International Journal of Approximate Reasoning 172 (2024) 109240

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Frequentist belief update under ambiguous evidence in social networks ☆

Michel Grabisch [a], M. Alperen Yasar [a,b,*]

[a] *Applied Mathematics, Université Paris 1, Panthéon-Sorbonne, France*
[b] *Economics, Ca' Foscari University of Venice, Italy*

ARTICLE INFO

Keywords:
Dempster-Shafer theory
Agent-based modeling
Social networks
Ambiguous evidence
Subjective belief update
Opinion dynamics

ABSTRACT

In this paper, we study a frequentist approach to belief updating in the framework of Dempster-Shafer Theory (DST). We propose several mechanisms that allow the gathering of possibly ambiguous pieces of evidence over time to obtain a belief mass assignment. We then use our approach to study the impact of ambiguous evidence on the belief distribution of agents in social networks. We illustrate our approach by taking three representative situations. In the first one, we suppose that there is an unknown state of nature, and agents form belief in the set of possible states. Nature constantly sends a signal which reflects the true state with some probability but which can also be ambiguous. In the second situation, there is no ground truth, and agents are against or in favor of some ethical or societal issues. In the third situation, there is no ground state either, but agents have opinions on left, center, and right political parties. We show that our approach can model various phenomena often observed in social networks, such as polarization or bounded confidence effects.

1. Introduction

Emerging behaviors of human beliefs are complex by nature. Both Bayesian and heuristic approaches have been used to study these emerging phenomena in social networks (e.g., polarization, bounded confidence, . . .) [1,2]. However, empirical observations suggest that such models, while robust, do not fully encapsulate the nuances of real-world belief evolution [3] (see [4] for a theoretical paper comparing Bayesian models to heuristic models).

Agents with similar priors may have vastly different posteriors after observing the same piece of evidence [5]. One of the explanations for this phenomenon is the *ambiguity* of evidence. Updating beliefs under ambiguity has been challenging in the Bayesian setting. For example, [6] assumes that an agent with a bias towards one of the events interprets ambiguous evidence as unequivocally supporting that favored event. It is within this context that the Dempster-Shafer Theory (DST) emerges as a potent alternative, incorporating the rationality of Bayesian models while still allowing for the intricacies of heuristics. An advantage of DST is that it need not model singleton evidence solely but rather can model evidence that might support multiple states of nature simultaneously. Therefore, two agents with identical betting probabilities, the same priors on the surface, might have two different mental models

☆ This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Actions grant agreement No 956107, "Economic Policy in Complex Environments (EPOC)".

* Corresponding author at: Economics, Ca' Foscari University of Venice, Italy.
E-mail address: alperen.yasar@unive.it (M.A. Yasar).

https://doi.org/10.1016/j.ijar.2024.109240
Received 30 January 2024; Received in revised form 11 June 2024; Accepted 12 June 2024

Available online 17 June 2024

leading to different posteriors [7] (see [8] for an explanation of how non-additive probability approaches are suitable for modeling ambiguous evidence).

DST has two mainstream belief update methodologies: combination rules [9] and conditioning rules [10] (see [11] for a comparison of both). Both of these methods can cause radical changes in one's belief under completely opposite evidence, which might violate *belief persistence* [12]. Belief persistence, also known as the *conservativity principle*, is the notion that an agent updating its beliefs in response to new evidence should remain as faithful as possible to its original beliefs [13]. More importantly, if the new information is not compatible with the old belief set, any change in beliefs must be minimal. Dempster's combination rule, on the other hand, might allow an agent to completely persuade another one, given enough certainty. For this reason, we propose a frequentist update methodology for DST. Our frequentist rule acts as an evidence aggregation mechanism where every atom of evidence only nudges an agent towards a direction, which is similar to heuristic approaches [14]. Using this methodology, we create several update rules for different cases and then show the emergent behavior of these updates in three different agent-based models.

The contribution of this paper is three-fold. Our first and foremost contribution is that we create a frequentist update framework for DST, addressing the challenges posed by the violation of *belief persistence*. Second, we propose several update rules for different scenarios and study their properties. Finally, we explore the application of DST in modeling belief dynamics within social networks, providing insights into emerging phenomena such as polarization or bounded confidence models. Polarization refers to situations where an opinion or belief in an event induces a clear partition of the society into two or more groups of radically different opinions/beliefs. On the other hand, bounded confidence models are characterized by the fact that agents do not listen to people whose opinions are too far from their own opinions. We demonstrate the efficacy of DST in managing ambiguous evidence, showing its unique ability to capture and process uncertainty.

The rest of the paper will be organized as follows: in Section 2, we will present the basic notation of DST that will be used later. In Section 3, we will create a frequentist framework for DST, suggest several update rules for this framework, and study the properties of these rules. Section 4 will study the emergent phenomena of DST in social networks using agent-based modeling, explaining the differences between the combination rule and the frequentist rule. Finally, Section 5 will conclude the paper with a summary of our results.

2. Basic concepts and notation

2.1. The Dempster-Shafer framework

We present basic notions of Dempster-Shafer theory (DST), which are necessary for our exposition. A detailed presentation can be found in [15]. Let Ω be a finite set of n *outcomes* (or states of nature, etc.), and 2^Ω be the power set denoting all possible subsets of outcomes, called *events*. In Dempster-Shafer Theory, an agent's belief about events is modeled by a *basic mass assignment* or *mass distribution*, abbreviated hereafter by bma, which is a function $m : 2^\Omega \to [0,1]$, satisfying $\sum_{X\subseteq\Omega} m(X) = 1$. Given some evidence received by the agent, $m(X)$ represents the (quantity of) belief committed exactly to the event X and not to any proper subset of X. The amount of belief committed to Ω represents the level of ignorance of the agent, while the amount of belief committed to the empty set represents the belief that Ω does not contain all possible outcomes (open world hypothesis). Throughout this paper, we assume $m(\emptyset) = 0$. We denote by $\mathcal{M}(\Omega)$ the set of all bmas on Ω satisfying the latter property.

Following [9], we introduce, for a given mass distribution m of some agent,

- The *belief function* $Bel : 2^\Omega \to [0,1]$, assigning to every event X the quantity $Bel(X) = \sum_{Y\subseteq X} m(Y)$, interpreted as the degree of certainty/belief that event X realizes, i.e., the certainty that the true outcome or state of nature lies in X.
- The *plausibility function* $Pl : 2^\Omega \to [0,1]$, assigning to every event X the quantity $Pl(X) = \sum_{Y\cap X\neq\emptyset} m(Y)$, interpreted as the degree of plausibility that event X realizes, i.e., the plausibility that the true outcome or state of nature lies in X.

2.2. Combination rule

Consider two agents i, j, with bmas m_i, m_j, coming from two sets of evidence supposed to be both trustable and independent. *Dempster's rule of combination* allows to combine these two sets of evidence and to produce a new mass distribution $m_i \oplus m_j$ defined as follows:

$$(m_i \oplus m_j)(X) = \frac{\sum_{Y\cap Z=X} m_i(Y)m_j(Z)}{1-\kappa}, \quad \emptyset \neq X \subseteq \Omega, \tag{1}$$

and $(m_i \oplus m_j)(\emptyset) = 0$, provided $1-\kappa \neq 0$. The quantity $\kappa = \sum_{Y\cap Z=\emptyset} m_i(Y)m_j(Z)$ is the *level of conflict*, and quantifies to what extent the two pieces of evidence contradict each other.[1] It is easy to check that $m_i \oplus m_j$ is indeed a bma.

Finally, we borrow Smets and Kennes's [7] transferable beliefs to model how decisions are made from beliefs (see [16] for a detailed literature review on different-making frameworks). Formally, a betting probability, denoted by BetP, indicates that this is the probability P the agent would bet on. BetP is defined as:

[1] Smets and Kennes [7] do not perform normalization by the denominator, since they do not assume that the mass given to $\emptyset$ has to be equal to 0. This approach can be used when the open-world hypothesis is assumed.

$$\mathtt{BetP}(\mathtt{x}) = \sum_{\mathtt{A}\ni\mathtt{x}} \frac{\mathtt{m}(\mathtt{A})}{|\mathtt{A}|}. \tag{2}$$

It is well known that BetP is nothing other than the Shapley value [17] of Bel, the belief function of the credal level.

3. Frequentist belief updates in DST

We consider as above a finite set Ω of n possible states of nature or outcomes. Each agent has a prior belief in the true state of nature, which is represented by a bma denoted by m_0. At each time step $t = 1, 2, \ldots$, agents receive a signal or evidence e_t giving information about the true state of nature under the form of a non-empty subset of Ω. If $e_t = \{\omega\}$ for some $\omega \in \Omega$, the evidence is said to be *non-ambiguous*, while if e_t is not a singleton, then the evidence is said to be *ambiguous*. We do not discard the case $e_t = \Omega$, considered as fully ambiguous, and bringing in principle no information.

Upon receiving a signal, each agent updates her bma by some mechanism, and we denote by m_t the bma of the agent at time t, after having received signal e_t and updated.

In the sequel, we consider several updating rules, which are based on a frequentist view rather than on Dempster's rule of combination, mimicking the way a probability is estimated by repeated observations via the law of large numbers. To the best of our knowledge, we did not find any previous attempts in this direction, as all updating mechanisms in DST are based either on Dempster's rule of combination or one of its variants or by conditioning like in Bayesian updating.

The presence of ambiguous evidence, together with different possible attitudes of the agent towards new information, opens the field of many possible updating rules. We begin by proposing a general form, from which several particular cases will be drawn.

3.1. The general rule

We start by stating two principles which will lead us to the expression of a general updating rule.

Our first principle is the *frequentist principle*: we *accumulate with equal weights* noisy and possibly ambiguous observations, which are supposed to be independent realizations produced by nature and reflecting some unknown true state. Accumulation means that, upon reception of a new piece of evidence, no state of nature is discarded, contrary to conjunctive combination (e.g., Dempster's rule). It differs also from disjunctive combination, since receiving twice the same observation counts twice. On the other hand, the equal weight assumption implies that there is no decay nor forgetting effect in the accumulation (all weights of the observations are equal). In accordance with the mainstream of the literature on DST and in particular on evidence combination, we consider that updating operates on the bma level.

Formalizing the previous considerations, let m_0 denote the initial bma of the agent under consideration, and suppose that the observation/evidence received at time $t = 1$ is e_1. The frequentist principle assesses that this observation is equally combined with m_0, that is:

$$m_1 = \frac{m_0 + f(e_1)}{2}$$

where f is some function translating the evidence e_1 into something similar or compatible with a mass assignment function, i.e., $f(e_1)$ is a mapping assigning to any nonempty subset $X \subseteq \Omega$ a number $f(e_1)(X)$. Consistently with the principle of accumulation, we suppose that $f(e_1)(\cdot)$ takes nonnegative values.

At the next time step $t = 2$, we combine the new evidence e_2 with m_1, which counts twice in the summation as being already a combination of two bmas:

$$m_2 = \frac{2m_1 + f(e_2)}{3}.$$

Iterating the process and putting f as a superindex to emphasize dependency on f, we find:

$$m_t^f = \frac{t \cdot m_{t-1}^f + f(e_t)}{t+1}. \tag{3}$$

Lemma 1. *For any $t > 0$, m_t^f is a bma iff $f(e)$ is a bma for every observation $e \subseteq \Omega$.*

Proof. $\Rightarrow$) Recall that we made the assumption that $f(e)$ is a nonnegative function for any e.

Let us prove that $f(e) \leqslant 1$. Suppose on the contrary that at some time t with evidence e_t, we have $f(e_t)(X) > 1$ for some $\varnothing \neq X \subseteq \Omega$. Then from (3), we get

$$t \cdot m_{t-1}^f(X) + f(e_t)(X) = (t+1)m_t^f(X) \leqslant t+1$$

since m_t^f is a bma, which is equivalent to

$$\underbrace{f(e_t)(X) - 1}_{>0} \leqslant \underbrace{t(1 - m_{t-1}^f(X))}_{\leqslant 0},$$

a contradiction.

Let us prove that $\sum_{X\subseteq\Omega} f(e)(X) = 1$ for every e. We have for every e_t, by assumption on m_t^f, m_{t-1}^f:

$$1 = \sum_{X\subseteq\Omega} m_t^f(X) = \frac{1}{t+1}\Big(t\cdot\sum_{X\subseteq\Omega} m_{t-1}^f(X) + \sum_{X\subseteq\Omega} f(e_t)(X)\Big) = \frac{1}{t+1}\Big(t + \sum_{X\subseteq\Omega} f(e_t)(X)\Big),$$

implying that $\sum_{X\subseteq\Omega} f(e_t)(X) = 1$.

$\Leftarrow$) We proceed by induction on t. By assumption, m_0 is a bma. For the induction step, for some $t \geqslant 1$, suppose that m_{t-1}^f is a bma and let us prove that m_t^f is a bma.

