrust other agents, especially if they are unreliable

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.

Ways to Mistrust

Holman and Bruner (2015) consider actors in epistemic network models who share **biased data** in an attempt to sway the community (i.e., data drawn from a biased bandit arm).

As they point out, agents do better when they discount data from those whose results are **statistically unlikely given their current beliefs**. This still can cause a sort of polarization, though.

Ideally, in these models, agents would discount data from a source that **consistently is a statistical outlier**.

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
4) In cases where it is necessary, it is best done on statistical grounds

Thank you!