As f and m_{t-1}^f take values in $[0,1]$, we have clearly $m_t^f \geqslant 0$. Also, for every $X \subseteq \Omega$, $X \neq \varnothing$,

$$m_t^f(X) \leqslant \frac{t\cdot m_{t-1}^f(X)+1}{t+1} \leqslant \frac{t+1}{t+1} = 1.$$

Let us prove that $\sum_{X\subseteq\Omega} m_t^f(X) = 1$. We have

$$\sum_{X\subseteq\Omega} m_t^f(X) = \frac{1}{t+1}\left(t\sum_{X\subseteq\Omega} m_{t-1}^f(X) + \sum_{X\subseteq\Omega} f(e_t)(X)\right) = \frac{1}{t+1}(t+1) = 1. \quad \square$$

Given a time step t with current bma m_{t-1}^f, it remains to specify to which variables should the bma $f(e_t)$ depend. For the sake of generality, and in particular to model some bias of the agent induced by her current state of knowledge, we assume that $f(e_t)$ may depend on the current bma m_{t-1}^f, and for this reason we write explicitly $f(m_{t-1}^f, e_t)$. However, we do not assume any dependency on time. That is, we state the principle of *time invariance*: for any positive integer Δ,

$$f(m_{t-1}^f, e_t) = f(m_{t-1+\Delta}^f, e_{t+\Delta})$$

provided that $m_{t-1}^f = m_{t-1+\Delta}^f$ and $e_{t+\Delta} = e_t$.

Summarizing the above discussion, the *general rule* is defined by

$$m_t^f(X) = \frac{t\cdot m_{t-1}^f(X) + f(m_{t-1}^f, e_t)(X)}{t+1}, \quad (\varnothing \neq X \subseteq \Omega), \tag{4}$$

where $f : \mathcal{M}(\Omega) \times (2^\Omega \setminus \{\varnothing\}) \times (2^\Omega \setminus \{\varnothing\}) \to [0,1]$ is a function satisfying the property

$$\sum_{\varnothing\neq X\subseteq\Omega} f(m,Y)(X) = 1.$$

Such a function f satisfying the above condition is called an *updating function* or *updator*. Lemma 1 tells us that $f(m,e)$ can be fully considered as a bma, modeling the observation e in the current state of knowledge m. Consequently, our general updating rule is clearly of the averaging type, and therefore differs from more classical conjunctive and disjunctive combination rules.

The next proposition gives some basic properties of this class of updating rules. First, for some nonempty $X \subseteq \Omega$, we say that an updating rule based on updating function f is *strictly monotone at* X if for all $t > 0$, $m_t^f(X) > m_{t-1}^f(X)$ whenever $e_t = X$ and $m_{t-1}^f(X) < 1$. It means that receiving a piece of evidence X always increases the mass on X. Note that X may be equal to Ω, in which case receiving one piece of totally ambiguous evidence augments the ignorance of the agent. By contrast, the third property tells under which condition fully ambiguous evidence does not change the state of knowledge of the agent.

Proposition 1. *Let f be an updating function and its corresponding updating rule. The following holds:*

1. *If m_{t-1}^f is a bma, then m_t^f is a bma.*
2. *The updating rule is strictly monotone at some nonempty $X \subseteq \Omega$ iff*

$$f(m_{t-1}^f, X)(X) > m_{t-1}^f(X), \forall t = 1, 2, \ldots$$

3. *Suppose $e_t = \Omega$. Then $m_t^f = m_{t-1}^f$ iff $f(m_{t-1}^f, \Omega) = m_{t-1}^f$.*

Proof. 1. See Lemma 1.

2. By definition we have

$$m_t^f(X) - m_{t-1}^f(X) = \frac{1}{t+1}\Big[-m_{t-1}^f(X) + f(m_{t-1}^f, X)(X)\Big]$$

which is positive iff $f(m_{t-1}^f, X)(X) > m_{t-1}^f(X)$.

3. By definition we have

$$(t+1)m_t^f = tm_{t-1}^f + f(m_{t-1}^f, \Omega),$$

hence letting $m_t^f = m_{t-1}^f$ yields the result. □

We postpone to Section 3.5 the study of convergence, and we hereafter introduce some particular cases of the general rule.

3.2. The unbiased rule

We start by taking one of the most simple forms for the updating function, which does not depend on the current bma:

$$f(m_{t-1}, e_t)(X) = \delta_{e_t}(X)$$

with $\delta_{e_t}(X) = 1$ if $e_t = X$, and 0 otherwise. Observe that if a piece of evidence $e_t = X$ is received, only the value of the bma at X is modified, nothing else. Moreover, the modification is maximal (value 1). This can be seen as an exact counterpart of the estimation of probabilities by frequencies. The rule is called “unbiased” because the current belief of the agent, m_{t-1}, is not used in the update process.

Its explicit form is:

$$m_t(X) = \begin{cases} \dfrac{t \cdot m_{t-1}(X) + 1}{t+1}, & \text{if } e_t = X \\ \dfrac{t \cdot m_{t-1}(X)}{t+1}, & \text{otherwise} \end{cases}, \quad (X \subseteq \Omega, X \neq \emptyset). \tag{5}$$

It can be easily checked that

$$m_t(X) = \frac{m_0(X) + n_t(X)}{t+1}, \quad \emptyset \neq X \subseteq \Omega, \tag{6}$$

where $n_t(X)$ is the number of occurrences of X in the sequence of evidence $e_0, \dots, e_t$. This formula shows that the order in the sequence of signals is unimportant. Also, as t tends to infinity, the influence of the prior belief tends to 0, and the mass distribution tends to the frequency of the events in the sequence, as for the law of large numbers.

Let us illustrate the rule by the following running example.

Example 1. Consider $\Omega = \{a, b\}$ with two states of nature, e.g., a means that climate change is real and b means that there is no climate change. A piece of evidence e_t is brought by regular meteorological bulletins, supporting either a or b, or being ambiguous ($e_t = \Omega$).

Consider 3 agents, with the following prior bmas m_0^1, m_0^2, m_0^3:

	$\{a\}$	$\{b\}$	Ω
m_0^1	1	0	0
m_0^2	0	1	0
m_0^3	0	0	1

Agent 1 is fully convinced that climate change is real, while Agent 2 is fully convinced that this is a conspiracy theory, and Agent 3 comes from the Moon and has absolutely no opinion on the topic.

Consider first the following sequence of non-ambiguous evidence (omitting braces and commas for short): a, a, b, a, b, where there are slightly more pieces of evidence of a than b. Applying the unbiased rule 5 times yields the following (using (6)):

	$\{a\}$	$\{b\}$	Ω
m_5^1	$2/3$	$1/3$	0
m_5^2	$1/2$	$1/2$	0
m_5^3	$1/2$	$1/3$	$1/6$

One can see how the mass distribution of the agents tends to the frequentist distribution of the events: $3/5$ for a and $2/5$ for b. Observe also how the mass of ignorance diminishes for Agent 3.

Let us now consider that in the previous sequences, the two last pieces of evidence become ambiguous: a, a, b, ab, ab. We obtain the following:

	$\{a\}$	$\{b\}$	Ω
m_5^1	1/2	1/6	1/3
m_5^2	1/3	1/3	1/3
m_5^3	1/3	1/6	1/2

Ambiguity raises the level of ignorance and slows down the convergence to the frequency of appearances of a and b.

The next proposition gives an important property of this rule.

Proposition 2. *The unbiased rule is strictly monotone for every nonempty $X \subseteq \Omega$. More precisely, when $e_t = X$,*

$$m_t(X) - m_{t-1}(X) = \frac{1 - m_t(X)}{t} \geqslant 0.$$

Moreover, $m_t(Y) \leqslant m_{t-1}(Y)$ for $Y \neq X$, more precisely

$$m_t(Y) - m_{t-1}(Y) = -\frac{m_t(Y)}{t} \leqslant 0.$$

Also, $\frac{m_t(Y)}{m_{t-1}(Y)} = \frac{t}{t+1}$.

Proof. Strict monotonicity is obvious from Proposition 1. In addition, from the definition of m_t, we get

$$(t+1)m_t(X) = t \cdot m_{t-1}(X) + 1$$

from which we get the desired expression (and similarly for $m_t(Y)$). □

The property shows a desirable feature: the mass of the received event increases while the other ones decrease. Repeating the same evidence X infinitely many times makes the mass of X converge to 1.

An interesting consequence of this property is the following. Suppose $e_1, \ldots, e_t, \ldots$ are pieces of evidence different from Ω. Then we obtain:

$$m_1(\Omega) = \frac{m_0(\Omega)}{2}, \quad m_2(\Omega) = \frac{m_0(\Omega)}{3}, \ldots, m_t(\Omega) = \frac{m_0(\Omega)}{t+1}, \ldots \tag{7}$$

showing the rapid convergence of $m_t(\Omega)$ to 0.

3.3. Distributive belief update

A distinctive feature of the unbiased rule is that every event is treated similarly, and the fact that the evidence is ambiguous does not matter. In particular, receiving the totally ambiguous signal Ω just augments the mass of Ω. However, we might argue that $e_t = \Omega$ does not bring any information, so why m_t should be different from m_{t-1}?

The distributive rule considers that when an ambiguous piece of evidence X is received, it could have been any subset of X, and the agent updates the masses of every subset of X *in proportion of her belief in these subsets*: the more the agent believes that Z is true, the more she is inclined to "interpret" any new evidence $X \supseteq Z$ as a manifestation of Z. Hence, ambiguous evidence serves to reinforce the belief in less ambiguous evidence.

To distinguish from the previous rule, we denote by $\hat{m}_t$ the updated bma. The form of the updating function is:

$$f(\hat{m}_{t-1}, e_t)(X) = \delta_{\subseteq e_t}(X)\hat{m}_{t-1}(X : e_t),$$

with $\delta_{\subseteq e_t}(X) = 1$ if $X \subseteq e_t$, and 0 otherwise, and

$$\hat{m}(X : Z) := \begin{cases} \dfrac{\hat{m}(X)}{\sum_{Y \subseteq Z} \hat{m}(Y)}, & \text{if } \sum\limits_{Y \subseteq Z} \hat{m}(Y) > 0 \\ 1, & \text{otherwise} \end{cases}, \quad (\emptyset \neq X \subseteq Z). \tag{8}$$

$\hat{m}(X : Z)$ represents the share of the mass of X among all masses of subsets of evidence Z, supposing $X \subseteq Z$, and masses not all zero. The two extreme cases are noteworthy: If $Z = X$ (minimal case), then $\hat{m}(X : Z) = 1$, while if $Z = \Omega$ (maximal case), then $\hat{m}(X : Z) = m(X)$.

The explicit expression of the rule reads:

$$\hat{m}_t(X) = \begin{cases} \dfrac{t \cdot \hat{m}_{t-1}(X) + \hat{m}_{t-1}(X : e_t)}{t+1}, & \text{if } X \subseteq e_t \\ \dfrac{t \cdot \hat{m}_{t-1}(X)}{t+1}, & \text{otherwise} \end{cases}, \quad (X \subseteq \Omega, X \neq \emptyset). \tag{9}$$

Recall the two extreme cases for e_t: If $e_t = X$, then $\hat{m}_{t-1}(X : e_t) = 1$, which is the maximal value that can take the updating function, therefore causing a maximal impact of the evidence on the mass of X; if $e_t = \Omega$, then $\hat{m}_{t-1}(X : e_t) = m(X)$, which causes $\hat{m}_t(X) = \hat{m}_{t-1}(X)$ (see Proposition 3 below). Note that this should not be interpreted as a zero impact on the mass of X, since $\hat{m}_{t-1}(X : e_t) = 0$ would cause a *diminution* of the mass of X.

We give a number of properties of the distributive rule.

Proposition 3. *The distributive rule has the following properties:*

1. *If e_t is a singleton, then $\hat{m}_t = m_t$, i.e., the distributive rule coincides with the unbiased rule.*
2. *If $e_t = \Omega$, then $\hat{m}_t = \hat{m}_{t-1}$.*
3. *Suppose $Z \subseteq X \subseteq Y$. Then*

$$\hat{m}_t(Z \mid e_t = X) \geqslant \hat{m}_t(Z \mid e_t = Y),$$

where the conditioning bar indicates the observed evidence.

Proof. 1. Suppose $e_t = \{\omega\}$. Then $\hat{m}_{t-1}(X : \{\omega\})$ reduces to $\hat{m}_{t-1}(\{\omega\} : \{\omega\}) = 1$. Therefore

$$\hat{m}_t(X) = \begin{cases} \dfrac{t \cdot \hat{m}_{t-1}(\{\omega\}) + 1}{t+1}, & \text{if } X = \{\omega\} \\ \dfrac{t \cdot \hat{m}_{t-1}(X)}{t+1}, & \text{otherwise} \end{cases},$$

which coincides with the unbiased rule.

2. Suppose $e_t = \Omega$. By Proposition 1, it suffices to check that $f(\hat{m}_{t-1}, \Omega) = \hat{m}_{t-1}$. We have

$$f(\hat{m}_{t-1}, \Omega)(X) = \delta_{\subseteq \Omega}(X)\hat{m}_{t-1}(X : \Omega) = \frac{\hat{m}_{t-1}(X)}{\sum_{Y \subseteq \Omega} \hat{m}_{t-1}(Y)} = \hat{m}_{t-1}(X).$$

3. Suppose $Z \subseteq X \subseteq Y$. Observe that

$$\hat{m}_{t-1}(Z : X) = \frac{\hat{m}_{t-1}(Z)}{\sum_{K \subseteq X} \hat{m}_{t-1}(K)} \geqslant \frac{\hat{m}_{t-1}(Z)}{\sum_{K \subseteq Y} \hat{m}_{t-1}(K)} = \hat{m}_{t-1}(Z : Y)$$

because $\hat{m}_{t-1}$ is nonnegative as it is a bma. The desired result follows from this inequality. □

The last property shows that belief increases more when the evidence is less ambiguous.

We may wonder if a kind of monotonicity property holds, as for the unbiased rule (see Proposition 2), but this is not true. Indeed, supposing $e_t = Z$ is received and $\sum_{Y \subseteq Z} \hat{m}_{t-1}(Y) > 0$, it is not possible to tell whether $\hat{m}_t(X)$ is greater or smaller than $\hat{m}_{t-1}(X)$ for $X \subseteq Z$, since we get from (9)

$$\hat{m}_t(X) - \hat{m}_{t-1}(X) = \frac{1}{t}\Big(\frac{\hat{m}_{t-1}(X)}{\sum_{Y \subseteq e_t} \hat{m}_{t-1}(Y)} - \hat{m}_t(X)\Big)$$

when $\sum_{Y \subseteq e_t} \hat{m}_{t-1}(Y) \neq 0$. The sign of the expression in parentheses may be positive or negative, depending on the values taken by $\hat{m}_t$ and $\hat{m}_{t-1}$. However, as for the unbiased rule, if $X \not\subseteq e_t$, then $\hat{m}_t(X)$ decreases.

Example 2 *(Example 1 cont'd).* Let us apply the distributive rule to our previous example. Since for the sequence a, a, b, a, b of non-ambiguous events, the result is identical (see Proposition 3 (2)), we turn to the second sequence a, a, b, ab, ab. We obtain:

	$\{a\}$	$\{b\}$	Ω
$\hat{m}_5^1$	3/4	1/4	0
$\hat{m}_5^2$	1/2	1/2	0
$\hat{m}_5^3$	1/2	1/4	1/4

From Proposition 3 (2) and (3), we get in fact $\hat{m}^i_5 = m^i_3$ for $i = 1, 2, 3$. We observe that in contrast to the unbiased rule, agents 1 and 2 have no ignorance, as the ambiguous evidence is not taken into account.

We present a second example to illustrate the computation with three states of nature.

Example 3. Let $\Omega = \{a, b, c\}$ be the set of all possible outcomes. Let

Event X	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	Ω
$m_0(X)$	0.3	0.1	0.2	0.1	0.1	0.1	0.1

be the prior distribution m_0 of an agent. Let $e_1 = \{a, b\}$ be the evidence at time 1. Then, the distributive update is obtained as follows (we omit braces and commas):

$$\begin{aligned}
\hat{m}_1(a) &= \frac{m_0(a) + \frac{m_0(a)}{m_0(a)+m_0(b)+m_0(ab)}}{2} = \frac{0.3 + \frac{0.3}{0.3+0.1+0.1}}{2} = 0.45\\
\hat{m}_1(b) &= \frac{0.1 + \frac{0.1}{0.3+0.1+0.1}}{2} = 0.15\\
\hat{m}_1(c) &= \frac{m_0(c)}{2} = 0.1\\
\hat{m}_1(ab) &= \frac{0.1 + \frac{0.1}{0.3+0.1+0.1}}{2} = 0.15\\
\hat{m}_1(ac) &= \frac{m_0(ac)}{2} = 0.05\\
\hat{m}_1(bc) &= \frac{m_0(bc)}{2} = 0.05\\
\hat{m}_1(abc) &= \frac{m_0(abc)}{2} = 0.05.
\end{aligned} \tag{10}$$

Notice that the mass of a increased more than the mass of b, while the relative ratio remained the same. Now let us calculate the $\mathtt{BetP}$ of these masses for the previous turn: $\mathtt{BetP}_{t-1}(\mathtt{a}) = 0.4333$, $\mathtt{BetP}_{t-1}(\mathtt{b}) = 0.2333$, $\mathtt{BetP}_{t-1}(\mathtt{c}) = 0.3333$. And the updated betting probabilities are: $\mathtt{BetP}_{t}(\mathtt{a}) = 0.56666$, $\mathtt{BetP}_{t}(\mathtt{b}) = 0.26666$, $\mathtt{BetP}_{t}(\mathtt{c}) = 0.1666$. Even though the evidence supported both states of nature a and b, the agent's prior caused the agent to support evidence a more. This result parallels [6] in terms of ambiguous evidence.

3.4. The biased rule when $\Omega = \{a, b\}$

We place ourselves in the case of two possible states of nature, as in our running example with climate change. The idea behind this rule is the following: when the agent is faced with ambiguous evidence Ω, she updates her belief on a and b in proportion to her current belief as in the distributive rule, but also in proportion to her level of certainty, that is, $1 - m_{t-1}(\Omega)$ (recall that $m_{t-1}(\Omega)$ is the level of ignorance). In other words, the more certain is the agent, the greater will be the update. We use the term "biased" to indicate that the agent, when updating, is inclined to interpret ambiguous evidence as something supporting her current belief in states of nature. In this respect, the distributive rule is also biased.

Let us denote by $\tilde{m}_{t+1}$ the bma updated by the biased rule. When $e_t = \{a\}$ or $e_t = \{b\}$, the formula is identical to the unbiased one: $\tilde{m}_t(X) = m_t(X)$ for all $X \subseteq \Omega$, $X \neq \emptyset$. In other words,

$$f(\tilde{m}_{t-1}, e_t) = \delta_{e_t}, \quad \text{if } e_t = \{a\} \text{ or } e_t = \{b\}.$$

When $e_t = \Omega$, we set, according to the above description:

$$f(\tilde{m}_{t-1}, \Omega)(X) = \tilde{m}_{t-1}(X)(1 - \tilde{m}_{t-1}(\Omega)), \quad (X = \{a\} \text{ or } X = \{b\}).$$

In order to ensure the condition $\sum_{\emptyset \neq X \subseteq \Omega} f(\tilde{m}_{t-1}, e_t)(X) = 1$, we put

$$\begin{aligned}
f(\tilde{m}_{t-1}, \Omega)(\Omega) &= 1 - \tilde{m}_{t-1}(\{a\})(1 - \tilde{m}_{t-1}(\Omega)) - \tilde{m}_{t-1}(\{b\})(1 - \tilde{m}_{t-1}(\Omega))\\
&= 1 - (1 - \tilde{m}_{t-1}(\Omega))^2 = \tilde{m}_{t-1}(\Omega)(2 - \tilde{m}_{t-1}(\Omega)).
\end{aligned}$$

Remark 1. It seems difficult to generalize this rule to more than two states of nature. One reason is to properly define a level of ignorance for ambiguous evidence different from Ω.

In summary, when $e_t = \Omega$, the formula is:

$$\tilde{m}_t(X) = \begin{cases} \dfrac{t \cdot \tilde{m}_{t-1}(X) + \tilde{m}_{t-1}(X)(1 - \tilde{m}_{t-1}(\Omega))}{t+1}, \text{ if } X = \{a\} \text{ or } X = \{b\} \\ \tilde{m}_{t-1}(X)\Big(\dfrac{t+2-\tilde{m}_{t-1}(X)}{t+1}\Big), \text{ if } X = \Omega. \end{cases} \tag{11}$$

This rule has a number of interesting properties, given in the next proposition below.

Proposition 4. *The biased rule has the following properties:*

1. *When the evidence is a singleton, all three rules coincide.*
2. *It is strictly monotone for every nonempty $X \subseteq \Omega$.*
3. *Supposing $e_t = \Omega$ for all t and $0 < m_0(\Omega) < 1$, we have $\tilde{m}_t(\Omega) < m_t(\Omega)$ for all t.*
4. *Suppose $e_t = \Omega$. Then*

$$\frac{\tilde{m}_t(\{a\})}{\tilde{m}_t(\{b\})} = \frac{\tilde{m}_{t-1}(\{a\})}{\tilde{m}_{t-1}(\{b\})}.$$

Proof. 1. Clear.

2. When $e_t \neq \Omega$, the result comes from the fact the biased rule coincides with the unbiased rule (see 1.). Suppose now $e_t = \Omega$. We have from the definition

$$\tilde{m}_t(\Omega) - \tilde{m}_{t-1}(\Omega) = \frac{\tilde{m}_{t-1}(\Omega)}{t+1}(1 - \tilde{m}_{t-1}(\Omega)).$$

As $0 \leq \tilde{m}_{t-1}(\Omega) \leq 1$ for every t, the sequence $\tilde{m}_t(\Omega)$ is nondecreasing, and is strictly increasing iff $0 < \tilde{m}_{t-1}(\Omega) < 1$ for every t. Observe that $m_0(\Omega) = 0$ entails $\tilde{m}_t(\Omega) = 0$ for every t, and $m_0(\Omega) = 1$ entails $\tilde{m}_t(\Omega) = 1$ for every t. It follows that, under that assumption $0 < m_0(\Omega) < 1$, $\tilde{m}_t(\Omega)$ converges to 1.

3. Let us show by induction that starting from the same prior $0 < m_0(\Omega) < 1$, $\tilde{m}_t(\Omega) < m_t(\Omega)$ for every t. From (5) and (11)) with $t = 1$, we obtain:

$$\begin{aligned} \tilde{m}_1(\Omega) &= m_0(\Omega)\frac{3 - m_0(\Omega)}{2} = \frac{3m_0(\Omega) - (m_0(\Omega))^2}{2} \\ &= \frac{m_0(\Omega) + 1 - (m_0(\Omega) - 1)^2}{2} < \frac{m_0(\Omega) + 1}{2} = m_1(\Omega). \end{aligned}$$

Let us assume the hypothesis till $t - 1$. We have from (11) again

$$\begin{aligned} \tilde{m}_t(\Omega) &= \tilde{m}_{t-1}(\Omega)\Big(\frac{t+2-\tilde{m}_{t-1}(\Omega)}{t+1}\Big) = \frac{\tilde{m}_{t-1}(\Omega)(t+2) - (\tilde{m}_{t-1}(\Omega))^2}{t+1} \\ &= \frac{-(\tilde{m}_{t-1}(\Omega) - 1)^2 + 1 + \tilde{m}_{t-1}(\Omega)t}{t+1} \\ &< \frac{\tilde{m}_{t-1}(\Omega) + 1}{t+1} < \frac{m_{t-1}(\Omega)t + 1}{t+1} = m_t(\Omega), \end{aligned} \tag{12}$$

where the last inequality comes from the induction hypothesis.

4. Let us assume a piece of ambiguous evidence, the updated beliefs are:

$$\begin{aligned} \tilde{m}_t(\{a\}) &= \frac{\tilde{m}_{t-1}(\{a\}) \cdot t + \tilde{m}_{t-1}(\{a\}) \cdot (1 - (1 - \tilde{m}_{t-1}(\{a\}) - \tilde{m}_{t-1}(\{b\})))}{t+1} \\ &= \frac{\tilde{m}_{t-1}(\{a\}) \cdot t + \tilde{m}_{t-1}(\{a\})(\tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}{t+1} \end{aligned}$$

and

$$\begin{aligned} \tilde{m}_t(\{b\}) &= \frac{\tilde{m}_{t-1}(\{b\}) \cdot t + \tilde{m}_{t-1}(\{b\}) \cdot (1 - (1 - \tilde{m}_{t-1}(\{a\}) - \tilde{m}_{t-1}(\{b\})))}{t+1} \\ &= \frac{\tilde{m}_{t-1}(\{b\}) \cdot t + \tilde{m}_{t-1}(\{b\})(\tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}{t+1}. \end{aligned}$$

The ratio between new masses is:

$$\frac{\tilde{m}_t(\{a\})}{\tilde{m}_t(\{b\})} = \frac{\frac{\tilde{m}_{t-1}(\{a\}) \cdot t + \tilde{m}_{t-1}(\{a\})(\tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}{t+1}}{\frac{\tilde{m}_{t-1}(\{b\}) \cdot t + \tilde{m}_{t-1}(\{b\})(\tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}{t+1}}$$

$$= \frac{\tilde{m}_{t-1}(\{a\}) \cdot t + \tilde{m}_{t-1}(\{a\})(\tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}{\tilde{m}_{t-1}(\{b\}) \cdot t + \tilde{m}_{t-1}(\{b\})(\tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}$$
$$= \frac{\tilde{m}_{t-1}(\{a\})(t + \tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}{\tilde{m}_{t-1}(\{b\})(t + \tilde{m}_{t-1}(\{a\}) + \tilde{m}_{t-1}(\{b\}))}$$
$$= \frac{\tilde{m}_{t-1}(\{a\})}{\tilde{m}_{t-1}(\{b\})}. \quad \square \tag{13}$$

Suppose $e_t = \Omega$. As observed, the distributive rule yields $\hat{m}_t = \hat{m}_{t-1}$, which is different from the biased rule. Observe that if $\tilde{m}_{t-1}(\Omega) = 1$ (total uncertainty) or $\tilde{m}_{t-1}(\Omega) = 0$ (no uncertainty), then $\tilde{m}_t = \tilde{m}_{t-1}$. In particular, $\tilde{m}_t(\Omega)$ remains 1 or 0, respectively.

Example 4 *(Example 1 cont'd).* Let us apply the biased rule to our previous example. As for the sequence a, a, b, a, b of non-ambiguous events, the result is identical to the previous ones (see Proposition 4 (1)), we turn to the second sequence a, a, b, ab, ab. We obtain:

	$\{a\}$	$\{b\}$	Ω
$\tilde{m}_5^1$	3/4	1/4	0
$\tilde{m}_5^2$	1/2	1/2	0
$\tilde{m}_5^3$	0.4522	0.2261	0.3216

One can see that as expected, $\tilde{m}_5(\Omega)$ remains smaller than $m_5(\Omega)$ (see Proposition 4 (4), and recall that $\tilde{m}_3 = m_3$).

3.5. *Convergence properties of updating rules*

We begin by considering the simple case where all pieces of evidence are identical $e_1 = e_2 = \cdots = e_t = \cdots = X$. The convergence is then given by the strict monotonicity property: if the updating rule m_t^f is strictly monotone at X, then $m_t^f(X)$ converges to 1. This is the case for the unbiased rule and the biased rule for every nonempty $X \subseteq \Omega$. So far, nothing can be said for the distributive rule; however, our analysis below will permit us to derive a result of convergence for the distributive rule.

Let us now study the convergence properties for arbitrary sequences of observations. Let f be an updating function and m_0 some initial bma. Given a sequence of evidence $e_1, e_2, \ldots, e_t$, we have the following:

$$m_1^f = \frac{m_0 + f(m_0, e_1)}{2}$$
$$m_2^f = \frac{m_0 + f(m_0, e_1) + f(m_1^f, e_2)}{3}$$
$$m_3^f = \frac{m_0 + f(m_0, e_1) + f(m_1^f, e_2) + f(m_2^f, e_3)}{4}$$
$$\vdots \quad \vdots$$
$$m_t^f = \frac{m_0 + \sum_{k=0}^{t-1} f(m_k^f, e_{k+1})}{t+1} \leqslant \frac{m_0 + t \cdot 1}{t+1}.$$

It follows that

$$\lim_{t\to\infty} m_t^f = \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} f(m_{k-1}^f, e_k), \tag{14}$$

provided the limit exists. Let us see some particular cases.

1. Suppose $e_t = X$, $t = 1, 2, \ldots$, and consider the unbiased rule. We have in this case $f(m_k, X)(X) = 1$, therefore by (14),
$$\lim_{t\to\infty} m_t(X) = \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} 1 = 1.$$
2. Suppose e_t is a random sequence where $e_t = X$ with probability p, and consider the unbiased rule again. Then, by the strong law of large numbers, $m_t(X)$ converges to p almost surely.
3. Suppose again that $e_t = X$, $t = 1, 2, \ldots$, and consider the distributive rule. The following can be shown.

Proposition 5. *Suppose $m_0(X) > 0$ and that $e_t = X$ for $t = 1, 2, \ldots$. Then*

$$\lim_{t\to\infty} \hat{m}_t(X) = m_0(X : X).$$

The proof relies on the following lemma.

Lemma 2. *Suppose $m_0(X) > 0$ and that $e_t = X$ for $t = 1, 2, \ldots$. Then, by the distributive rule, for any $t = 1, 2, \ldots$,*

$$\hat{m}_t(X : X) = m_0(X : X), \quad \hat{m}_t(Y : X) = m_0(Y : X) \text{ for every } Y \subset X.$$

Proof. We show the result by induction. Let us put

$$m_0(X : X) = \frac{m_0(X)}{\sum_{Y\subset X} m_0(Y) + m_0(X)} =: \alpha,$$

$$m_0(Y : X) = \frac{m_0(Y)}{\sum_{Y\subset X} m_0(Y) + m_0(X)} =: \alpha_Y \text{ for any } Y \subset X.$$

Note that $\alpha_Y = 0$ iff $m_0(Y) = 0$ and that $\sum_{Y\subset X} \alpha_Y + \alpha = 1$. We have

$$\hat{m}_1(X) = \frac{1}{2}\big(m_0(X) + \alpha\big), \quad \hat{m}_1(Y) = \frac{1}{2}\big(m_0(Y) + \alpha_Y\big).$$

Therefore

$$\begin{aligned} m_1(X : X) &= \frac{\hat{m}_1(X)}{\sum_{Y\subset X} \hat{m}_1(Y) + \hat{m}_1(X)} \\ &= \frac{m_0(X) + \alpha}{\sum_{Y\subset X} m_0(Y) + m_0(X) + 1} \\ &= \frac{m_0(X) + \alpha}{\frac{1}{\alpha} m_0(X) + 1} = \alpha. \end{aligned}$$

Similarly, for any $Y \subset X$:

$$\begin{aligned} m_1(Y : X) &= \frac{\hat{m}_1(Y)}{\sum_{Y\subset X} \hat{m}_1(Y) + \hat{m}_1(X)} \\ &= \frac{m_0(Y) + \alpha_Y}{\sum_{Y\subset X} m_0(Y) + m_0(X) + 1} \\ &= \frac{m_0(Y) + \alpha_Y}{\frac{1}{\alpha_Y} m_0(Y) + 1} = \alpha_Y. \end{aligned}$$

Let us suppose the property is true till t and let us prove it for $t + 1$. We have, using the induction hypothesis:

$$\hat{m}_{t+1}(X) = \frac{(t+1)\hat{m}_t(X) + \alpha}{t+2},$$

$$\hat{m}_{t+1}(Y) = \frac{(t+1)\hat{m}_t(Y) + \alpha_Y}{t+2}$$

for any $Y \subset X$. Therefore

$$\begin{aligned} m_{t+1}(X : X) &= \frac{\hat{m}_{t+1}(X)}{\sum_{Y\subset X} \hat{m}_{t+1}(Y) + \hat{m}_{t+1}(X)} \\ &= \frac{(t+1)\hat{m}_t(X) + \alpha}{(t+1)\Big(\sum_{Y\subset X} \hat{m}_t(Y) + \hat{m}_t(X)\Big) + 1} \\ &= \frac{(t+1)\hat{m}_t(X) + \alpha}{\frac{t+1}{\alpha}\hat{m}_t(X) + 1} = \alpha. \end{aligned}$$

Similarly, for any $Y \subset X$:

$$\begin{aligned} m_{t+1}(Y : X) &= \frac{\hat{m}_{t+1}(Y)}{\sum_{Y\subset X} \hat{m}_{t+1}(Y) + \hat{m}_{t+1}(X)} \\ &= \frac{(t+1)\hat{m}_t(Y) + \alpha_Y}{(t+1)\Big(\sum_{Y\subset X} \hat{m}_t(Y) + \hat{m}_t(X)\Big) + 1} \end{aligned}$$

$$= \frac{(t+1)\hat{m}_t(Y) + \alpha_Y}{\frac{t+1}{\alpha_Y}\hat{m}_t(Y) + 1} = \alpha_Y. \quad \square$$

Proof of Proposition 5. By using (14) and Lemma 2, we find, under constant evidence $e_t = X$, $t = 1, 2, \ldots$:

$$\lim_{t\to\infty} \hat{m}_t(X) = \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} f(\hat{m}_{k-1}, X)(X)$$

$$= \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} \hat{m}_{k-1}(X : X)$$

$$= \lim_{t\to\infty} \frac{1}{t} t \cdot m_0(X : X) = m_0(X : X).$$

Similarly,

$$\lim_{t\to\infty} \hat{m}_t(Y) = m_0(Y : X)$$

for any $Y \subset X$. $\square$

4. Agent-based models

In order to understand how beliefs evolve in a social network via different settings and update rules, we create three agent-based models. In the first setting, there is a true state of nature, like in Example 1. In each turn, nature constantly sends signals. After updating their beliefs, agents discuss with their neighbors. In the second setting, we create a model where agents form opinions about an ethical issue with no true state; therefore, agents form opinions via their discussions with their neighbors. Finally, we extend our second model to a three-party setting, e.g., $\Omega = \{l, c, r\}$.

To study opinion dynamics, we create social networks via the Barabasi-Albert algorithm in all of our models [18], which proceeds as follows. In order to generate a network with $N = \{1, 2, \ldots, n\}$, we initially assume $|N| = 1$. Then, in each step t, we add a new agent n_t until $|N| = n$. Agent n_t connects to agent i of degree k_i with probability $p_i = \frac{k_i}{\sum_j k_j}$. This algorithm gives us a scale-free network whose degree distribution follows a power law. Its main characteristic is that few nodes will have many connections, and many nodes will have only a few connections.[2] The main advantage of scale-free networks is that there are few extremely influential agents, which sometimes poses a challenge for asymptotic learning [22].

In this section, we will use three of the rules we described in Section 3: Dempster's combination rule from Equation (1), the distributive rule from Equation (9) when $|\Omega| = 3$, and the biased rule from Equation (11) when $|\Omega| = 2$. The precise way of combination and belief updating will be explained in each subsection. We nevertheless make an essential remark about Dempster's rule of combination. Recall that it is not defined when the level of conflict κ is equal to 1 in order to avoid division by 0. Practically, we do not perform the combination if κ exceeds $d = 0.9$. It means that agents with very different beliefs do not update; hence, their opinions are not affected, as if they would not communicate at all. This reminds the *bounded confidence models* in social networks, where agents delete links with agents with which they have too distant opinions (see [23] for an extensive literature review about bounded confidence models and the effect of the distance).

In our agent-based model, we use a total of 9 parameters. We run each simulation on average $T = 2000$ turns, where beliefs stabilize. Moreover, we create $R = 1000$ simulations for each parameter combination and then take the average results. In order to facilitate network visualization, we take $N = 200$ agents in the network. We have also tested the robustness of our results with 2000 agents, where we have observed no noticeable difference in results. We create agents with a prior belief mass assignment randomly selected from a Dirichlet distribution in each run, e.g., $m_0^i \sim \text{Dir}(\alpha_\emptyset, \ldots, \alpha_\Omega)$, $\forall i \in N$, where $\alpha_\emptyset, \ldots, \alpha_\Omega$ are Dirichlet parameters that denote the weight of outcome of $\emptyset, \ldots, \Omega$ in the prior (see Appendix A for a detailed explanation on our reasoning for using Dirichlet distributions). Table 1 describes the parameters used in the simulation.

4.1. Existence of a true unknown state

This section studies the process of belief update when there is a true state of nature that is unknown, while nature constantly sends a signal revealing the true state with a probability p, the signal being possibly ambiguous. We take Example 1 as framework, with $\Omega = \{a, b\}$ where a means that climate change is real, while b indicates the that there is no climate change. We distinguish two different procedures for updating beliefs, detailed in the following subsections.

[2] There is a dispute in the literature based on the shape of networks encountered in empirical data. Some studies claim that real data follows a scale-free shape [19], while a recent paper challenges that the real data is rarely scale-free [20]. However, this challenge is based on how empirical observations are more complex and have underlying unique properties. Therefore, we have also tested our model using a Watts-Strogatz algorithm [21], where agents are in a "small world", i.e., they can reach one another in very few steps. The significance of our results remains the same.

Table 1
Parameters of simulations.

Parameter	Value range	Description
T	2000	Number of turns per run;
R	1000	Number of runs per parameter combination;
N	200	Number of agents in the network;
α_H	20	A high weight Dirichlet parameter;
α_M	1	An unbiased agent's Dirichlet parameter;
α_L	0.1	A low weight Dirichlet parameter;
p	$(0,5,1]$	The probability of observing the true state if the evidence is non-ambiguous;
q	$[0,1)$	The probability of observing a non-ambiguous evidence;
π	0.01	The proportion of edges that are selected to communicate;
d	0.9	Bounded confidence parameter;
k	1	Neighbors to be connected in Barabasi-Albert algorithm (scale-free).

4.1.1. Communication via Dempster's rule of combination

Supposing the true state of nature is a, nature sends an ambiguous signal Ω with probability 0.1; otherwise, nature sends a signal a with probability 0.8 and b with probability 0.2. The signals sent by nature are independent and identically distributed, and the probabilities of a and b are constant over time. At each time step t:

1. Each agent i updates its bma m^i_{t-1} with the signal X sent by Nature by using the biased rule (Eq. (11)).
2. Select a proportion π of links. For each link ij selected, the agents i, j update their bmas m^i_t, m^j_t by replacing them with $m^i_{t-1} \oplus m^j_{t-1}$ (Dempster's rule), provided the level of conflict κ does not exceed d (bounded confidence). Otherwise, no update occurs.

Algorithm 1 summarizes the whole procedure.

Algorithm 1 Dempster's rule with a true state.

A network $\mathcal{G}$ is randomly generated via Barabasi-Albert algorithm;
for every time step t **do**
  Nature determines e_t:
  if $e_t = X \neq \Omega$ **then**
    $\mathbb{P}(e_t = X) = p_X * q$; $X \neq \emptyset$;
  else
    $\mathbb{P}(e_t = \Omega) = 1 - q$.
  end if
  for every agent i **do**
    $m^i_t(Y)$ is computed by Eq. (11), $\forall Y \subseteq \Omega$;
  end for
  A proportion π of links is selected randomly;
  for every selected link ij **do**
    Compute the level of conflict κ between m^i_t and m^j_t;
    if $\kappa \leqslant d$ **then**
      $m^i_t = m^j_t \leftarrow m^i_t \oplus m^j_t$;
    else
      No update occurs.
    end if
  end for
end for

Fig. 1 shows that agents start fully believing in climate change if nature consistently sends evidence for it. In this case, nature is a constant persuader with a rigid opinion. This persuader effect parallels the literature, where a constant persuader often prevails [4].

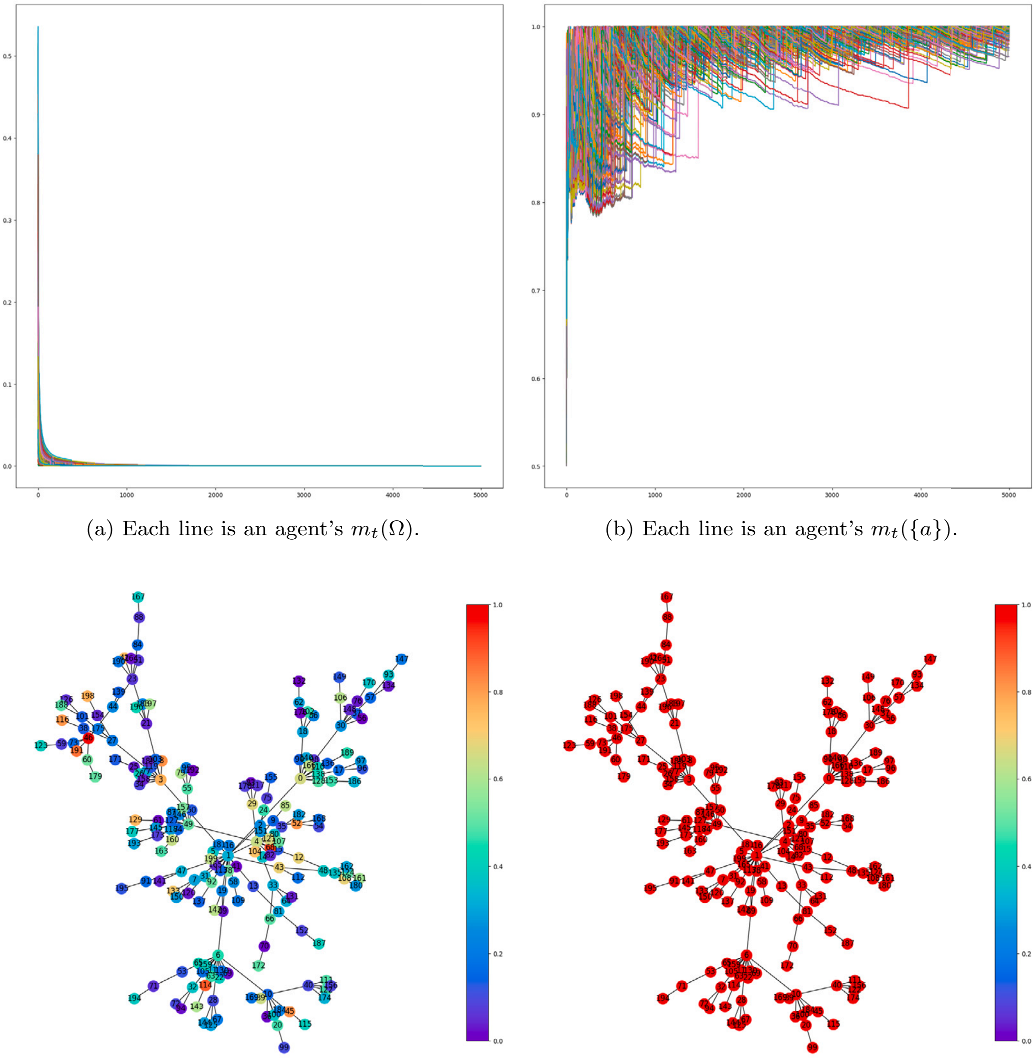
One can see that, despite that there is a uniform distribution for the masses on a and on Ω at the origin, $m_t(\Omega)$ tends rapidly to 0, and eventually $m_t(\{a\})$ converges to 1, for every agent (consensus for climate change). Let us explain these results.

Concerning $m_t(\Omega)$: since $q = 0.9$, most of the time, the evidence is a singleton, therefore the biased rule is equivalent to the unbiased rule (see Proposition 4 (1)), and we can use then Eq. (7), showing that $m_t(\Omega)$ tends to 0 in $1/t$. The fact that Dempster's rule is used also for some agents accelerates this convergence to 0. Indeed, suppose that at time t, agents 1 and 2 have masses $m^1_t(\Omega) = \alpha_1$, $m^2_t(\Omega) = \alpha_2$, with $0 < \alpha_1, \alpha_2 < 1$, and $m^1_t(\{a\}) = 1 - \alpha_1$, $m^2_t(\{b\}) = 1 - \alpha_2$ (conflicting views, i.e., most unfavorable case for the mass of Ω). Then

$$(m^1_t \oplus m^2_t)(\Omega) = \frac{\alpha_1\alpha_2}{1-(1-\alpha_1)(1-\alpha_2)} = \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2-\alpha_1\alpha_2}. \tag{15}$$

Taking, e.g., $\alpha_1 = \alpha_2$, we get

$$(m^1_t \oplus m^2_t)(\Omega) = \frac{\alpha}{2-\alpha} < \alpha.$$



(a) Each line is an agent's $m_t(\Omega)$.

(b) Each line is an agent's $m_t(\{a\})$.

(c) Initial distribution

(d) Final distribution

Fig. 1. Created via Algorithm 1. The color scale indicates the value of BetP({a}) (red: value 1 (climate change), blue: value 0 (no climate change)). $p = 0.8, q = 0.9$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Note that the fact that $m_t(\Omega)$ tends rapidly to 0 implies that $m_t(\{a\})$ and $\mathrm{BetP_t}(\{\mathrm{a}\})$ coincide.

Concerning $m_t(\{a\})$: despite some erratic behavior where the mass of $\{a\}$ decreases slowly, there is always a time t leading to a sudden convergence to 1. The slow decrease of $m_t(\{a\})$ is due to the biased rule. Neglecting the ambiguous evidence ($q = 0.9$), the biased rule coincides with the unbiased rule. It is proved in Section 3.5 that in case of receiving a sequence of identical pieces of evidence a with probability p', $m_t(\{a\})$ converges to p' with the unbiased rule. In our case, $p' = 0.9 \times 0.8 = 0.71$. It seems indeed that $m_t(\{a\})$ is attracted by this value. On the other hand, the sudden step to value 1 is caused by Dempster's rule of combination. As in each turn a pair (i, j) of agents is selected at random for combination, sooner or later every agent is selected, and along with time the proportion of agents with mass $m_t(\{a\}) = 1$ is augmenting. Take agents i, j with $m_t^i(\{a\}) = \alpha < 1$ and $m_t^j(\{a\}) = \beta$, with β close to 1, and suppose t is large enough so that the masses of Ω are 0. Then

$$(m_t^i \oplus m_t^j)(\{a\}) = \frac{\alpha\beta}{1-(1-\alpha)\beta} \approx \frac{\alpha}{1-(1-\alpha)} = 1, \tag{16}$$

whatever the value of α. This shows the drastic behavior of the Dempster rule. See also Appendix B for other simulations with different values of parameter p.

4.1.2. Communication via frequentist update

This section assumes that agents communicate via the biased update rule from Equation (11). Frequentist update ensures that agents are not communicating thoroughly as they did via Dempster's combination rule; instead, they communicate with a softer approach where agents change their opinions slowly. More precisely, at each time step t:

1. Updating with a signal of nature is identical to the previous case.
2. Select a proportion π of links. For each link ij selected:
   (a) Agent i sends a signal $X \subseteq \Omega$ to Agent j with probability $m_t^i(X)$, while simultaneously Agent j sends a signal $Y \subseteq \Omega$ to Agent i with probability $m_t^j(Y)$.
   (b) Both agents i and j update via the frequentist biased rule (11).

A summary of this procedure can be found in Algorithm 2. Then, we present our results in Fig. 2.

Algorithm 2 Frequentist rule with a true state.

A network $\mathcal{G}$ is randomly generated via Barabasi-Albert algorithm;
for every time step t **do**
  Nature determines e_t:
  if $e_t = X \neq \Omega$ **then**
    $\mathbb{P}(e_t = X) = p_X * q$; $X \neq \emptyset$;
  else
    $\mathbb{P}(e_t = \Omega) = 1 - q$
  end if
  for every agent i **do**
    $m_t^i(Y)$ is computed by Eq. (11), $\forall Y \subseteq \Omega$;
  end for
  A proportion π of links is selected randomly;
  for every selected link ij **do**
    Agent i sends signal $X \subseteq \Omega$ to agent j with probability $m_t^i(X)$
    Agent j sends signal $Y \subseteq \Omega$ to agent i with probability $m_t^i(Y)$
    $m_t^i(Z)$ is computed by Eq. (11) with $e_t = X$ and m_t^i, $\forall Z$;
    $m_t^j(Z)$ is computed by Eq. (11) with $e_t = Y$ and m_t^j, $\forall Z$;
  end for
end for

We observe from Fig. 2 that, as in Fig. 1, $m_t(\Omega)$ rapidly tends to 0, although in a slower and more erratic way. This is due to the communication between agents. At the beginning, the mass of Ω for agents is distributed uniformly. Therefore, two randomly picked agents may both send with some probability an ambiguous message, producing a reinforcement of the mass on Ω (see Proposition 4 (2)). As for $m_t(\{a\})$, since only the biased rule governs the convergence, as explained in Section 4.1.1, $m_t(\{a\})$ converges to a value close to $pq = 0.72$. However, it is not easy to separate the effect of communication when there is a constant persuader: nature. The difference here is that agents take tens of thousands of turns to converge to the true state, while Dempster's rule stabilizes after 3000 turns on average. Results of this model are similar to [24] due to the existence of consensus achieved from communication.

Finally, the opinion difference between agents gets smaller over time in Fig. 2b, which indicates no polarization as everyone has a similar belief.

4.2. No evidence from nature

In this section, we study an example where agents try to form opinions about an issue without obtaining evidence from nature. This situation is similar to forming opinions about political or social issues where there is no true state. Let us consider $\Omega = \{a, b\}$ with two ethical propositions, e.g., a indicates a pro-choice opinion, and b indicates a pro-life opinion. We test the belief evolution in a network using the combination and frequentist rules.

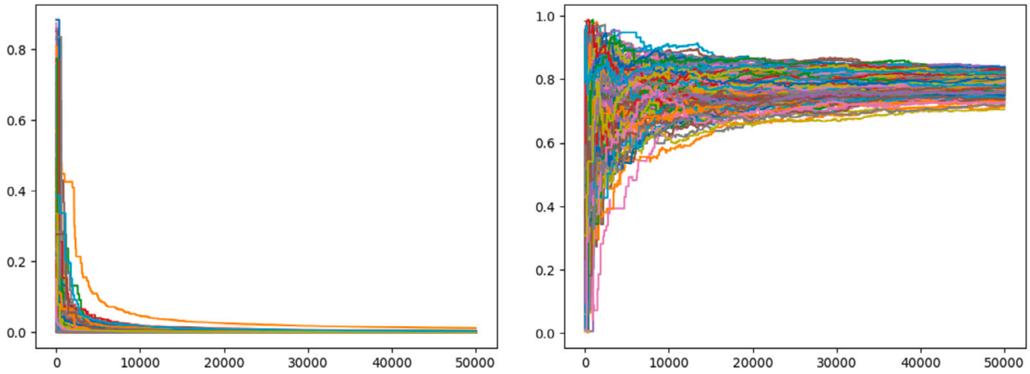
4.2.1. Communication via the Dempster's rule of combination

In this model, at each time step t:

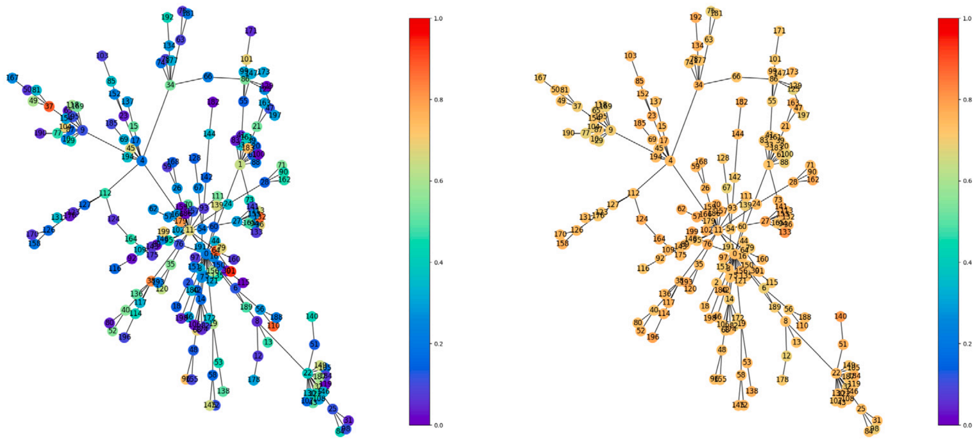
1. Select a proportion π of links.
2. For each link ij selected, the agents i, j update their `bmas` m_t^i, m_t^j by replacing them by $m_{t-1}^i \oplus m_{t-1}^j$ (Dempster's rule), provided the level of conflict κ does not exceed d (bounded confidence). Otherwise, no update occurs.

Algorithm 3 summarizes this procedure.

Fig. 3 reveals total polarization if agents communicate with their neighbors using Dempster's rule. Surprisingly enough, this polarization remains even when $d = 0.999$, indicating that agents are highly tolerant and communicate with each other unless they are on the opposite side of the spectrum. Moreover, we tested the robustness of our results under small-world graph generation, and the results remain robust.



(a) Each line is an agent's $m_t(\Omega)$. (b) Each line is an agent's $m_t(\{a\})$.



(c) Initial distribution (d) Final distribution

Fig. 2. Created via Algorithm 2. Red indicates full belief in climate change ($\mathrm{BetP}(\{a\}) = 1$).

Algorithm 3 Dempster's rule with no true state.

A network $\mathcal{G}$ is randomly generated via Barabasi-Albert algorithm;
for every time step t **do**
    A proportion π of links is selected randomly;
    for every selected link ij **do**
        Compute the level of conflict κ between m_t^i and m_t^j;
        if $\kappa \leqslant d$ **then**
            $m_t^i = m_t^j \leftarrow m_t^i \oplus m_t^j$;
        else state No update occurs;
        end if
    end for
end for

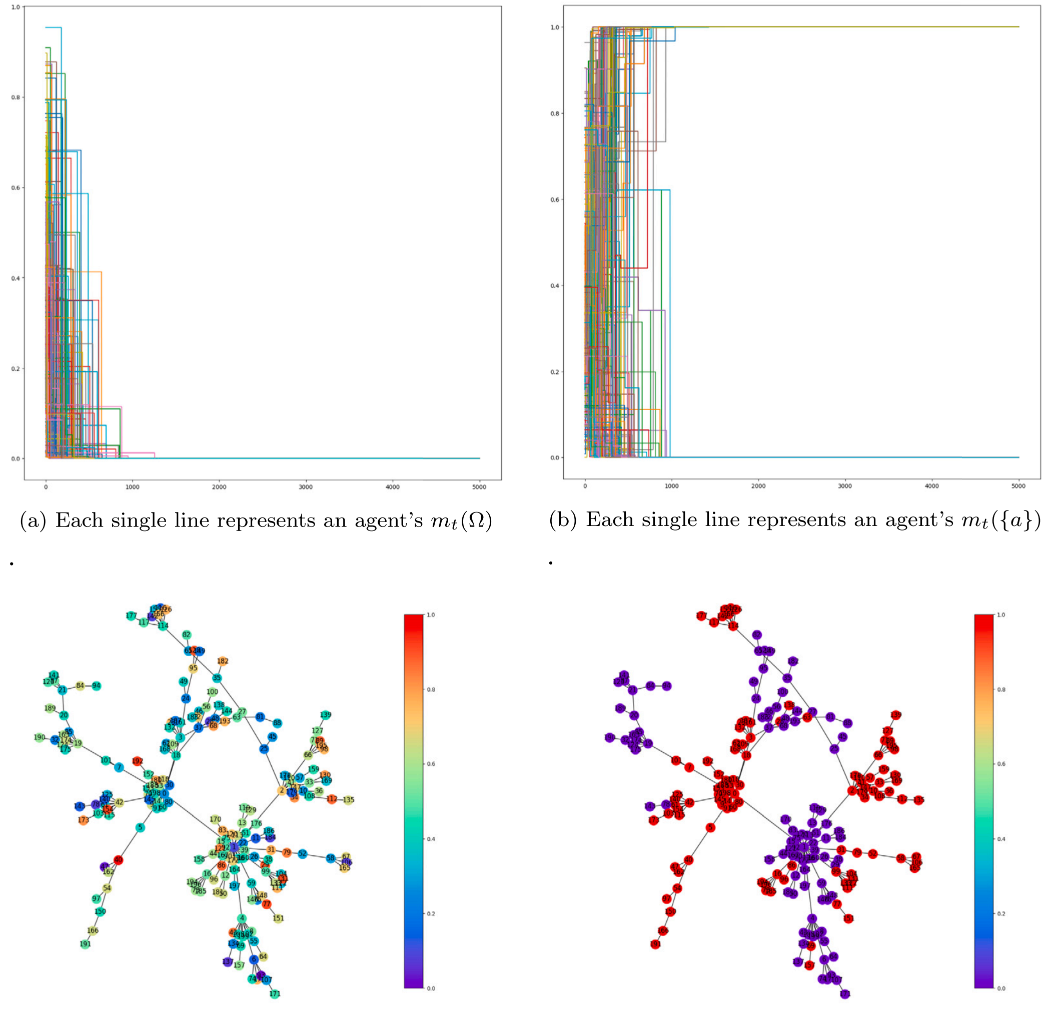
More specifically, we see that $m_t(\Omega)$ rapidly reaches the value 0 for each agent, but in a way very different from previous cases. Instead of a smooth curve in $1/x$, we have a stair-case shape with high steps. This can be explained via Eq. (15). Let us denote as above by α_1, α_2 the masses on Ω for agents 1 and 2. Suppose that α_1 is close to 0, while α_2 is much larger. Then we can write

$$(m_t^1 \oplus m_t^2)(\Omega) = \frac{\alpha_1\alpha_2}{\alpha_1 + \alpha_2 - \alpha_1\alpha_2} \approx \frac{\alpha_1\alpha_2}{\alpha_2} = \alpha_1.$$

Hence $m_t(\Omega)$ for agent 2 changes in one step from a high value to a value close to 0. Let us now explain the behavior of $m_t(\{a\})$. Taking two agents i, j with masses α, β respectively on $\{a\}$, and mass 0 on Ω, we obtain after combination (see (16)):

$$(m_t^i \oplus m_t^j)(\{a\}) = \frac{\alpha\beta}{1 - (1-\alpha)\beta}.$$

Suppose α is far from the extremes 0 and 1, and β is close to the extremes. Then we obtain:



(a) Each single line represents an agent's $m_t(\Omega)$

(b) Each single line represents an agent's $m_t(\{a\})$

(c) Initial distribution

(d) Final distribution

Fig. 3. Created via Algorithm 3. Red indicates full belief in a ($\mathrm{BetP}(\{a\}) = 1$).

$$(m_t^i \oplus m_t^j)(\{a\}) \approx \frac{\alpha}{\alpha} = 1 \ (\beta \text{ close to } 1); \quad (m_t^i \oplus m_t^j)(\{a\}) \approx \frac{\beta}{1} \approx 0 \ (\beta \text{ close to } 0).$$

Therefore, the opinion of agent i is attracted in one step towards the opinion of agent j, which explains polarization.

When comparing Fig. 3 to Fig. 1, we realize that the constant flow of evidence eventually wears down radical agents in our previous model. The lack of a constant persuader, nature, is a potent source of polarization.

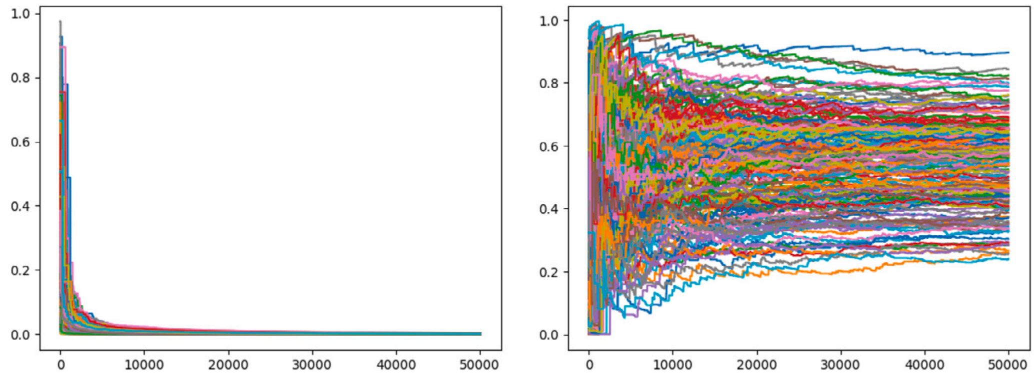
4.2.2. *Frequentist communication*

The procedure here is very similar to the previous case, except that Agent i sends a signal $X \subseteq \Omega$ with a probability m_{t-1}^i to a randomly chosen neighbor j, who updates its bma via the frequentist biased rule. Algorithm 4 describes this process.

Fig. 4 shows a quite different behavior compared to previous figures. As for $m_t(\Omega)$, we observe without surprise the same shape in $1/x$ as before, however, the convergence is much slower and less regular, due to the random selection of agents to be updated, instead of updating all agents at the same time. In addition, since the distribution of $m_0(\Omega)$ is uniform over agents, there is some probability that two selected agents exchange ambiguous signals, reinforcing their mass on Ω. By contrast, the evolution of $m_t(\{a\})$ is radically different from previous cases, as no convergence nor clear structure is apparent, especially in the beginning. Similarly to the case of $m_t(\Omega)$, this is due to the initial uniform distribution and the random selection of links. After 10,000 periods, it seems that the belief of each agent more or less stabilizes. The reason for this is that, as t increases, the impact of a new piece of evidence decreases in $1/t$. Therefore, for $t > 10,000$, the impact becomes negligible.

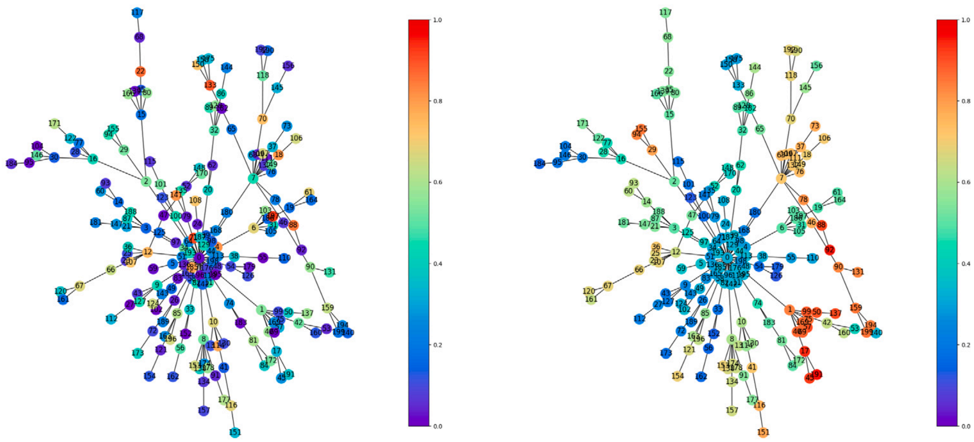
Algorithm 4 Frequentist rule with no true state.

A network $\mathcal{G}$ is randomly generated via Barabasi-Albert algorithm;
for every time step t **do**
  A proportion π of links is selected randomly;
  for every selected link ij **do**
    Agent i sends signal $X \subseteq \Omega$ to agent j with probability $m^i_{t-1}(X)$
    Agent j sends signal $Y \subseteq \Omega$ to agent i with probability $m^i_{t-1}(Y)$
    $m^i_t(Z)$ is computed by Eq. (11) with $e_t = X$ and m^i_{t-1}, $\forall Z$;
    $m^j_t(Z)$ is computed by Eq. (11) with $e_t = Y$ and m^j_{t-1}, $\forall Z$;
  end for
end for



(a) Each line represents a single agent's $m_t(\Omega)$ value.

(b) Each line represents a single agent's $m_t(\{a\})$ value.



(c) Initial distribution

(d) Final distribution

Fig. 4. Created via Algorithm 4. Red indicates full belief in $\{a\}$ ($\mathtt{BetP}(\{a\}) = 1$).

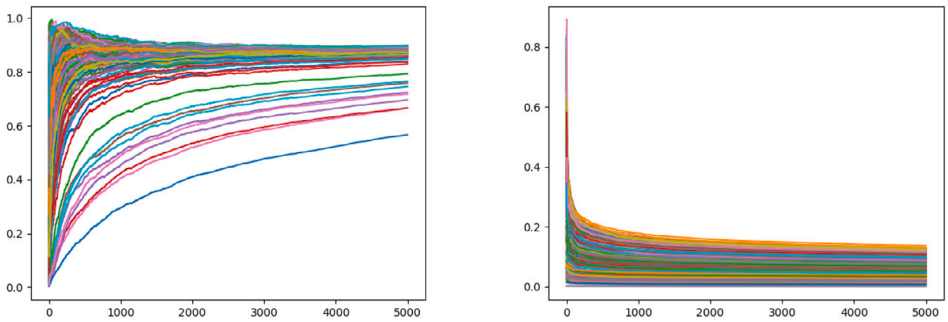
4.3. Ambiguous evidence and recognition of ignorance

An essential contribution of this paper is that it implements a degree of recognition of ignorance. One question that arises from this fact is what happens to the ambiguity of agents under our belief update rules.

In order to understand the role of communication in self-certainty under ambiguous evidence, we study two cases where agents attempt to learn the true state of nature. In the first scenario, we again employ Algorithm 2 where there is a true state and agents use the frequentist rule to update their beliefs under communication. Then, we create a new model where there is only nature and no communication. Agents update their beliefs using the biased update rule. Algorithm 5 summarizes the model.

Algorithm 5 Belief update under constant evidence.

```
A network 𝒢 is randomly generated via Barabasi-Albert algorithm;
for every time step t do
    Nature determines e_t:
    if e_t = X ≠ Ω then
        ℙ(e_t = X) = p_X * q; X ≠ ∅;
    else
        ℙ(e_t = Ω) = 1 − q
    end if
    for every agent i do
        m_t^i(Y) is computed by Eq. (11), ∀Y ⊆ Ω;
    end for
end for
```



(a) No communication (b) Communication via frequentist update

Fig. 5. Fig. 5b is created via Algorithm 2, while Fig. 5a is created via Algorithm 5. Each line shows the level of ignorance of an agent, e.g., $m(\Omega)$. $q = 0.1$, meaning that agents receive almost only ambiguous evidence.

To understand how communication impacts the self-confidence of agents, we study an extreme case where nature *almost* only sends ambiguous evidence, e.g., $q = 0.1$. Our goal here is to emphasize the difference between using communication as a treatment and no communication as a benchmark.

In Fig. 5a (a), we see a slow convergence of $m_t(\Omega)$ towards the value $1 - q = 0.9$. This can be explained, proceeding as in Section 4.1.1. Given that the evidence is not ambiguous, let p be the probability of sending evidence a. Then $m_t(\{a\})$ should converge towards pq, and $m_t(\{b\})$ should converge towards $(1-p)q$. Therefore, $m_t(\Omega)$ must converge to $1 - pq - (1-p)q = 1 - q$. Let us now explain why $m(\Omega)$ is decreasing when communication occurs. Consider agents i, j selected at $t = 0$, with masses of Ω being equal to α and β, respectively. Then, signal Ω is mutually sent by both agents with probability $\alpha\beta$. As α, β are independent random variables with uniform distribution, the distribution of their product follows the law $-\log z$, and the expected value of $\alpha\beta$ is the product of the expected values of α and β, hence $0.5 \times 0.5 = 0.25$. Consequently, at $t = 0$, only for a quarter of selected links the mass of Ω will increase, otherwise it will decrease. When $t > 0$, the expected values of α, β is decreasing with t due to the previous fact, therefore there will be fewer and fewer selected links where the mass of Ω is increasing.

4.4. A political example

Finally, we study an example where $|\Omega| = 3$. Let us assume a political election with three parties: $\Omega = \{l, c, r\}$. In order to observe the impact of certainty, we study two types of agents. First, an agent can be the voter of a party with extreme certainty, e.g., $m(r) > 0.9$. Second, an agent can be indifferent between two parties while having a strong distaste for the third, e.g., $m(\{l, c\}) > 0.9$. We study under which circumstances the party with a certain voter base would prevail. For this section, we use the distributive update rule from Equation (9) since its advantage becomes apparent when we have more than two possible outcomes. Algorithm 6 summarizes the model.

Fig. 6 shows that the right party wins most elections. The difference in initial ambiguity beliefs is apparent. More interestingly, we observe that the impact of increasing the starting age is exceptionally high. This effect is expected since our model increases *belief persistence* with the amount of evidence one receives. Therefore, left and centre-party voters tend to change their opinions since their initial beliefs are more ambiguous, while right-party voters eventually act as persuaders. We also observe that the more connected a network gets, the differences become less noticeable. This indicates that network bottlenecks work in favor of the party that is more certain of herself. Finally, we observe that initial ambiguity helps the party that is certain of themselves (see [25] for an opinion dynamics paper that shows under which conditions polarization benefits the majority candidate).

Algorithm 6 Frequentist rule with three parties.

A network $\mathcal{G}$ is randomly generated via Barabasi-Albert algorithm;
for every time step t **do**
  A proportion π of links is selected randomly;
  for every selected link ij **do**
    Agent i sends signal $X \subseteq \Omega$ to agent j with probability $m^i_{t-1}(X)$
    Agent j sends signal $Y \subseteq \Omega$ to agent i with probability $m^i_{t-1}(Y)$
    $m^i_t(Z)$ is computed by Eq. (11) with $e_t = X$ and m^i_{t-1}, $\forall Z$;
    $m^j_t(Z)$ is computed by Eq. (11) with $e_t = Y$ and m^j_{t-1}, $\forall Z$;
  end for
end for

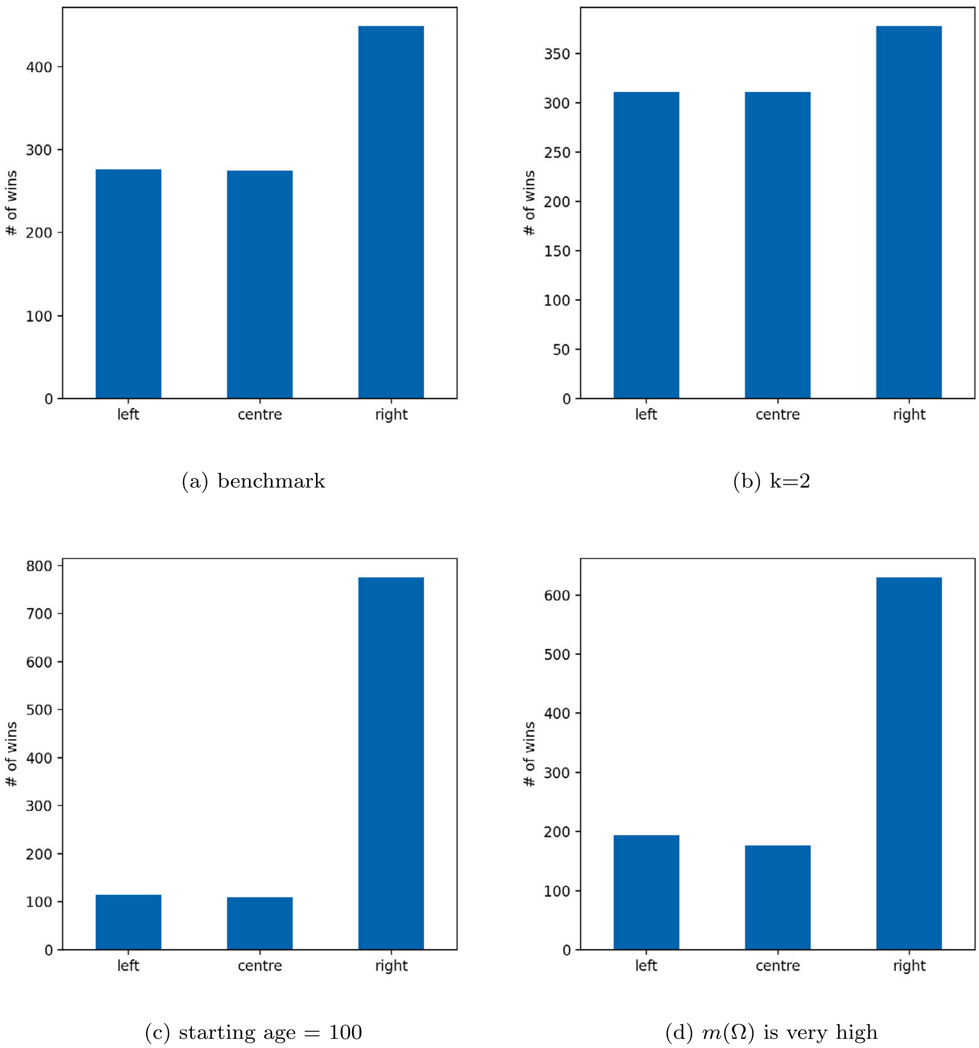


Fig. 6. The effect of initial age, ambiguity, and network connectivity. In the benchmark model, k = 1, starting age is 10, and initial $m(\Omega)$ is low for all agents.

5. Concluding remarks

Dempster-Shafer Theory in opinion dynamics had been relatively barren, with an emerging interest in the last decade [26,27,24]. So far, to our knowledge, our paper is the first study that creates a frequentist model in DST. We study two approaches created in the works of literature of Bayesian and non-Bayesian models, using various update rules in DST. We show that DST is strong and versatile enough to generate various phenomena of social networks.

Climate change was a polarized topic of discussion in the mid-late twentieth century. However, the number of people who believe that climate change is real and that humans are the leading cause is increasing over time [28]. Ethical discussions like abortion [29] or political discussions like fraud in elections, however, cause significant polarization in the population, which we do not observe vanishing. Our preliminary results predict this behavior: agents tend to form a consensus on issues with a constant source of meaningful, accurate information. Meanwhile, a social issue without a definitive true answer creates significant polarization. To summarize, nature eventually acts as a persistent persuader, which has been shown to prevail.

Furthermore, the ambiguity of the evidence changes behavior depending on whether there is communication or not. When there is no communication, agents realize that the evidence is mostly noise. However, communication polarises the community, and agents start believing they are entirely correct and there is no ambiguity. The combination rule might cause the violation of *belief persistence* [30] when two agents with vastly different opinions collide. The proposed model is similar to subtle belief update methodologies observed in heuristic-based models, where agents take one step towards others instead of immediately meeting somewhere in the middle. Hence, this approach is immune to the violation of *belief persistence*.

CRediT authorship contribution statement

Michel Grabisch: Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **M. Alperen Yasar:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Muhammed Alperen Yasar reports financial support was provided by EU Framework Programme for Research and Innovation Marie Sklodowska-Curie Actions. Michel Grabisch reports financial support was provided by EU Framework Programme for Research and Innovation Marie Sklodowska-Curie Actions. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A. Application of Dirichlet distribution

The Dirichlet distribution is a generalization of the beta distribution to higher dimensions and is commonly used in Bayesian statistics for generating random probability vectors that sum to 1 [31,32]. We use these distributions in this study for the construction of priors in Section 4. Formally, the Dirichlet distribution, denoted as $\mathrm{Dir}(\alpha_1, \dots, \alpha_K)$ for a $K-1$ dimensional simplex, is a family of continuous multivariate probability distributions parameterized by a vector $\alpha = (\alpha_1, \dots, \alpha_K)$ of positive reals.

$$f(x_1, x_2, \dots, x_K; \alpha_1, \alpha_2, \dots, \alpha_K) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1} \tag{A.1}$$

where $x_1, x_2, \dots, x_K$ are non-negative real numbers that sum to 1 ($\sum_i^K x_i = 1$), and $B(\alpha)$ is the multinomial beta function, which serves as a normalization constant to ensure that the total probability integrates to 1. The function $B(\alpha)$ is defined as:

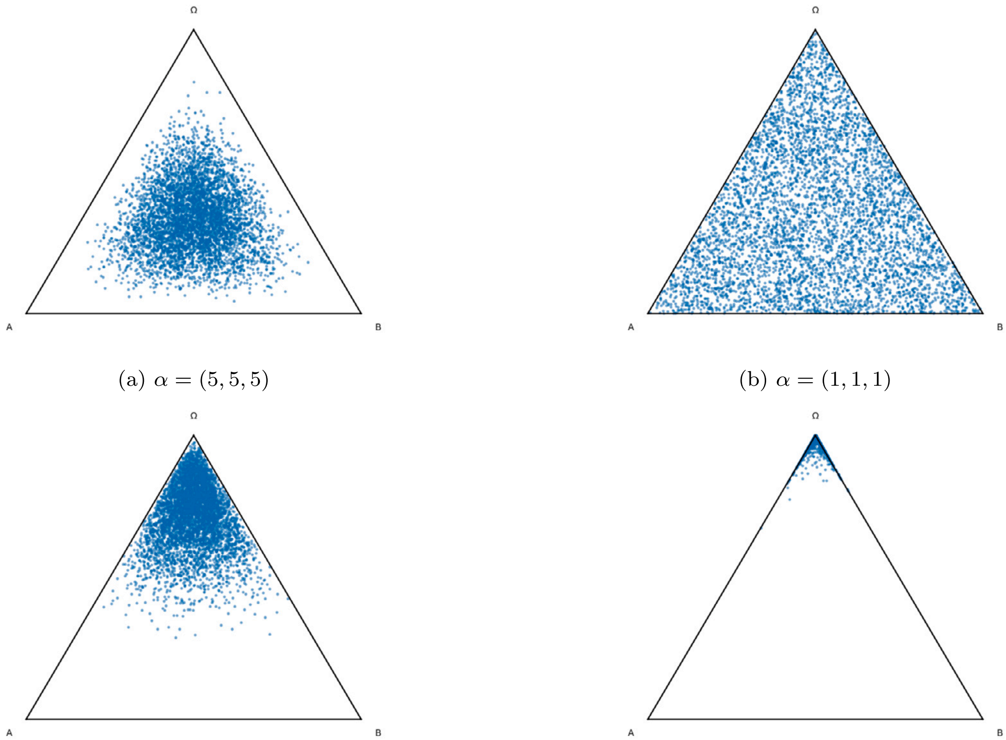
$$B(\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)} \tag{A.2}$$

where $\Gamma(\cdot)$ denotes the gamma function. Its suitability for our model is based on two properties:

1. Conformity to normalization constraint: Given the requirement in DST that the sum of masses across all subsets of the frame of discernment must be unity, the Dirichlet distribution is apt due to its inherent property of generating vectors: $m_0 = (m_0(X_1), \dots, m_0(X_K))$ where $\sum_{i=1}^{K} m_0(X_i) = 1,\ X \subseteq \Omega$.
2. Flexibility in encoding beliefs: The parameters α_i of the Dirichlet distribution directly influence the distribution's shape, allowing for modeling a wide range of initial belief states. This is vital in representing diverse agent perspectives in DST, where α_i essentially controls the agent's initial confidence in the corresponding outcome.

Let $\Omega = \{A, B\}$, in order to create a prior mass $m_0 = \{m(\emptyset), m(\{a\}), m(\{b\}), m(\Omega)\}$, we use a Dirichlet distribution with parameters $(\alpha_a, \alpha_b, \alpha_\Omega)$. Please note that we assume $\alpha_\emptyset = 0$ due to the closed-world hypothesis.

In Section 4, we first create agents randomly in Algorithms 1, 2, 3, 4, and 5. In these models, we have a power set $2^\Omega = \{\emptyset, \{\text{Climate change is real}\}, \{\text{It is not real}\}, \{\text{Either can be true}\}\} = \{\emptyset, \{A\}, \{B\}, \Omega\}$. We assume that an agent will always consider $\alpha = 1$ for all non-empty set events, e.g., $\alpha = (1, 1, 1)$ (we omit the case of the empty set for the sake of simplification). On the other



(a) $\alpha = (5, 5, 5)$ (b) $\alpha = (1, 1, 1)$

(c) $\alpha = (2, 2, 10)$ (d) $\alpha = (0.1, 0.1, 20)$

Fig. A.7. Each triangle represents a 2-simplex where each point inside is a vector (A, B, Ω), where $A + B + \Omega = 1$.

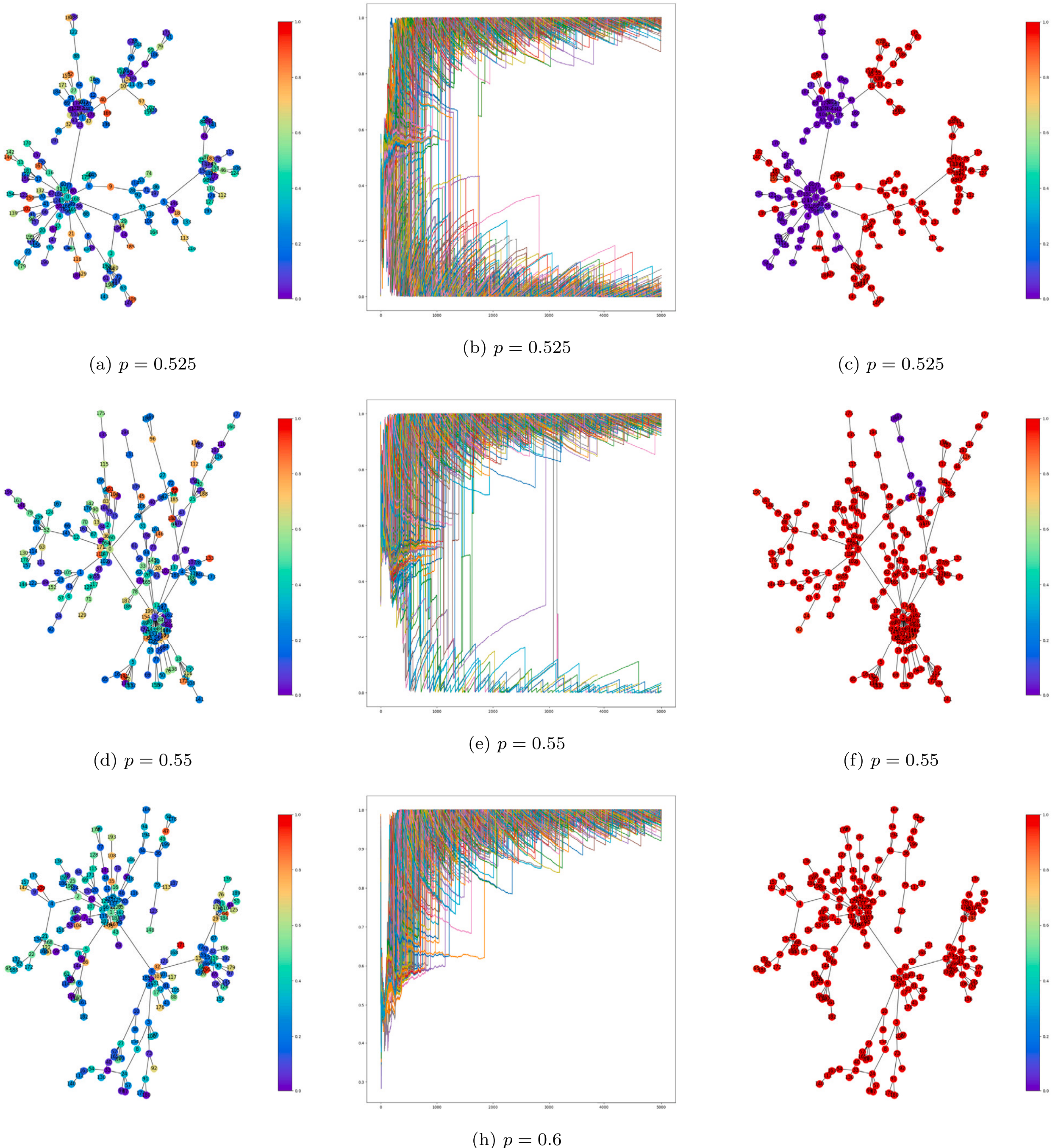
hand, in Algorithm 6, we assume biased agents towards one of the outcomes. Let $\Omega = \{l, c, r\}$; for example, a voter who supports the right party with almost full certainty would have a prior selected from a distribution with parameters $\alpha = (0.1, 0.1, 20, 0.1, 0.1, 0.1, 0.1)$, while a voter who hates the right party while being indifferent between left and center parties would have a prior selected from $\alpha = (0.1, 0.1, 0.1, 20, 0.1, 0.1, 0.1, 0.1)$.

To illustrate how different Dirichlet parameters lead to different distributions, we generate four different plots in Fig. A.7. All parameters are equal in Figs. A.7a and A.7b. It is possible to see that $\alpha = (1, 1, 1)$ from Fig. A.7b is equivalent to a multivariate uniform distribution, where all points are equally likely. In contrast, $\alpha = (5, 5, 5)$ from Fig. A.7a causes more central points to be selected. Then, we create two plots where the distribution favors one of the outcomes in Figs. A.7c and A.7d. We observe that many agents from Fig. A.7c have beliefs that are still close to the center, while Fig. A.7d causes all agents to be biased. In this paper, we used parameters from Fig. A.7b for unbiased agents and those from Fig. A.7d for biased agents.

Appendix B. Persistence of nature

In Section 4, we used several parameter combinations to explain the emergent behavior that arises from various frequentist and combinatory rules. Here, we give an example where we study how the persistence and strength of nature impact the model. Using Algorithm 1, we study three parameters. Figs. B.8a, B.8b, and B.8c use $p = 0.525$, a value close to 0.5, where nature sends a signal suggesting climate change is real almost half the time. Then, we slowly increase this value to first 0.55 in Figs. B.8d, B.8e, and B.8f and then to 0.6 in Figs. B.8g, B.8h, and B.8i.

We observe that agents almost fully polarize when $p = 0.525$. This result is not surprising, since the evidence sent by nature is almost noise. When the evidence from nature is *almost* convincing, we observe that a group of agents stay radical, and they continue not believing in climate change. The reason their beliefs are not being affected by nature is due to the strength of Dempster's rule being overwhelmingly self-reinforcing. We observe that some agents get closer to the center over time when they do not communicate with their neighbors for a while, and then they are immediately brought back to being radical. Over time, frequentist rule causes their beliefs to be more stiffed against evidence from nature.



(a) $p = 0.525$ (b) $p = 0.525$ (c) $p = 0.525$

(d) $p = 0.55$ (e) $p = 0.55$ (f) $p = 0.55$

(g) $p = 0.6$ (h) $p = 0.6$ (i) $p = 0.6$

Fig. B.8. Each of these figures is created via Algorithm 1.

References

[1] A.F. Peralta, M. Neri, J. Kertész, G. Iñiguez, Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions, Phys. Rev. E 104 (4) (2021) 044312.

[2] A. Jern, K.-M.K. Chang, C. Kemp, Belief polarization is not always irrational, Psychol. Rev. 121 (2) (2014) 206.

[3] A.G. Chandrasekhar, H. Larreguy, J.P. Xandri, Testing models of social learning on networks: evidence from two experiments, Econometrica 88 (1) (2020) 1–32.

[4] D. Acemoglu, A. Ozdaglar, Opinion dynamics and learning in social networks, Dyn. Games Appl. 1 (2011) 3–49.

[5] D. Acemoglu, A. Ozdaglar, A. ParandehGheibi, Spread of (mis) information in social networks, Games Econ. Behav. 70 (2) (2010) 194–227.

[6] R.G. Fryer Jr, P. Harms, M.O. Jackson, Updating beliefs when evidence is open to interpretation: implications for bias and polarization, J. Eur. Econ. Assoc. 17 (5) (2019) 1470–1501.

[7] P. Smets, R. Kennes, The transferable belief model, in: Classic Works of the Dempster-Shafer Theory of Belief Functions, 2008, pp. 693–736.
[8] I. Gilboa, D. Schmeidler, Updating ambiguous beliefs, J. Econ. Theory 59 (1) (1993) 33–49.
[9] G. Shafer, A Mathematical Theory of Evidence, vol. 42, Princeton University Press, 1976.
[10] G. Shafer, Jeffrey's rule of conditioning, Philos. Sci. 48 (3) (1981) 337–362.
[11] D. Dubois, T. Denoeux, Conditioning in Dempster-Shafer theory: prediction vs. revision, in: Belief Functions: Theory and Applications: Proceedings of the 2nd International Conference on Belief Functions, Compiègne, France, 9-11 May 2012, Springer, 2012, pp. 385–392.
[12] A. Dominiak, M.S. Lee, Coherent Dempster–Shafer equilibrium and ambiguous signals, J. Math. Econ. 68 (2017) 42–54.
[13] P. Battigalli, G. Bonanno, The logic of belief persistence, Econ. Philos. 13 (1) (1997) 39–59.
[14] A.E. Allahverdyan, A. Galstyan, Opinion dynamics with confirmation bias, PLoS ONE 9 (7) (2014).
[15] M. Grabisch, et al., Set Functions, Games and Capacities in Decision Making, vol. 46, Springer, 2016.
[16] T. Denoeux, Decision-making with belief functions: a review, Int. J. Approx. Reason. 109 (2019) 87–110.
[17] L.S. Shapley, A Value for n-Person Games, Princeton University Press, Princeton, 1953.
[18] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[19] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, 2007, pp. 29–42.
[20] A.D. Broido, A. Clauset, Scale-free networks are rare, Nat. Commun. 10 (1) (2019).
[21] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[22] D. Acemoglu, M.A. Dahleh, I. Lobel, A. Ozdaglar, Bayesian learning in social networks, Rev. Econ. Stud. 78 (4) (2011) 1201–1236.
[23] M. Grabisch, A. Rusinowska, A survey on nonstrategic models of opinion dynamics, Games 11 (4) (2020) 65.
[24] T. Denoeux, Distributed combination of belief functions, Inf. Fusion 65 (2021) 179–191.
[25] M. Grabisch, A. Mandel, A. Rusinowska, On the design of public debate in social networks, Oper. Res. 71 (2) (2023) 626–648.
[26] X. Lu, H. Mo, Y. Deng, An evidential opinion dynamics model based on heterogeneous social influential power, Chaos Solitons Fractals 73 (2015) 98–107.
[27] Z. Huang, W. Li, J. Yang, Belief updating under ambiguity: a numerical simulation analysis, Econ. Lett. 233 (2023) 111359.
[28] T.L. Milfont, M.S. Wilson, C.G. Sibley, The public's belief in climate change and its human cause are increasing over time, PLoS ONE 12 (3) (2017).
[29] A. Adamczyk, C. Kim, L. Dillon, Examining public opinion about abortion: a mixed-methods systematic review of research over the last 15 years, Sociol. Inq. 90 (4) (2020) 920–954.
[30] M.J. Ryan, Violations of belief persistence in Dempster–Shafer equilibrium, Games Econ. Behav. 39 (1) (2002) 167–174.
[31] K.W. Ng, G.-L. Tian, M.-L. Tang, Dirichlet and Related Distributions: Theory, Methods and Applications, John Wiley & Sons, 2011.
[32] C.P. Robert, et al., The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, vol. 2, Springer, 2007.